AIR QUALITY MODELING ANALYSIS FOR THE DENVER EARLY ACTION OZONE COMPACT:

Evaluation of MM5 Simulations of The Summer '02 Denver Ozone Season and Embedded High 8-hr Ozone Episodes

Prepared for:

Mr. Gerald Dilley

Denver Regional Air Quality Council
1445 Market Street, # 260
Denver, CO 80202

Prepared by:

Dennis E. McNally
T. W. Tesche

Alpine Geophysics, LLC
3479 Reeves Drive
Ft. Wright, KY 41017

and

Ralph E. Morris

ENVIRON International Corporation, Inc.,
101 Rowland Way, Suite 220
Novato, CA 94945-5010

1 July 2003

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1.0 INTRODUCTION

This report by describes the results of a meteorological model evaluation study carried out as part of the Denver-Northern Front Range Early Action Compact Study (Denver EAC Study), described in detail in the modeling protocol by Tesche et al., (2003a). As part of the Denver EAC study, the PSU/NCAR Mesoscale Meteorological Model (MM5) was applied to a fifty (50) day long summer ozone period in central Colorado spanning the 6 June-25 July 2002 timeframe. Within this so-called Summer '02 episode, three embedded high 8-hr ozone episodes occurred in the Denver-Northern Front Range Region (DNFRR). These were: (a) Episode 1: (16-22 July 2002), (b) Episode 2: (24 June–2 July 2002), and (c) Episode 3: (8-12 June 2002). MM5 nested meteorological simulations were performed by modelers at Alpine Geophysics in technical consultation with staff at ENVIRON International Corporation (the modeling prime contractor). In this report, we present the results of an operational and limited scientific evaluation of the MM5 model for the Summer '02 episode and the first two intensive embedded periods[1] and assess whether the model's performance in simulating three-dimensional fields of wind, temperature, and moisture (i.e. mixing ratio) are adequate for use in 8-hr ozone modeling over the DNFRR. We also compare the MM5's performance in Episodes 1 and 2 with results from fifty (5) other recent regional modeling studies carried out across the U.S. over the past several years using the MM5 or other contemporary prognostic models.

1.1 Background

As described in the ozone modeling protocol (Tesche et al., 2003a), the goal of the Denver EAC 8-hr Ozone Study is to conduct a comprehensive photochemical modeling study for the Denver-Northern Front Range Region (DNFRR) that can be used as the technical basis for 8-hr ozone SIP development. The modeling study, guided by the protocol, is specifically designed to identify the processes responsible for 8-hr ozone exceedances in the region and to develop realistic emissions reduction strategies for their control. Major objectives of the Denver EAC study include:

> Prepare an Ozone Modeling Protocol, consistent with EPA requirements, that provides direction to the 8-hr ozone modeling of the Denver-Northern Front Range. Collaborate with the CDPHE in the identification and justification of one or more 8-hr ozone modeling episodes for the Denver study;

> Construct dynamically and thermodynamically consistent MM5 meteorological inputs at appropriate grid scales for direct input to the emissions and photochemical models (*the subject of this report)*;

> Produce the model-ready base-year and future-year emissions inventories suitable for input to the CAMx model and perform additional quality assurance (QA) of the emissions data sets beyond that conducted by the CDPHE;

> Develop photochemical model base case modeling inputs for the selected modeling episode(s) and carry out base case model performance testing, diagnostic analysis, and pertinent sensitivity studies, including a check on mass consistency;

> Evaluate the photochemical model's performance for the selected episode(s) and compare

1 The evaluation for Episode 3 is underway and will be reported in a stand-alone technical memorandum issued later.

the results with EPA's performance objectives (EPA, 1991; 1999) for ozone modeling;

> Perform across-the-board VOC and NOx emissions reduction sensitivity simulations to explore the ozone response for the modeling episode(s);

> Perform additional future-year (2007 or 2012) control scenario simulations to estimate ozone levels in the Denver region under different local control regimes (if the future year baseline modeling does not show attainment with the 8-hr NAAQS);

> Develop suitable "weight of evidence" analyses supporting the ozone attainment demonstration, consistent with EPA guidance and assist the RAQC and CDPHE in developing the technical information to support the documentation required for the Denver 8-hr ozone Early Action Compact protocol;

> Provide for a thorough and efficient transfer of modeling codes, data sets, and related information to other stakeholders in the process including the EPA Region VIII and the CDPHE; and

> Set up the full suite of models and databases developed in this study on CDPHE computers and provide on-site training in the use of the modeling system(s).

These main and other subsidiary study objectives will be met following technical approach set forth in the Denver EAC protocol.

Following EPA guidance (EPA, 1991, 1999) the study team recommended three embedded ozone episodes (i.e., 16-22 July; 24 June–2 July; and 8-12 June) within a larger two-month long summer ozone period during 2002. The recommended episodes were reviewed and approved by the Colorado Department of Public Health and Environment (CDPHE) and the Denver Regional Air Quality Council (RAQC) and a Technical Modeling Subcommittee that included representatives from the U.S. EPA, local industry and governments, and other private sector stakeholders.

1.2 MM5 Evaluation Objectives and Methodology

The specific objective of the MM5 evaluation study was to compare the three-dimensional meteorological fields predicted by the model for the three episodes with available surface and aloft data routinely collected by the National Weather Service (NWS) and other reporting agencies. In particular, we attempted to assess the accuracy and reliability of the meteorological fields produced by the model for input into the CAMx nested regional photochemical model. Meteorological inputs required by CAMx include hourly estimates of surface pressure and clouds; the three-dimensional distribution of winds, temperatures, and mixing ratio; and other physical parameters or diagnosed quantities such as turbulent mixing rates (i.e., eddy diffusivities) and planetary boundary layer heights. As described below, the evaluation centered on comparisons between surface and aloft meteorological measurements routinely collected over the DNFRR area the air quality model-ready meteorological fields derived from the MM5 model outputs.

A number of recent studies describe the theoretical formulation and operational features of the MM5 model (see, for example, Dudhia, 1993; Grell et al., 1994; 1998; Seaman, 1995, 1998; Pielke and Pearce, 1994; Seaman et al., 1997) and discuss its performance capabilities under a range of atmospheric conditions (e.g., Cox et al., 1998; Emery et al., 2002; Hanna et. al., 1998; Seaman and Michelson, 1998; Seaman et al., 1992, 1995, 1996; Seaman and Stauffer, 1996;Tesche and McNally, 1993a-f; 1996a McNally and Tesche, 1996a-b, 1998, Tesche et. al., 1997a; 2001, 2003). The results of the present analysis add to this considerable body of knowledge.

As noted, a principal aim of the MM5 evaluation was to assess whether the simulated fields from the meteorological modeling systems may be relied upon to provide wind, temperature, mixing, moisture, and radiation inputs to the CAMx model for typical high 8-hr ozone periods in the Denver-Northern Front Range Region. Here we use the term "modeling system" refers to the model's source code, its preprocessor and data preparation programs, the underlying data base, and the post-processor programs that "map" (i.e., interpolate) the simulated meteorological fields onto the air quality model grid meshes. Ideally, a comprehensive evaluation of the MM5 model would include at least seven steps (Tesche, 1994):

> Evaluate and inter-compare the scientific formulation of the modeling systems via a thorough peer-review process;

> Assess the fidelity of the computer code(s) to scientific formulation, governing equations, and numerical solution procedures;

> Evaluate the predictive performance of individual process modules and preprocessor modules (e.g., advection scheme, subgrid scale processes, closure schemes, planetary boundary layer parameterization, FDDA methodology);

> Carry out diagnostic and/or sensitivity analyses to assure conformance of the modeling systems with known or expected behavior in the real world;

> Evaluate the full modeling system's predictive performance;

> Evaluate the direct meteorological output from the models as well as the "mapped" fields that are processed into air quality model-ready inputs; and

> Implement a quality assurance activity.

Such an intensive evaluation process is rarely, if ever, carried out due to time, resource and data base limitations. The stringent time frame associated with EPA's Early Action Compact process would not allow such a research effort. Nevertheless, it is useful to identify the *ideal* evaluation framework so that the results of the current evaluation can be judged in the proper perspective. This also allows one to set realistic expectations for the reliability and robustness of the actual evaluation findings.

The MM5 modeling system is well established with a rich development and refinement history spanning more than two decades (Seaman, 1998). The model has seen extensive use worldwide by many agencies, consultants, university scientists and research groups. Thus, the current version of MM5 as well as its predecessor versions have been extensively "peer-reviewed" and considerable algorithm development and module testing has been carried out with all of the important process components. Accordingly, the MM5 evaluation in the Denver EAC study focused on the last three steps in the ideal testing process.

Performance testing of the meteorological model is divided into two general categories: operational and scientific. The *operational evaluation* refers to an assessment of the models ability to estimate correctly the atmospheric observations independent of whether the actual process descriptions in the model are accurate (Tesche, 1991a,b). It is an examination of how well the model reproduces the observed meteorological fields in time and space consistent with the input needs of the air quality model. Here, the primary emphasis is on the model's ability to reproduce hourly surface wind speed, wind direction, temperature, and mixing ratio observations across the 36/12/4/1.33 km grid domains. The operational evaluation provides only limited information about whether the results are correct from a scientific perspective or whether they are the fortuitous product of compensating errors; thus a "successful" operational evaluation is a necessary but insufficient condition for achieving a sound, reliable performance testing exercise. An additional, scientific evaluation is also needed.

The *scientific evaluation* attempts to elucidate the realism of the basic meteorological processes simulated by the model. This involves testing the model as an entire system (i.e., not merely focusing on surface wind predictions) as well as its component parts. The scientific evaluation seeks to determine whether the model's behavior in the aggregate and in its component modules is consistent with prevailing theory, knowledge of physical processes, and observations. The main objective is to reveal the presence of bias and internal (compensating) errors in the model that, unless discovered and rectified, or at least quantified, may lead to erroneous or fundamentally incorrect technical or policy decisions.

Unfortunately the scope of the scientific evaluation in the Denver EAC 8-hr ozone study is severely limited by the lack of special meteorological observations (radar profiler winds, turbulence measurements, PBL heights, precipitation and radiation measurements, inert tracer diffusion experiments, and so on). The scientific evaluation component in this study is limited to an appraisal of the model's ability to reproduce wind speed, wind direction, and temperatures aloft based on data collected from available NWS twice-daily soundings. This evaluation is further constrained by the fact that portions of this aloft information was used in the data assimilation scheme used to produce the model's three-dimensional, time dependent fields.

1.3 Report Structure

In chapter 2, we present an overview of the MM5 model followed in Chapter 3 with a brief summary of the 16-22 July and the 24 June–2 July 2002 modeling episodes. Chapter 4 identifies the meteorological modeling domains and the data sets available for model testing. Procedures for model input development and quality assurance are discussed in Chapters 5 and 6, respectively. The specific statistical measures and graphical tools used for performance testing are discussed in Chapter 7. Results from the operational and scientific evaluation exercises are presented in Chapter 8 for the two-month long simulation on the 36/12 km grid and in Chapters 9 and 10 for embedded Episodes 1 and 2. These latter episodes are evaluated on the higher resolution 4/1.33 km grids and the results are compared with other recent prognostic meteorological studies throughout the U.S. In Chapter 11 we assess the suitability of the meteorological modeling fields developed in this study for use in 8-hr ozone modeling in the Denver-Northern Front Range Region. Our summary and conclusions are presented in Chapter 12.

2.0 THE PSU/NCAR MESOSCALE METEOROLOGICAL MODEL

This chapter summarizes the general features of the MM5 prognostic model. For a detailed scientific description of the model the reader is referred to the citations appearing in the Reference section of this report. Table 2-1 identifies the general technical attributes and recent applications of the MM5 model pertinent to air quality studies.

The non-hydrostatic MM5 model (Dudhia, 1993; Grell et al., 1994) is a three-dimensional, limited-area, primitive equation, prognostic model which has been used widely in regional air quality model applications (see, for example, Russell and Dennis, 1997; Seaman et al., 1995,1997; Seaman and Stauffer, 1996; Tesche et al., 1998a; 2003a-d). The basic model has been under continuous development, improvement, testing and open peer-review for more than 20 years (see, for example, Anthes and Warner, 1978; Anthes et al., 1987) and has been used world-wide by hundreds of scientists for a variety of mesoscale studies, including cyclogenesis, polar lows, cold-air damming, coastal fronts, severe thunderstorms, tropical storms, subtropical easterly jets, mesoscale convective complexes, desert mixed layers, urban-scale modeling, air quality studies, frontal weather, lake-effect snows, sea-breezes, orographically induced flows, and operational mesoscale forecasting.

MM5 is based on the prognostic equations for three-dimensional wind components (u, v, and w), temperature (T), water vapor mixing ratio (q_v), and the perturbation pressure (p'). Use of a constant reference-state pressure increases the accuracy of the calculations in the vicinity of steep terrain. The model uses an efficient semi-implicit temporal integration scheme and has a nested-grid capability that can use up to ten different domains of arbitrary horizontal and vertical resolution. The interfaces of the nested grids can be either one-way or two-way interactive. The model is also capable of using a hydrostatic option, if desired, for coarse-grid applications.

MM5 uses a terrain-following non-dimensionalized pressure, or "sigma", vertical coordinate similar to that used in many operational and research models. In the non-hydrostatic MM5 (Dudhia, 1993), the sigma levels are defined according to the initial hydrostatically-balanced reference state so that the sigma levels are also time-invariant. The gridded meteorological fields produced by MM5 are directly compatible with the input requirements of air-quality models using this coordinate, such as SAQM. The fields can be used in other regional air quality models with different coordinate systems (e.g., CAMx, URM, UAM-V, and MAQSIP) by performing a vertical interpolation, followed by a mass-conservation re-adjustment (McNally, 1997).

Distinct planetary boundary layer (PBL) parameterizations are available for air-quality applications, both of which represent sub-grid-scale turbulent fluxes of heat, moisture and momentum. These parameterizations each have a surface energy budget equation to predict the ground temperature (T_g), based on the insolation, atmospheric path length, water vapor, cloud cover and longwave radiation. The surface physical properties of albedo, roughness length, moisture availability, emissivity and thermal inertia are defined as functions of land-use for 14 categories via a look-up table. One scheme uses a first-order eddy diffusivity formulation for stable and neutral environments and a modified first-order scheme for unstable regimes. The other uses a prognostic equation for the second-order turbulent kinetic energy, while diagnosing the other key boundary layer terms.

Initial and lateral boundary conditions are specified from mesoscale three-dimensional analyses performed at 12-hour intervals on the outermost grid mesh selected by the user. Additional surface fields are analyzed at three-hour intervals. A Cressman-based technique is used to analyze standard surface and radiosonde observations, using the National Meteorological Center's (NMC) spectral analysis as a first guess. The lateral boundary data are introduced into MM5 using a relaxation technique applied in the outermost five rows and columns of the most coarse grid domain.

A major feature of the MM5 is its use of state-of-science methods for Four Dimensional Data Assimilation (FDDA). The theory underlying this approach and details on how it has been applied in a variety of applications throughout the country are described in depth elsewhere (Seaman et al., 1992, 1995, 1996, 1997; Tesche and McNally, 1996a,b, 1997; Emery et al., 1999a,b: 2002).

Table 2-1. Attributes of the PSU/NCAR MM5 Prognostic Meteorological Model.

Model Attribute	MM5
Model Name	Mesoscale Meteorological Model, (Version 5)
Developer	Pennsylvania State University, National Center for Atmospheric Research Dudhia (1993); Grell, Dudhia and Stauffer (1994)
Availability	Free, public-domain model
Forecast Variables	Three dimensional wind components, temperature, water vapor, cloud water/ice, rain water/ice, and the perturbation pressure.
Input Requirements	3-hourly surface data and 12-hourly soundings plus gridded pressure level data set (horizontal winds, temp., R.H. as a function of pressure) for model initialization, BC's and FDDA. Also requires topography, sea-surface temp., and land use.
Computer Platforms	Most popular workstations (e.g., SUN SPARCstation, IBM RISC); PC's running LINUX, including clusters.
Hardware Requirements	RAM = ~512 MB; Free hard disk = 100Gb;
Software Requirements	UNIX, FORTRAN 77, NCAR Graphics
Evaluation Studies for Air Quality Model Applications	***Gulf Coast:*** Tesche and McNally (1998c); Douglas et al, (1999) ***NARSTO-NE***: Seaman and Michelson (1998); Tesche and McNally (1996b,f); McNally and Tesche 1996d); Zhang and Rao (1999) ***RADM***: Dennis et al., (1990) ***OTAG***: McNally and Tesche (1996a,b);Tesche and McNally (1996b,d) ***SARMAP***: Seaman, Stauffer, and Lario (1995); Seaman and Stauffer (1996); Tesche and McNally (1993e,f); Tanrikulu (1999); Tesche et al., (1998b) ***LMOS***: Shafran and Seaman (1998); Tesche and McNally (1999c) ***Los Angeles:*** Seaman et al. (1996, 1997); Tesche and McNally (1997c); Pai et al. (1998); Steyn and McKendry (1988); Tesche et al., (1997e) ***Denver Front Range:*** McNally and Tesche (1998c); Tesche et al., (2003b) ***Florida:*** Green et al. (1998) ***Texas Gulf Coast:*** Emery et al., (2001); TCEQ (2002); McNally and Tesche (2002) ***Cincinnati-Hamilton SIP***: Tesche and McNally (1998e) ***Pittsburgh-Beaver Valley SIP:*** McNally and Tesche (1996c) ***Kansas City/St. Louis SIP:*** Emery et al., (1999);McNally and Tesche (1999c).
Peer Review	Pielke (1984); Barchet and Dennis (1990); Tesche and McNally (1993e,f); Pielke and Pierce (1994); Seaman (1995, 2000)
Documentation	5-Volume User's Manuals (Gill, 1992); Twice-annual tutorial classes for new outside users; on-line consultant helpline (NCAR).
Noted Strengths	Supports multi-scale FDDA for both analysis and special asynoptic data; turbulent exchange based on TKE; selection of advanced convective parameterizations.
Noted Limitations	Extended computational time, particularly for smaller (i.e., 4 km or less) grid scales
Equations	Primitive equation model, Non-hydrostatic (non-hydrostatic option)
Grid Differencing Scheme	Arakawa-B staggered.

Model Attribute	MM5
Spatial Resolution -Horizontal -Vertical - Nesting	 Variable (1 to 200 km) Variable/stretched Multiple/2-way/movable during simulation
Coordinate System - Horizontal - Vertical	 Mercator; Lambert Conformal; Polar Stereographic Sigma-p (terrain-following)
Nesting Scheme	Multiple, moving, overlapping nesting with two-way interaction and pre-defined nest ratios of 3:1
Initialization	Cressman objective analysis on pressure surfaces (independent data analysis)
Numerics - Time differencing - Advection	 Leapfrog; split semi-implicit time differencing 4^{th}-order leapfrog
Boundary Conditions - Top - Surface - Lateral	 Absorbing layer Prognostic surface temperature; NCEP/OSU soil-moisture scheme (Seaman, 1998) based on land use. Time-dependent and inflow/outflow dependent
Parameterizations - Radiation - Explicit moist physics - Deep convection -Surface layer - Boundary layer	 Shortwave and longwave schemes that interact with the atmosphere, including cloud and precipitation fields as well as with the surface (Dudhia, 1989). Liquid, ice, and mixed phase Large-scale processes treated explicitly. Various convective precipitation modules available including Kuo (1974), Kain-Fritsch (1990, 1993), Fritsch-Chappell, Betts-Miller, modified Arakawa-Schubert (1974), Grell et al., (1991), and Anthes-Kuo (Anthes, 1977). Heat, momentum, and water vapor fluxes (Blackadar; Gayno-Seaman) Simple bulk aerodynamic parameterization (Blackadar), revised non-local Blackadar (Zhang and Anthes, 1982), Level-2.5 Mellor-Yamada (1974, 1982), or 1.5-order turbulent kinetic energy (TKE) scheme (Gayno et al., 1994).
FDDA Capability	Multi-scale, both analysis-nudging and observation-nudging, data use sensitive to orography, 3-D weighting functions Nudged parameters: *u,v* winds, temperature, water vapor mixing ratio.

3.0 THE MODELING EPISODES

The CDPHE performed an initial identification and screening of candidate 8-hr modeling episodes for the Denver EAC. Key elements of the Denver episode selection process included:

> Identification of the policy and technical issues influencing episode selection for regulatory 8-hr ozone attainment modeling;

> An objective episode selection process based on: (a) analysis of historical air quality and meteorology in the region, (b) synthesis of past studies, and (c) the consideration of the conceptual nature of the types, character, and frequency of occurrence of 8-hr ozone episodes in the Denver-Northern Front Range region; and

> Development of a prioritized list of recommended episodes complete with supporting air quality and meteorological analyses of the preferred period(s).

The ENVIRON/Alpine study team supplemented the agency's analyses with investigations focusing on the suitability of candidates modeling episodes for regulatory emissions, meteorological and ozone modeling. During the 1999 to 2002 time frame, there were five episodes of two days or greater identified by the CDPHE as potentially suited for modeling. The five episodes recommended by the CDPHE (in priority order), are as follows:

> 18-21 July 2002
> 25 June-1 July 2002
> 8-12 June 2002
> 4-9 July 2001
> 3-4 August 2001

The first two episodes, 18-21 July 2002 and 25 June-1 July 2002 were viewed by CDPHE as equally important. As discussed in the Denver EAC modeling protocol (Tesche et a., 2003a), it was decided to exercise the combined emissions, meteorological and photochemical models for the models for the months of June-July (55 days) in one large numerical since it would capture the top three 8-hr modeling episodes and would fulfill EPA's emerging recommendations to model entire ozone seasons (EPA, 2002).

3.1 Characterization of the Denver EAC 8-hr Ozone Modeling Episodes

Meteorological data for the top three episodes were examined by CDPHE staff to establish a general conceptual model for each period. Back trajectories were calculated by for each episode using the NOAA HYSPLIT model. Back trajectories were calculated from the NREL site starting from the location of the 8-hour ozone exceedance and at three different heights above ground level (AGL): surface, 100-m, and 800-m. Different height levels allowed the for the assessment of the transport of low-level air parcels into the area as well as air parcels aloft above ground level. It also provided an indication of the level of wind shear in the atmosphere. The conceptual summaries of the three 8-hr modeling episodes are given below.

3.1.1 16-22 July 2002

The highest ozone levels recorded at Rocky Flats North and NREL over the 1999 through 2002-time period characterized this episode. NREL recorded an 8-hour ozone concentration of 92 ppb on July 18. Rocky Flats North recorded a high 8-hour ozone concentration of 92 ppb on July 19. On July 19, seven monitors had monitored concentrations over 84 ppb including Highlands Ranch (86 ppb), South Boulder County (86 ppb), Chatfield (89 ppb), Rocky Flats North (92 ppb), NREL (91 ppb), and Rocky Mountain National Park (92 ppb). Two monitors, Carriage (83 ppb) and Arvada (84 ppb) had 8-hour ozone concentration greater than 80 ppb but less than 85 ppb.

This period had nine days of temperatures greater than or equal to 90 degrees F. from July 12 through July 20th. On the last day of the episode (July 21) the temperature made it up to 85 degrees. Dryness, subsidence, and stable conditions predominated the episode. An upper level ridge was centered over Colorado. This ridge was nearly stationary for several days. Despite southeast surface flow along the Front-Range, dew point levels were low enough to inhibit thunderstorm activity. There was some thunderstorm activity in the mountains, though.

On Thursday, July 18th, there was a slight increase in mid-level moisture during the day. There was too much stability in the atmosphere for thunderstorm development despite the increase in moisture. The strength of the upper ridge peaked on Friday, January 19. The peak strength of the upper ridge coincided with the highest area wide ozone concentrations. Eastern Colorado appeared to be in a dry subsident hole as subtropical moisture extended from Mexico north into Utah and southern Canada. On Saturday, July 20, a quick moving Canadian/Pacific short wave pushed through Montana and North Dakota. The result of this short wave weakened the northern section of the upper ridge. Subtropical moisture migrated over eastern Colorado. Northeastern Colorado began to see stronger diurnal east to northwest surface flow late on Saturday. The diurnal surface flow was enhanced by rising surface heights overnight.

The ozone episode essentially ended on Sunday, July 21. A short wave crossed the northeastern Colorado plains during the early morning hours with some rain shower activity. The effect of this short wave was to suppress afternoon convective activity. A second, weaker short wave crossed over northeastern Colorado during the afternoon hours. Some shower activity in Larimer and Weld Counties resulted from the passage of the second short wave. For the most part, cooler temperatures resulting from a moist and cooler northeasterly flow suppressed convective activity. Winds aloft were also weak during the day, and, for most of the episode as well.

Backward trajectories were computed for July 17 through July 21, 2002. These composite backward trajectory analyses indicated that at lower levels up to 100 meters, the origin of the air mass was from the south and east during the early days of the episode and then from the north during the late part of the episode. Upper layers of flow were from the northwest and may have originated from Salt Lake City but this may be misleading as the 36-hour back trajectories were from the south. Thirty-six hour trajectories for each day indicated that the Northwesterly flow might be an artifact of the long period the trajectory analysis was ran. The mid-level air mass was mixed down to ground level by the time it reached the Denver area. The flow from the various layers (surface, 100m, and 800m) were generally from the south. Flow at 800 m was very light and did not mix down to the ground. Even at 100m the flow did not mix down to the ground either.

On July 19, the flow at all levels were again from the south. Winds speeds were less than the previous day and apparently the flow had a tendency to go around the Palmer Divide. The 800m winds did not mix down to ground level but flowed over the Palmer Divide. On July 20, when NREL had its highest reading over the episode, the ground level flow was very light from the west. There is some indication that flow in the lower levels circulated along the front range. The flow on this day very likely brought in smoke from the Big Elk Fire that was burning near Estes Park. At the 800m level, the flow was from the south over the Palmer Divide. The general flow shifted on the last day of the episode. Winds at the surface were from the north. At the 800m level, winds were from the northwest.

3.1.2 24 June- 2 July 2002

This episode was lengthy when compared to the other episodes. There were seven days in a row where at least one monitor exceeded 80 ppb. This episode had the highest 8-hour average ozone concentration recorded at Rocky Mountain National Park of 93 ppb recorded on June 30. Three days had at least one monitored concentration that exceeded 85 ppb. On June 29, Rocky Flats North recoded an 8-hour average concentration of 89 ppb. However the rest of the monitors in the network had values less than 80 ppb on this date. On June 30, Rocky Mountain National Park recorded a 93 ppb and Rocky Flats north recorded a value of 88 ppb. Both NREL and South Boulder County had 8-hour ozone concentrations of 80 ppb.

The highest ozone concentration occurred in southwest Denver on July 1 where Chatfield recorded a value of 94 ppb. This is the highest ozone concentration recorded over the entire network during the 1999-2002 periods. Highlands Ranch also exceeded the 8-hour ozone concentration at 86 ppb. Values greater than 85 ppb were also recorded at Rocky Flats North (88 ppb), NREL (91 ppb), and Rocky Mountain National Park (85 ppb). A value of 82 ppb was recorded at the Weld County Tower. It should be noted that several large wildfires were burning during this period including the Rodeo Fire in Arizona, the Mission Ridge Fire near Durango, the Hayman Fire near Denver, and other fire complexes in western Colorado. Flow during the later parts of this episode, as indicated by the trajectory plots, blew from one or more of these large fires. (Figures 3-1 through 3-3 give examples of some of these fire complexes).

A stretch of 13 consecutive days of 90 degree F or more occurred from June 21 through July 3. The maximum temperature exceeded 95 degrees F on June 26 (96^oF), June 29 (97^oF), and July 1 (99^oF). On June 25, a warm upper ridge dominated the southwest United States including Colorado through the period. Mid-level winds were weak north to easterly (upslope) to about 700 mb. Surface dew points were fairly moist at 40 to 50 degrees F. Winds aloft were weak and convective storm motion was slow. The upper level ridge remained intact along the Rockies from Mexico to southern Canada on June 26. Winds from the surface to 600 mb were light and from the east. The eastern plains had a fairly moist air mass (50 degree F dew points) but the stable atmosphere prevented much in the way of thunderstorms on the plains. Cooler air had advected into the 700 to 500 mb levels. Surface winds to 700 mb were more northerly and a bit stronger than the day before. The upper level ridge was slightly weaker than the day before and more disorganized but little movement was detected. Winds aloft were weak with slow moving convective storms.

The high-pressure ridge was again in control of the state on June 27. Surface southeast flow on the plains provided for slightly drier air. Convective storms that developed in the mountains died off quickly over the drier and capped air mass over the plains. Friday, June 28 continued the same weather pattern. The air mass was dry and capped over the eastern plains. Any convective storms that developed over the mountains, quickly dissipated over the eastern plains except for a few very slow moving storms. Heavy rain occurred in some areas because of the slow moving storms. The Platteville profiler indicated light and variable winds from the surface on up.

A convergence zone formed from southeastern Douglas County through eastern Adams County on the afternoon of Saturday, June 29. The convergence zone separated very dry air coming off of the foothills from moist (45-55 degree F) dew points to the south and east. Flow aloft was stronger and more organized than on previous days. The flow aloft was also more from the west and northwest than on the previous days. Despite a dry cold front sliding southward through eastern Wyoming, overnight temperatures did not fall much below 70 degrees F until the early morning hours.

Sunday, June 30 and Monday, July 1 had the highest ozone concentrations over the episode. On Sunday, the air mass over northeastern Colorado was very dry and stable following the cold front passage. Subsidence from the already warm and dry air mass pushed temperatures near the century mark over much of eastern Colorado. A mid-level inversion prevented any thunderstorms from building on the eastern plains. Moderate levels of smoke from several fires burning in the west (Hayman, Missionary Ridge, Rodeo in Arizona, and Million Fire) were reported along the northern Front Range. Monday, July 1 was more of the same. Strong mid-level subsidence over the northeastern plains continued to dominate the local weather pattern. High ozone readings were widespread over the network. Ozone levels decreased on July 2 and 3 with temperature continuing over 90 degrees F. Gulf moisture moved into the area across the mountains and foothills. A weak cap around 500 mb was still present over the area. Surface winds shifted to the northeast. No real strong indications why the ozone episode did not continue on July 2 and 3.

This episode was strongly influenced by flow from the southwestern United States including southern California and Arizona. Subsidence over Colorado mixes surface and 800m layers down to the surface by the time they reach the Front Range. The flow from June 25 through June 28 was generally from the south. Winds during this period were light at all levels, especially on June 25. During this period, ozone concentrations were the lowest during the episode. On June 28, winds became stronger from the south. Upper levels winds at 800m started to shift from the southwest. On June 29 through July 1, winds at 800m were from the southwest, originating in Arizona and Utah. Surface winds were light during this period originating from the west and southwest.

3.1.3 8-12 June 2002[1]

This period occurred just three days after the start of the Hayman fire. Very warm temperatures along with smoky conditions characterized this episode. Concentrations of 88 ppb occurred on two days, June 8 and 9, at Rocky Flats North. A value of 88 ppb was recorded at NREL on June 9 as well. A value of 83 ppb occurred at Rocky Mountain National Park on June 11 and at Rocky Flats North on June 12. During this episode, other monitors in the network were all below 80 ppb indicating that this episode was not widespread.

June 8 and 9, when the highest ozone readings occurred, the maximum temperature reached 96 degrees F and 95 degrees F, respectively. Despite cooler temperature on June 11 and 12, ozone readings above 80 ppb were monitored on these days. The maximum temperatures recorded on June 11 and 12 were

---

[1] The MM5 modeling results for the third episode (8-12 June 2002) were not completed as of this writing. They will be presented once the modeling and evaluation is finished. For completeness, however, we present a summary of the conditions of this episode here.

78 degrees F and 82 degrees F, respectively. Ozone readings were below 80 ppb on June 10 when the maximum-recorded temperature was 75 degrees F. On June 8, shallow moist air covered most of eastern Colorado during the morning hours. This moist air mass mixed out as the day progressed. Very little convective activity occurred over the mountains and northeastern plains. Smoke from the Hayman fire was observed over the Denver area on June 9. A weak short wave passed north of the area during the evening hours. As a result of the short wave passage, winds aloft shifted to a westerly direction. The inversion layer lowered to about 2000 feet overnight.

Much colder air moved into the area on June 10 with the maximum temperature in Denver recorded at 75 degrees F. Consequently, no ozone readings exceeded 80 ppb. Winds had shifted to the southeast for most of the day. An inversion persisted for most of the day on June 11. The height of the inversion was around 18 thousand feet. Although ozone readings were low network wide, Rocky Mountain National Park had a reading of 83 ppb. An ozone reading of 83 ppb was recorded at Rocky Flats north on June 12. Except for smoke in the area from the Hayman Fire, very little else can be said about this day. Warm temperatures over the mountains with cooler temperatures over the plains were indicative of a persistent inversion over the area.

The composite trajectory analyses revealed that the air parcels originated in very different areas at the three levels over the 120-hour simulation. At the surface, the flow was from the south from Texas, at 100m the flow was from the southwest from Arizona, and at 800m the flow was from the northwest from Salt Lake City. The 36- hour plots indicated the flow was from the southwest from Arizona on June 8th through June 10th at all levels. The southwesterly flow was fairly strong originating in Arizona and southern California at the start of each 36-hour period. On June 11 and 12, the flow became more westerly at 800m. The surface flow was from the Nebraska panhandle on June 11. The northeasterly flow was much less on this date. By June 12 the surface flow had shifted to the southwest with fairly light wind speeds.

3.2 Summary of the Conceptual Model of 8-hr Ozone Episodes for the DNFRR

The CDPHE developed a succinct conceptual model of 8-Hour Ozone formation in the Denver Northern Front Range Region (CDPHE, 2003). Salient features of the model are as follows. High ozone concentrations generally occur in the Denver region on days that are hot, cloud-free, and with stagnant to light wind speeds at both at the surface and aloft. Most high-ozone events occur on days when high temperatures are above 90 degrees F and when light, up-slope winds occur at the surface and mountaintop level. Episodic events of ozone occur when maximum daily temperatures above 90 degrees F persist for several days in a row. On most high ozone days, dew points on the eastern plains are in the 40-60 ^{O}F range. Relatively high dew point levels are probably necessary for efficient photochemistry production and differentiate those days that are above 90 ^{O}F with high ozone levels, and, dry hot days with lower ozone levels. The absence of cloud cover and thunderstorms promotes ozone formation. Conversely, typical late-afternoon thunderstorms and associated cloud cover retard the formation of ozone and help keep ozone concentrations at levels below the federal standard. Timing of thunderstorms off of the mountains in the late afternoon and evening hours during hot days is another critical piece in determining whether the 8-hour ozone standard is exceeded on a day-to-day basis in the Denver area. The highest ozone levels usually occur in June and July and sometimes-early August.

3.3 Summer '02 Episode and Intensive Study Periods

As set forth in the protocol, it was agreed that the Denver 8-hr ozone study would focus on the three 2002 episodes as a single MM5/CAMx regional simulation. This the 'Summer '02' episode runs from 5 June to 23 July 2002. Meteorological inputs from the MM5 model were produced on the 36/12 km grid for the period beginning 1200 UTC (0500 MST) on 5 June 2002 through 1200 UTC (0500 MST) on 23 July 2002. The higher resolution MM5 simulations will be active during the following periods:

- 1200 UTC (0500 MST) on 16 July through 1200 UTC (0500 MST) on 23 July;
- 1200 UTC (0500 MST) on 23 June through 1200 UTC (0500 MST) on 3 July; and
- 1200 UTC (0500 MST) on 5 June through 1200 UTC (0500 MST) on 14 June.

The CAMx model will be run at 12km resolution from 0000 MST on 7 June through midnight on 22 July 2002. The higher resolution 4 km and 1.33 km CAMx grid nests will be run from:

- 0000 MST on 8 June through 2400 MST on 12 June (i.e., 8-12 June);
- 0000 MST on 25 June through 2400 MST on 1 July (i.e., 25 June-1 July); and
- 0000 MST on 18 July through 2400 MST on 21 July 2002 (i.e., 18-21 July).

These latter two intensive study periods are the top two ranking episodes in the CDPHE's ozone episode selection scheme. As indicated previously, the 8-12 June episode is potentially confounded by the presence of the Hayman fire complex southwest of Denver. Accordingly, should it be determined that the uncertainties in the modeling of the 8-12 June 2002 episode are unacceptability large owing to the wildfires (thereby precluding reliable source-receptor modeling), this intensive period may not be used to assess compliance with the 8-hr ozone standard.

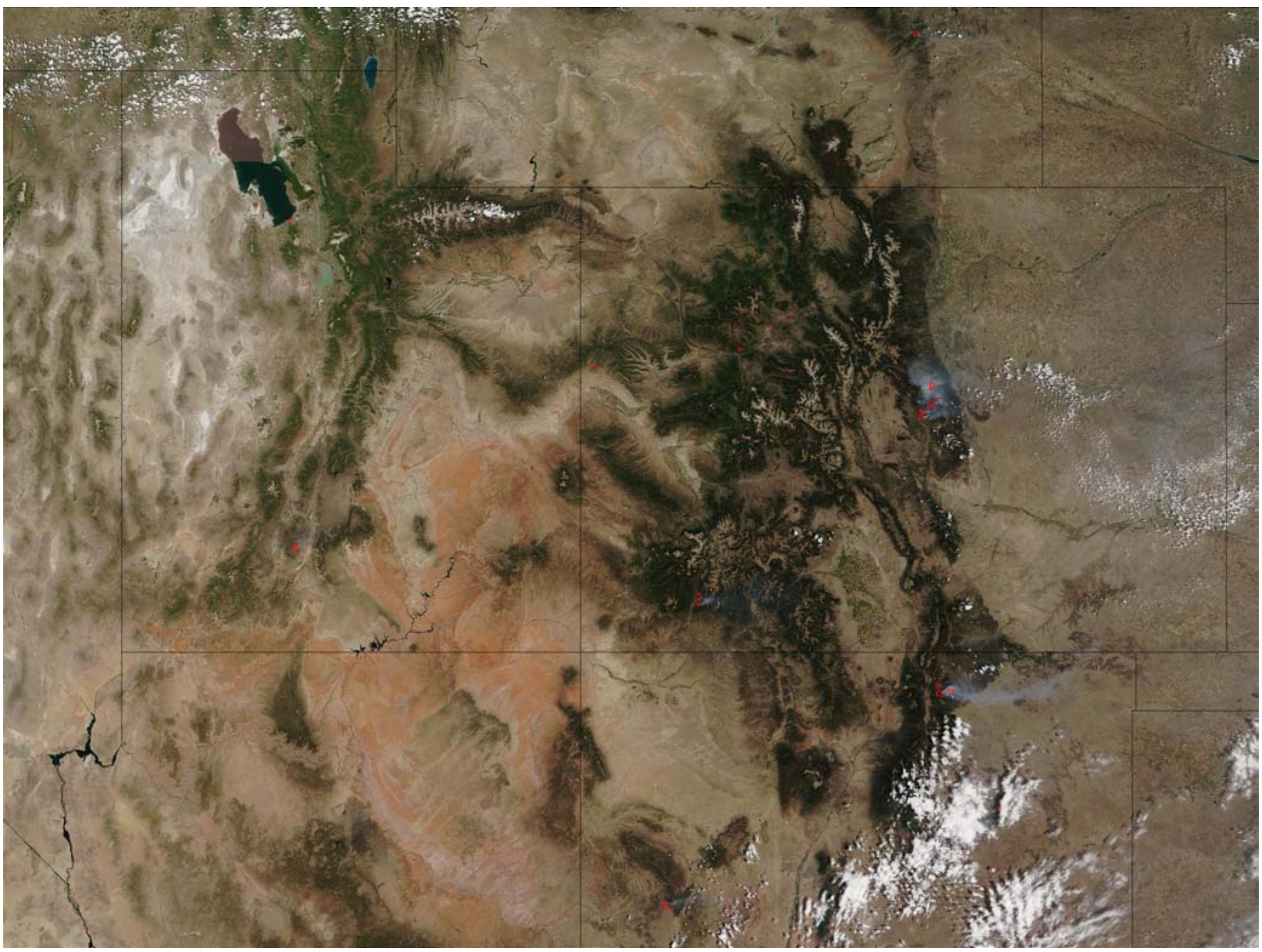


Figure 3-1. Satellite Imagery of Fires in Colorado and New Mexico on Day 163 (12 June 2002).

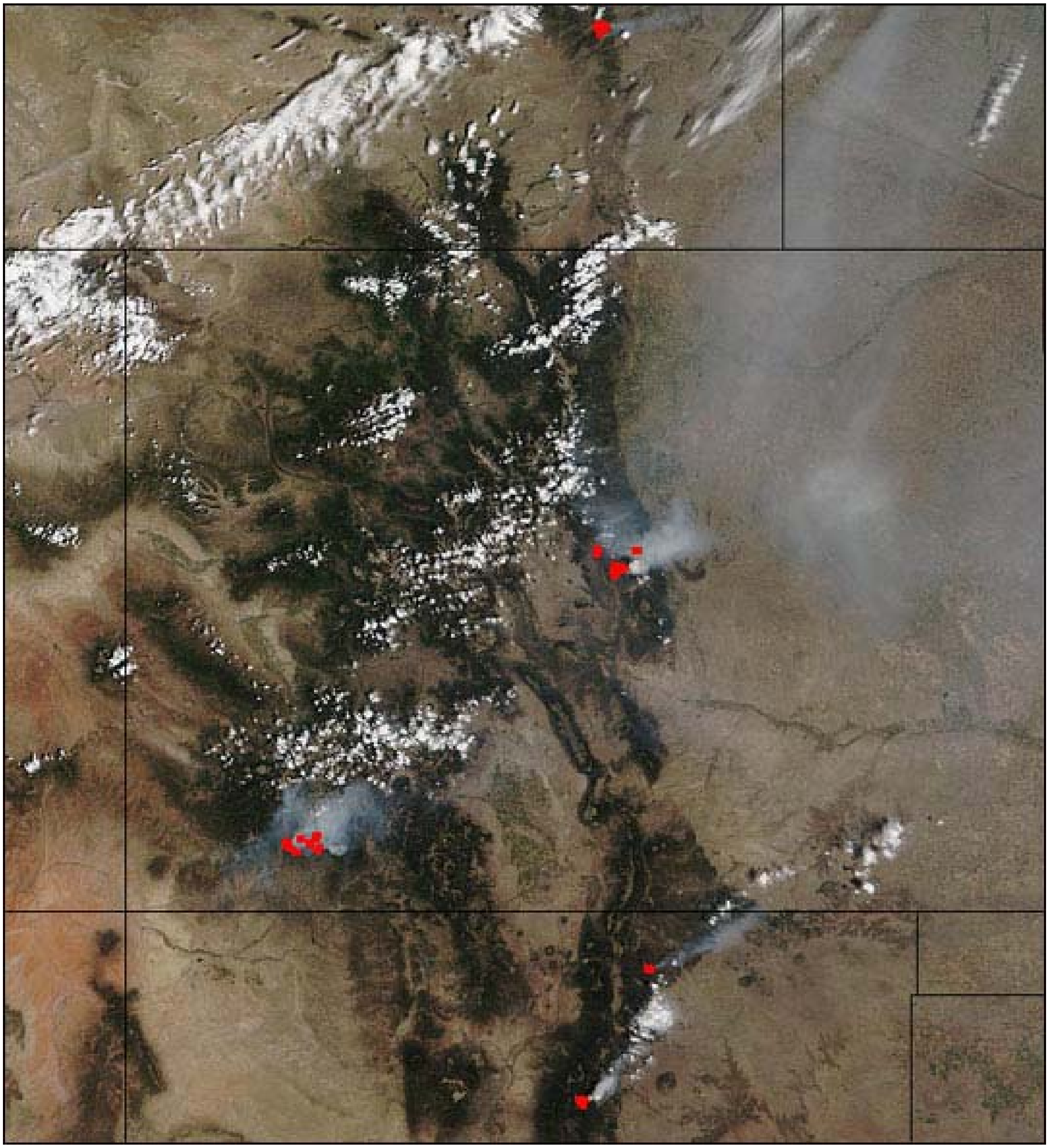


Figure 3-2. Satellite Imagery of Fires in Colorado and New Mexico on Day 170 (19 June 2002).

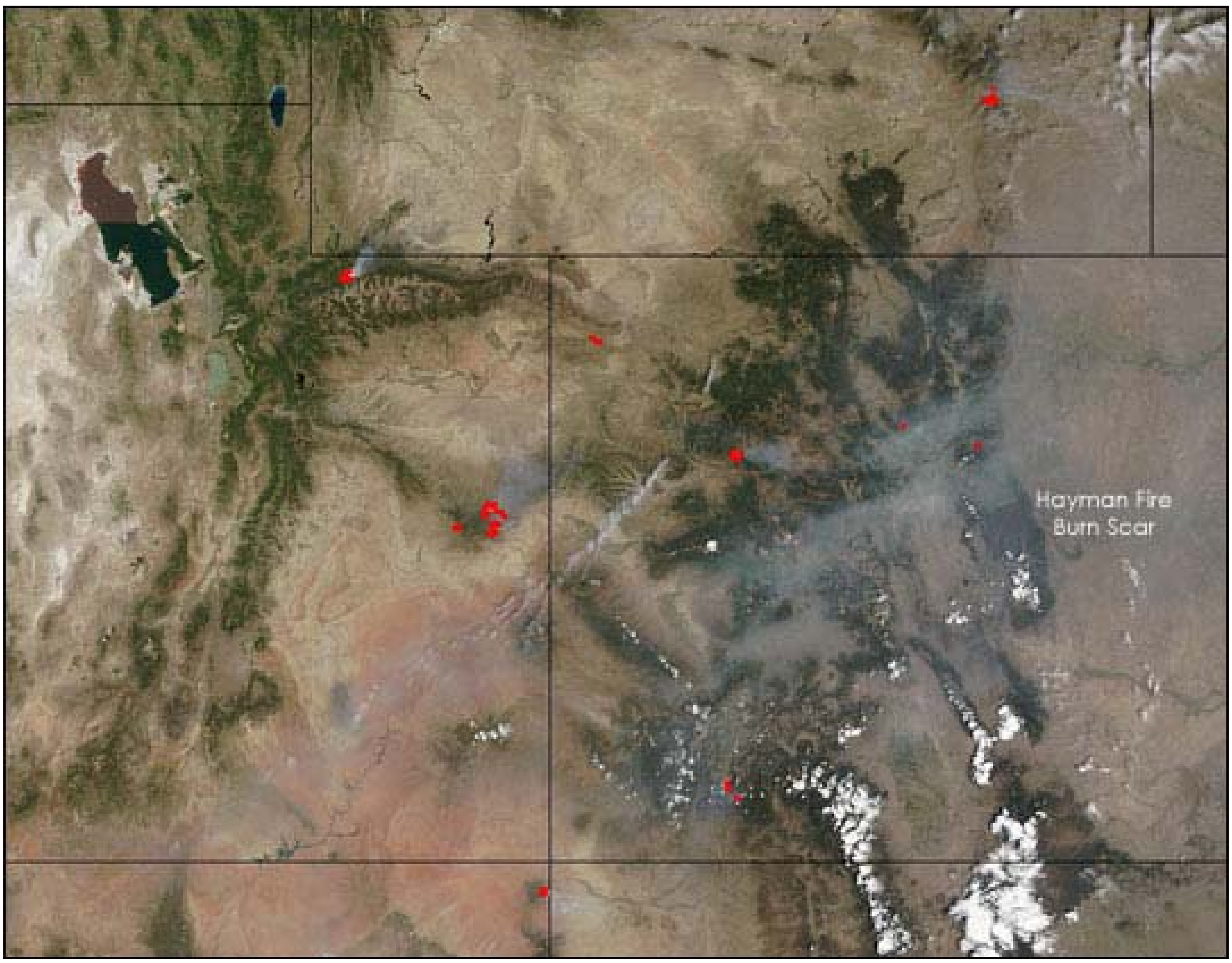


Figure 3-3. Satellite Imagery of Fires in Colorado and New Mexico on Day 181 (30 June 2002).

4.0 MODELING DOMAINS AND DATA AVAILABILITY

This section identifies the MM5 meteorological modeling domain and grid specifications used in the Denver 8-hr Early Action Compact Study. The meteorological domains (Figure 4-1) are consistent with EPA 8-hr ozone modeling guidance. Figure 4-1 through 4-3 and Tables 4-1 and 4-2 present the spatial definitions of the meteorological modeling domains. Below, we discuss the rationale underlying these selections.

4.1 MM5 Meteorological Modeling Domain

Figures 4-1 and 4-2 present the nested MM5 domains at four levels of nesting: 36/12/4/1.33 km horizontal resolutions. In this DNFRR application, the 1.33 km 'Hi-Res' domain (Figure 4-2b) is located over the Denver-Boulder metropolitan area. For the MM5 modeling, the outer 36 km grid domain covers the entire continental U.S. and large portions of Canada and Mexico. This region is consistent with the recent continental scale, annual MM5 modeling Alpine has performed for the U.S. EPA and EPA Region 8 (see, for example, McNally and Tesche, 2002; 2003).

By using four nested grids at these resolutions (3:1 ratios), the needs of synoptic-scale accuracy, fine resolution, and consistency with the requirements of regional photochemical models is achieved. In addition, the meteorological modeling domain is configured so that: (1) the MM5 grids will align properly with the CAMx air quality grids, with some overlap; (2) additional 4-km and 1.33 km fine nests will be established to cover the Denver-Northern Front Range focus area, and (3) the MM5 LCP grid will be defined to be centered over the intermountain west domain. The horizontal resolution of the four MM5 nests are listed in Table 4-1. In the vertical, thirty-four (34) layers for MM5 will be used (see Table 4-2). Based on our previous modeling of this region with the MM5/CAMx (Tesche and McNally, 2003a,b) we expect this grid layering scheme to offer sufficient vertical resolution over the study region.

4.2 Data Availability

The predominant types of meteorological data to be used in this study will be surface and upper air meteorological measurements reported by the National Weather Service (NWS), and large-scale (i.e., regional/global) analysis databases developed by the National Center for Environmental Prediction (NCEP). Both types of data are archived by, and currently available from, the National Center for Atmospheric Research (NCAR). Measurement data include surface and aloft wind speed, wind direction, temperature, moisture, and pressure. Hourly surface data are usually available from many Class I airports, i.e., larger-volume civil and military airports operating 24-hour per day. The standard set of upper air data are provided by rawinsonde soundings launched every 12 hours from numerous sites across the continent. The typical spacing of rawinsonde site is approximately 300 km.

ETA analysis databases include 3-hourly 40 km. resolution analysis fields of winds, temperature, moisture, and pressure. The analysis data will be combined as necessary with measurement data for the following purposes:

> Developing initial and boundary inputs to MM5;

> Developing nudging fields for the MM5 FDDA package; and

> Evaluating MM5 predictive performance over the central U.S., with particular focus on the central Colorado region.

Figure 4-3 shows the locations of the NWS upper air meteorological sounding data available from NCAR archives. Although this figure doesn't depict all of the currently available upper air sites, it does give a good overview of the number of sounding locations and their spatial distribution across the eastern U.S.

The MM5 requires inputs of gridded terrain elevation and landuse/landcover codes for each grid specified in a simulation. NCAR provides access to several global and continental-scale terrain elevation and landcover databases of varying resolution. For example, the 36-km grid will use 10 minute topographic information derived from the Geophysical Data Center global data set. The 12-km grid will use the 5 min (~9.25 km) Geophysical Data Center global data set. Even finer resolution databases are available from NCAR for limited areas of North America; these would be used for the finest 4 km and 1.33 km grid nests.

Table 4-1. Grid Definitions for the Denver EAC 8-hr Ozone Modeling Study.

Model	Grid Cells East-West	Grid Cells North-South	Lambert Grid Origin (km)From Pole (-93,40)
EPS2x/CAMx			
- 36 km Grid	74	56	-2304, -1404
- 12 km Grid	107	107	-1560, -912
- 4 km Grid	146	122	-1076, -292
- 1.33 km Grid	128	128	-733.3, -73.3
MM5			
- 36 km Grid	165	129	-2952, -2304
- 12 km Grid	127	127	-1656, -1008
- 4 km Grid	163	151	-1116, -372
- 1.33 km Grid	151	151	-748.6, -88.6

Table 4-2. MM5 Vertical Grid Structure.

k(MM5)	sigma	press.(mb)	height(m)	depth(m)
34	0.000	10000	15674	2004
33	0.050	14500	13670	1585
32	0.100	19000	12085	1321
31	0.150	23500	10764	1139
30	0.200	28000	9625	1004
29	0.250	32500	8621	900
28	0.300	37000	7720	817
27	0.350	41500	6903	750
26	0.400	46000	6153	693
25	0.450	50500	5461	645
24	0.500	55000	4816	604
23	0.550	59500	4212	568
22	0.600	64000	3644	536
21	0.650	68500	3108	508
20	0.700	73000	2600	388
19	0.740	76600	2212	282
18	0.770	79300	1930	274
17	0.800	82000	1657	178
16	0.820	83800	1478	175
15	0.840	85600	1303	172
14	0.860	87400	1130	169
13	0.880	89200	961	167
12	0.900	91000	794	82
11	0.910	91900	712	82
10	0.920	92800	631	81
9	0.930	93700	550	80
8	0.940	94600	469	80
7	0.950	95500	389	79
6	0.960	96400	310	78
5	0.970	97300	232	78
4	0.980	98200	154	39
3	0.985	98650	115	39
2	0.990	99100	77	38
1	0.995	99550	38	38
0	1.000	100000	0	0

Table 4-3. Comparison of MM5 and CAMx Vertical Grid Structures.

MM5 Layer K	Interface Heights Height (m)	CAMx Layer Interface Heights
28	6521	21
25	4660	20
22	3132	19
19	1911	18
17	1434	17
16	1280	16
15	1129	15
14	981	14
13	834	13
12	690	12
11	619	11
10	548	10
9	478	9
8	409	8
7	340	7
6	271	6
5	203	5
4	135	4
3	102	3
2	68	2
1	35	1

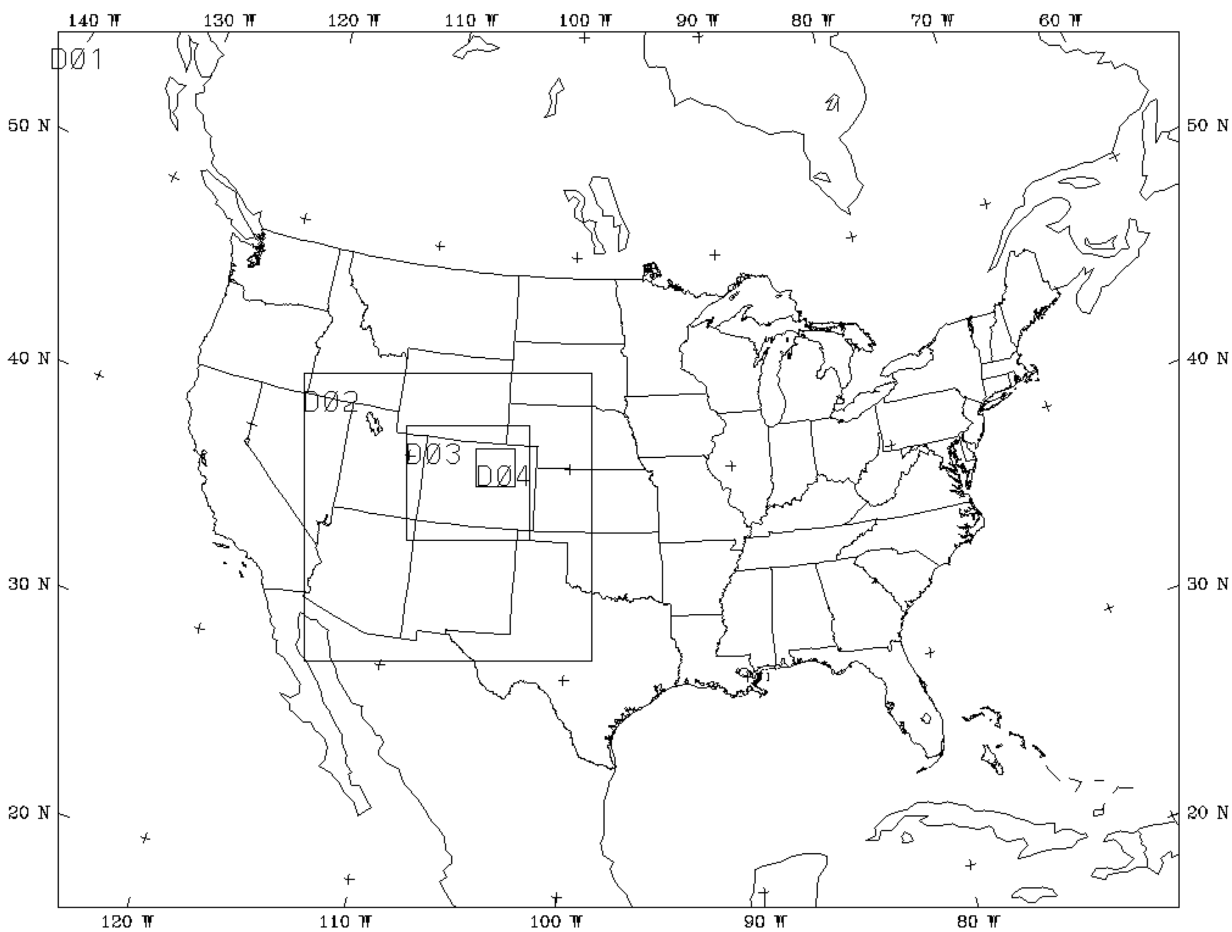
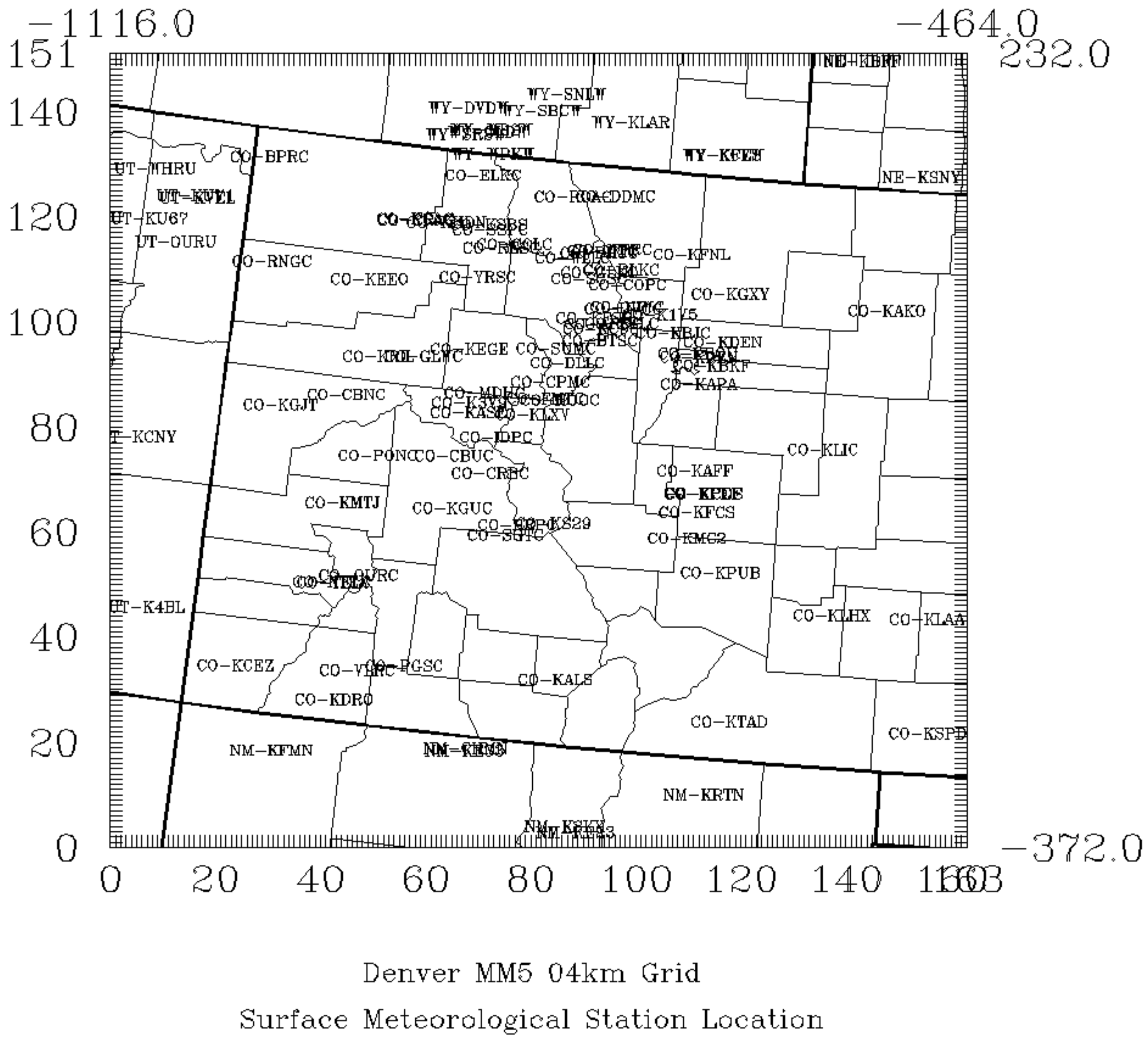
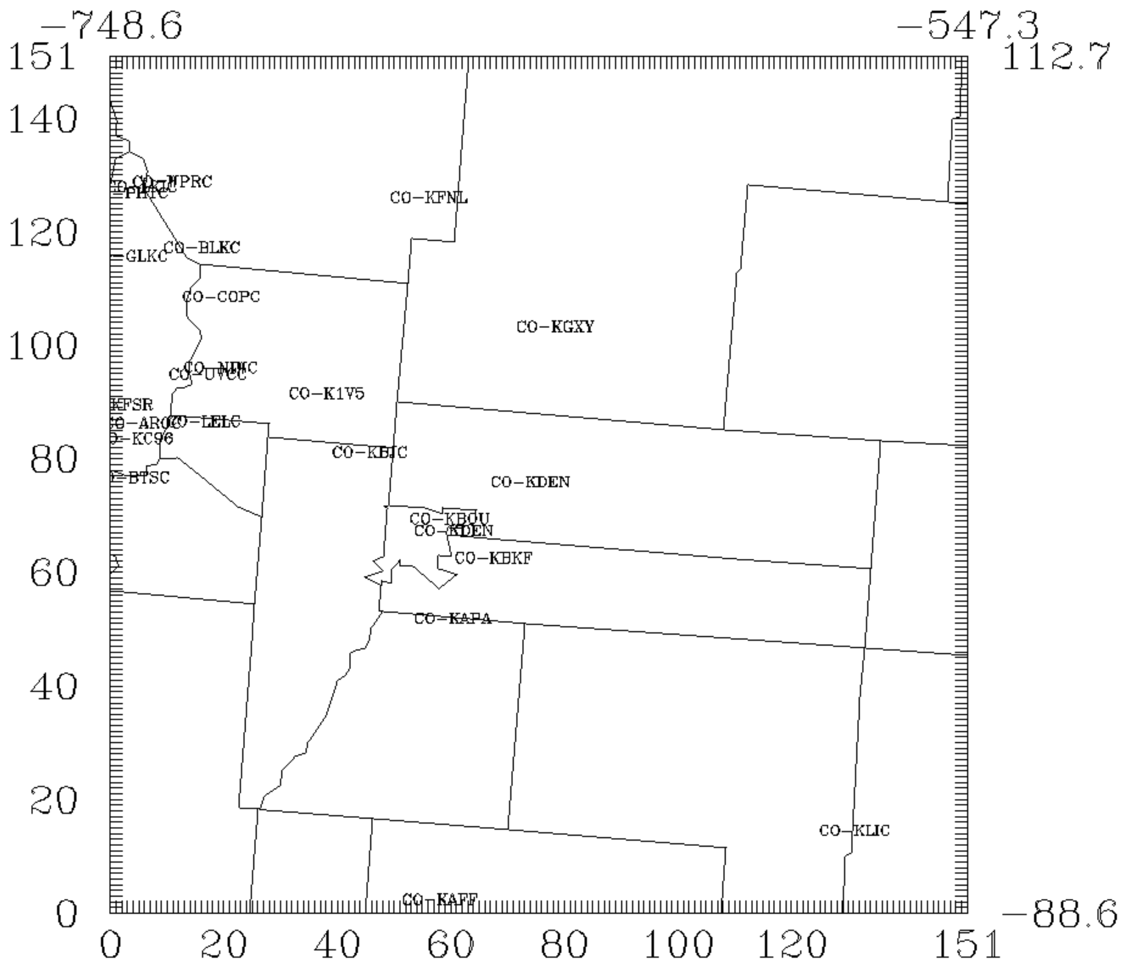


Figure 4-1. MM5 Nested 36/12/4/1.33-km Meteorological Modeling Domain for the Denver EAC 8-Hr Ozone study.



(a) 4 km Grid Domain

Figure 4-2. Location of Nested MM5 Grids and Air Quality Monitoring Stations for the Denver EAC 8-Hr Ozone Study.



Denver MM5 1.3 km Grid
Surface Meteorological Station Location

(b) 1.33 km Grid Domain

Figure 4-2. Continued.

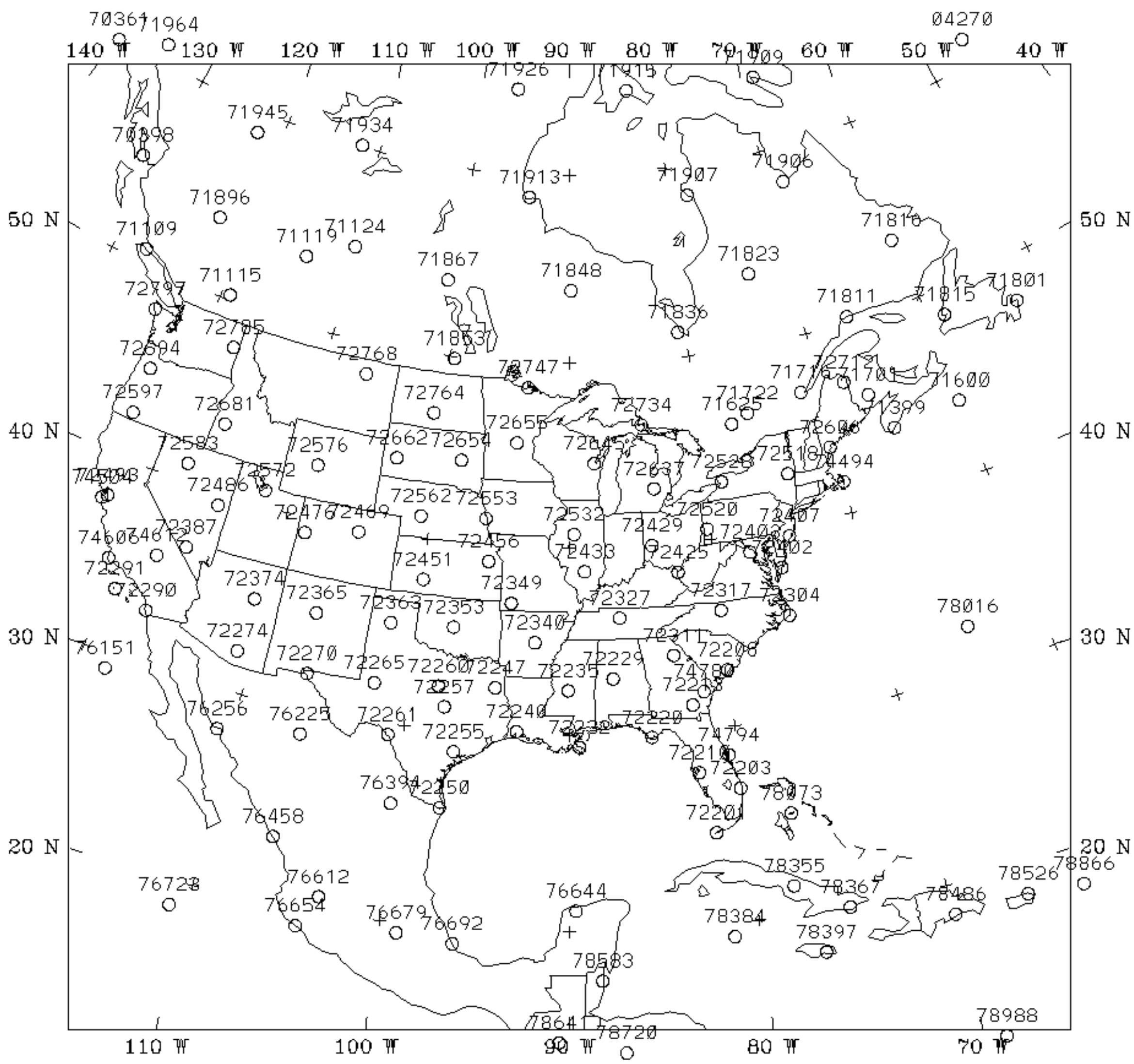


Figure 4-3. Location of Upper Air Sounding Sites Throughout the U.S. to be Used in the MM5 Prognostic Meteorological Modeling for the Denver EAC 8-hr Ozone Study.

5.0 INPUT DATA PREPARATION PROCEDURES

The modeling methodology used with the MM5 model for the three Denver EAC 8-hr ozone episodes is described in the protocol (Tesche et al., 2003a). In this section we summarize the input preparation procedures and the specific model features and/or algorithms used to produce the meteorological fields employed in the present study. Additional details describing the model algorithms and modeling procedures are contained in the literature citations.

5.1 Fixed Inputs

Topography Topographic information was developed using the National Center for Atmospheric Research (NCAR) terrain databases. The 36 km and 12km grids used 5 min topographic information derived from the Geophysical Data Center global data set while the 4 km and 1.33 km grids used the 30 sec resolution data set. Terrain data were interpolated to the model grids using a Cressman-type objective analysis scheme.

Vegetation Type and Land Use: Vegetation type and land use information was developed using the NCAR/PSU 10 min. (~18.5 km) databases for the 36 km grid and from the United States Geological Survey (USGS) data for the 12 km, 4 km and 1.33 km grids. Surface characteristics corresponding to each land use category in the MM5 modeling domain are consistent with those used in the photochemical model (CAMx) and are discussed in McNally and Tesche (2002, 2003).

5.2 Variable Data Inputs

Atmospheric Data: Initial conditions to the MM5 were developed from operationally analyzed fields derived from the National Center for Environmental Predictions (NCEP) ETA (40 km resolution) following the procedures outlined by Stauffer and Seaman (1990). The synoptic-scale data used in the initialization (and in the analysis nudging discussed below) were obtained from the conventional National Weather Service (NWS) twice-daily radiosondes and standard 3-hr NWS surface observations. These data included the horizontal wind components (u and v), temperature (T), and relative humidity (RH) at the standard pressure levels, plus sea-level pressure (SLP) and ground temperature (T_g). Here, T_g represents surface temperature over land and sea-surface temperature over water.

The so-called "first guess" NMC-analyzed fields were interpolated to several supplemental analysis levels (e.g., 950, 900, 800, and 600 mb) and then modified by blending in the NWS standard rawinsonde data using a successive-correlation type of objective analysis that accounts for enhanced along-wind correlation of variables in strongly curved flow (Benjamin and Seaman, 1985). Subsequently, the three-dimensional variable fields were interpolated onto the MM5's sigma vertical coordinate system. On the 36 km grid (Grid D01), the analyses were performed using a Cressman-type procedure and then interpolated to the 12 km, 4 km and 1.33 km grids (Grids D02, D03, and D04) shown previously in Figure 4-1.

Lateral boundary conditions to the MM5 were specified from observations by temporally interpolating the 12-hourly enhanced analyses described above. The inner meshes were operated in a two-way interactive mode with the next outer grids and received their boundary conditions at one-hour intervals. For each time step between the times for which new boundary conditions were available, a temporal interpolation was performed to provide smoothly changing boundary values to the appropriate nested meshes.

Water Temperature: Water temperatures were derived from the ETA skin temperature variable. These temperatures were then bi-linearly interpolated to each model domain and, where necessary, filtered to smooth out irregularities.

Clouds and Precipitation: While the non-hydrostatic MM5 treats cloud formation and precipitation directly through explicit resolved-scale and parameterized sub-grid scale processes, the model does not require precipitation or cloud input. The potential for precipitation and cloud formation enters through the thermodynamic and cloud processes formulations in the model. The only precipitation-related input required was the initial mixing ratio field that was developed from the NWS and NMC data sets previously discussed.

5.3 Multi-Scale FDDA

The multi-scale Four Dimensional Data Assimilation (FDDA) technique developed at Penn State (Stauffer and Seaman, 1990, 1994; Stauffer et al., 1985, 1991) is based on Newtonian relaxation, or nudging, which is a continuous assimilation method that relaxes the model state toward the observed state by adding to one or more of the prognostic equations artificial tendency terms based on the difference between the two states. It is basically a form of continuous data assimilation because the nudging term is applied at every time step, thereby minimizing "shock" to the model solutions that may occur in intermittent assimilation schemes. The standard FDDA methodology includes two options: (a) nudging toward gridded analyses which are interpolated to the model's current time step, and (b) nudging directly toward individual observations within a time-and-space "window" surrounding the data. These two approaches are referred to as "analysis nudging" and "obs-nudging", respectively. Analysis nudging is ideal for assimilating synoptic data that cover most or all of a model domain at discrete times. Obs-nudging does not require gridded analyses of observations and is better suited for assimilating high-frequency asynoptic data that may be distributed non-uniformly in space and time (i.e., the Lake Michigan Ozone Study intensive studies data).

A “multi-scale” data assimilation strategy was used with MM5 in this study. This methodology, developed by researchers at Penn State University (Shafran and Seaman, 1998) employs both FDDA methods. Standard “analysis nudging” was used on the outer grids using objectively analyzed three-dimensional fields produced every 3-hr from the NWS rawinsonde wind, temperature, and mixing ratio data, and similar analyses generated every three hours from the available NWS surface data. More specifically, analysis nudging was only used on the outer two grids (i.e., 36 km and 12 km) and the size of the nudging coefficient used for the assimilation of wind, temperature and moisture was 2.5×10^{-4} for winds and temperature and 1.0×10^{-4} for mixing ratio. These are modest nudging coefficients and certainly do not ‘drive’ the MM5 fields to the observations. It is more a gentle steering of the fields aimed at avoiding error growth.

5.4 Physics Options

The MM5 model physics options used in the Denver EAC application were as follows.

Planetary Boundary Layer Schemes. For the 12 and 36 km grids, the Pleim-Chang planetary boundary layer scheme was used. This PBL scheme is a derivative of the Blackadar PBL scheme called the Asymetric Convective Model using a variation on Blackadar's non-local mixings. For the 4 and 1.3 km grids, the Blackadar planetary boundary layer scheme was used.

Explicit Moisture Schemes. Resolved-scale precipitation processes were treated explicitly with a simple water/ice scheme (no supercooled water substance) following the approach of Dudhia (1989). For the 36 km and 12 km mesoscale grids, the Kain-Fritsch scheme was used. This parameterization achieves closure via convective available potential energy and an entraining/detraining cloud model. Furthermore, it parameterizes moist convective downdrafts. No convective parameterization was performed on the 4 km and 1.3km meshes since we assumed that convection is explicitly resolved at this scale.

Radiation Scheme. The Rapid Radiative Transfer Model (RRTM) longwave scheme was used. The RRTM is a new highly accurate and efficient method.

Land Surface Model. For the 12 and 36 km grids, the Pleim-Xiu (PX) land surface model was used. This scheme represents soil moisture and temperature in two layers (surface layer at 1cm and root zone at 1m) as well as canopy moisture. It handles soil surface, canopy and evapotranspiration moisture fluxes. The PX scheme was run in a continuous mode throughout the entire episode. For the 4 and 1.3km grids, no land surface model was employed. To account for the extreme drought in the Western United States during the summer of 2002, the soil moisture availability for all non-water/ice land uses was reduced by approximately 50%.

Grid Nesting: The 36km and 12km domains were run with continuous updating without feedback from the finer grid to the coarser grid. The 4km domain was run with hourly updating from the 12km domain. The 1.3km domain was run with continuous updating of the 4km domain.

Table 5-1. Description of Land Use Categories and Physical Parameters.

Land Use Integer Identification	Land Use Description	Albedo (%)	Moisture Avail. (%)	Emissivity (% at 9 micrometers)	Roughness Length (cm)	Thermal Inertia (cal cm-2 k-1 s-1/2)
1	Urban Land	18	5	88	50	0.03
2	Agriculture	17	30	92	15	0.04
3	Range-grassland	19	15	92	12	0.03
4	Deciduous Forest	16	30	93	50	0.04
5	Coniferous Forest	12	30	95	50	0.04
6	Mixed Forest and Wet Land	14	35	95	40	0.05
7	Water	8	100	98	0.01	0.06
8	Marsh or Wet Land	14	50	95	20	0.06
9	Desert	25	2	85	10	0.02
10	Tundra	15	50	92	10	0.05
11	Permanent Ice	55	95	95	5	0.05
12	Tropical or SubTropical Forest	12	50	95	50	0.05
13	Savannah	20	15	92	15	0.03

6.0 QUALITY ASSURANCE

The MM5 meteorological model inputs and outputs were plotted and examined to ensure: (a) accurate representation of the observed data in the model-ready fields, and (b) temporal and spatial consistency and reasonableness. As discussed in the next chapter, the MM5 was subjected to an operational/scientific evaluation and this facilitated, among other things, the quality assurance review of the meteorological modeling procedures. Data sets available to support this quality assurance of the aerometric inputs included the routine synoptic-scale data sets from the NWS 12-hourly rawinsondes and 3-hourly surface observations. These data include the horizontal wind components (u and v), temperature (T), and relative humidity (RH) at the standard pressure levels, plus sea-level pressure (SLP) and ground temperature (T_g). i.e., the surface temperature over land and sea-surface temperature over water.

7.0 MODEL PERFORMANCE EVALUATION

Model performance evaluation (MPE) is the process of testing a model's ability to estimate accurately observed atmospheric properties over a range of synoptic and geophysical conditions. When conducted thoughtfully and thoroughly, the process focuses and directs the continuing cycle of model development, data collection, model testing, diagnostic analysis, refinement, and re-testing. In this section we summarize the philosophy and objectives that governed the evaluation of the MM5 prognostic model for the DNFRR application. We then identify the specific evaluation methods that were employed to judge the suitability of the MM5 for input to CAMx in the present EAC regulatory application. In Appendix A, we summarize the statistical measures and graphical procedures used to elucidate MM5 model performance. This evaluation plan conformed to the procedures recommended by the EPA (1991, 1999) for meteorological modeling in support of 1-hr and 8-hr ozone attainment demonstrations.

7.1 Principles

We begin by establishing a framework for assessing whether the MM5 modeling system (i.e., the core model, data processing schemes, and supporting data sets) performs with sufficient reliability to justify its use with CAMx for developing ozone control strategies. The model's reliability will be assessed given consideration to the following principals:

> **The Model Should be Viewed as a System.** When we refer to evaluating a "model", we mean this in the broad sense. This includes not only the MM5 meteorological, but its various components: companion preprocessor models, the supporting geophysical data base, and other related analytical and numerical procedures used to produce meteorological modeling results. A principal emphasis in the MM5 model testing process is to identify and correct flawed model components;

> **Model Acceptance is a Continuing Process of Non-Rejection.** Over-reliance on explicit or implied model "acceptance" criteria should be avoided. This includes the emergent 'benchmarks' used as performance goals for meteorological modeling (Emery et al., 2002; Tesche et al., 2003c,d). Models should be accepted gradually as a consequence of successive non-rejections. Over time, confidence in a model builds as it is exercised in a number of different applications (hopefully involving stressful performance testing) without encountering major or fatal flaws that cause the model to be rejected;

> **Criteria for Judging Model Performance Must Remain Flexible.** The criteria for judging the acceptability of model performance should remain flexible, recognizing the challenging requirement of the Denver EAC application (i.e., use of a nested regional photochemical model, new emissions data sets developed by the CDPHE, and prognostic model simulations over complex terrain down to physical grid scales as fine as 1.33 km); and

> **Previous Experience Used as a Guide.** Previous photochemical modeling experience serves as a primary guide for judging model acceptability. Interpretation of the MM5 modeling results for each episode, against the backdrop of previous modeling experience, will aid in identifying potential performance problems and suggest whether the model should be tested further or rejected.

These principals have been incorporated into the following operational methodology for testing the performance of the MM5 modeling system to support regulatory 8-hr ozone modeling in accordance with EPA (1999, 2002) guidelines.

7.2 Meteorological Model Evaluation Process

Meteorological inputs required by CAMx include hourly estimates of surface pressure and clouds; the three-dimensional distribution of winds, temperatures, and mixing ratio; and other physical parameters or diagnosed quantities such as turbulent mixing rates (i.e., eddy diffusivities) and planetary boundary layer heights. Accordingly, the objective of the MM5 performance evaluation is to assess the adequacy of these surface and aloft meteorological fields. More specifically, we seek to assess the adequacy and reliability of the dynamic and thermodynamic meteorological fields for input to the CAMx regional photochemical model. The MM5 evaluation will be founded upon comparisons between hourly-averaged modeled predictions and surface and aloft meteorological measurements obtained principally from National Weather Service (NWS) sites and at various air monitoring stations.

7.2.1 Components of the MM5 Evaluation

The MM5 modeling system is well-established with a rich development and refinement history spanning more than two decades (Seaman, 2000). The model has seen extensive use worldwide by many agencies, consultants, university scientists and research groups. Thus, the current version of the model, as well as its predecessor versions, has been extensively "peer-reviewed" and considerable algorithm development and module testing has been carried out with all of the important process components. Given that the MM5 model code and algorithms have already undergone significant peer review, performance testing of the MM5 model in this study will be focused on an operational evaluation.

The *operational evaluation* refers to an assessment of a model's ability to estimate atmospheric observations independent of whether the actual process descriptions in the model are accurate (Tesche, 1991a,b). It is an examination of how well the model reproduces the observed meteorological fields in time and space consistent with the input needs of the air quality model. Here, the primary emphasis is on the model's ability to reproduce hourly surface wind speed, wind direction, temperature, and mixing ratio observations across the 12/4/1.33 km grid domains. The operational evaluation provides very useful information but is somewhat limited in revealing whether the results are correct from a scientific perspective or whether they are the fortuitous product of compensating errors.

A "successful" operational evaluation is a necessary but insufficient condition for achieving a sound, reliable performance testing exercise. An additional scientific evaluation is also needed. The *scientific evaluation* attempts to elucidate the realism of the basic meteorological processes simulated by the model. This involves testing the model as an entire system (i.e., not merely focusing on surface wind predictions) as well as its component parts. The scientific evaluation seeks to determine whether the model's behavior in the aggregate and in its component modules is consistent with prevailing theory, knowledge of physical processes, and observations. The main objective is to reveal the presence of bias and internal (compensating) errors in the model that, unless discovered and rectified, or at least quantified, may lead to erroneous or fundamentally incorrect technical or policy decisions. Typically, the scope of the scientific evaluation is limited by the availability of special meteorological observations (radar profiler winds, turbulence measurements, PBL heights, precipitation and radiation measurements, inert tracer diffusion experiments, and so on). Unfortunately, since none of these measurements were available over the Denver region during the summer of 2002, a meaningful scientific evaluation of the MM5 was not possible in this study. However, we believe the operational evaluation (presented in Chapters 8 through 11) is sufficient to serve as the basis for judging whether the model is operating with sufficient reliability to be used in the photochemical modeling portion of this study.

7.2.2 Data Supporting Model Evaluation

Hourly surface observations were obtained from the National Center for Atmospheric Research and the CDPHE to support the evaluation of MM5 near-surface temperature, water vapor, and wind speed fields. The specific NCAR data set used for this purpose was DS472.0 which is the hourly airways surface data. The primary data set available for comparing model performance aloft was the NOAA Forecast Systems Lab and National Climatic Data Center's Radiosonde Data of North America.

7.2.3 Evaluation Tools

The MM5 operational evaluation included calculation and analysis of numerous statistical measures of model performance and the plotting of specific graphical displays to elucidate the basic performance of the model in simulating atmospheric variables. Tables 7-1 and 7-2 identify the specific statistical and graphical procedures that were used to evaluate the MM5 model. These measures have been employed extensively in numerous other prognostic model evaluations (Seaman et al., 1997; Tesche et al., 2001a,b; 2003c,d; Emery and Yarwood, 2001; Emery et al., 2002). The procedures are incorporated into the Model Performance Evaluation, Analysis, and Plotting Software (MAPS) system (McNally and Tesche, 1994) which is described in Appendix A.

Table 7-1. Statistical Measures and Graphical Displays Considered in the MM5 Operational Evaluation.

Statistical Measure	Graphical Display
Surface Winds (m/s)	
Vector mean observed wind speed	Vector mean modeled and observed wind speeds as a function of time
Vector mean predicted wind speed	Scalar mean modeled and observed wind speeds as a function of time
Scalar mean observed wind speed	Modeled and observed mean wind directions as a function of time
Scalar mean predicted wind speed	Modeled and observed standard deviations in wind speed as a function of time
Mean observed wind direction	RMSE, $RMSE_s$, and $RMSE_u$ errors as a function of time
Mean predicted wind direction	Index of Agreement as a function of time
Standard deviation of observed wind speeds	Surface wind vector plots of modeled and observed winds every 3-hrs
Standard deviation of predicted wind speeds	Upper level wind vector plots every 3-hrs
Standard deviation of observed wind directions	
Standard deviation of predicted wind directions	
Total RMSE error in wind speeds	
Systematic RMSE error in wind speeds	
Unsystematic RMSE error in wind speeds	
Index of Agreement (I) in wind speeds	
$SKILL_E$ skill scores for surface wind speeds	
SKILLvar skill scores for surface wind speeds	
Surface Temperatures (Deg-C)	
Maximum region-wide observed surface temperature	Normalized bias in surface temperature estimates as a function of time
Maximum region-wide predicted surface temperature	Normalized error in surface temperature estimates as a function of time
Normalized bias in hourly surface temperature	Scatterplot of hourly observed and modeled surface temperatures
Mean bias in hourly surface temperature	Scatterplot of daily maximum observed and modeled surface temperatures

Statistical Measure	Graphical Display
Normalized gross error in hourly surface temperature	Standard deviation of modeled and observed surface temperatures as a function of time
Mean gross error in hourly surface temperature	Spatial mean of hourly modeled and observed surface temperatures as a function of time
Average accuracy of daily maximum temperature estimates over all stations	Isopleths of hourly ground level temperatures every 3-hr
Variance in hourly temperature estimates	Time series of modeled and observed hourly temperatures as selected stations
Surface Mixing Ratio (G/kg)	
Maximum region-wide observed mixing ratio	Normalized bias in surface mixing ratio estimates as a function of time
Maximum region-wide predicted mixing ratio	Normalized error in surface mixing ratio estimates as a function of time
Normalized bias in hourly mixing ratio	Scatterplot of hourly observed and modeled surface mixing ratios
Mean bias in hourly mixing ratio	Scatterplot of daily maximum observed and modeled surface mixing ratios
Normalized gross error in hourly mixing ratio	Standard deviation of modeled and observed surface mixing ratios as a function of time
Mean gross error in hourly mixing ratio	Spatial mean of hourly modeled and observed surface mixing ratios as a function of time
Average accuracy of daily maximum mixing ratio	Isopleths of hourly ground level mixing ratios every 3-hr
Variance in hourly mixing ratio estimates	Time series of modeled and observed hourly mixing ratios at selected stations

Table 7-2. Statistical Measures and Graphical Displays Considered in the MM5 Scientific Evaluation. (Measures and Displays Developed for Each Simulation Day).

Statistical Measure	Graphical Display
Aloft Winds (m/s)	
Vertically averaged mean observed and predicted wind speed aloft for each sounding	Vertical profiles of modeled and observed horizontal winds at each NWS sounding location and at each NOAA continuous upper-air profiler location in the 36, 12, and 4-km grid.
Vertically averaged mean observed and predicted wind direction aloft for each sounding	
Aloft Temperatures (Deg-C)	
Vertically averaged mean temperature observations aloft for each sounding	Vertical profiles of modeled and observed temperatures at each sounding location
Vertically averaged mean temperature predictions aloft for each sounding	

8.0 MM5 EVALUATION FOR THE SUMMER '02 EPISODE: 36/12 KM GRIDS

This chapter presents selected results of the operational evaluation of the MM5 model for the 6 June–25 July 2002 ozone episode over the Denver-Northern Front Range. In this evaluation we focus on the 36 km and 12 km grid results for the full Summer '02 episode. Chapters 9 and 10 present MM5 performance comparisons at 4 km and 1.33 km scales for the two embedded ozone episodes (16-22 July and 24 June-2 July 2002) to be used in the 8-hr ozone attainment demonstration and control strategy assessment.

8.1 Surface Comparisons

The surface meteorological data available from the National Weather Service (NWS)and other agencies for MM5 performance testing included mixing ratio (i.e. specific humidity), temperature, wind speed and direction. Figures 8-1 and 8-2 present the predicted noontime surface wind speeds on 20 July and 30 June 2002, over the 36 km and 12 km grid regions, respectively. The red arrows represent the surface NWS or other monitoring locations and the solid black arrows are the MM5 predictions. For clarity of presentation, we present only every fourth grid cell prediction. Note also that at the locations where surface winds are reported, measurements of mixing ratio and temperature are typically recorded as well. The CD archive of the Summer '02 simulation contains the full set of hourly predictions and observations for the full episode. Cursory examination of the 36/12 km surface winds did not reveal any unusual or obviously flawed wind predictions, behavior symptomatic of spurious model results.

8.1.1 Mixing Ratio

Figures 8-3 and 8-4 present the daily average gross error and bias in hourly near-surface mixing ratio for the full 6 June-25 July Summer '02 modeling episode. The figures present the 12 km grid results in red and the 36 km grid results in green. The mixing ratio (i.e., specific humidity) measurements are typically taken at a 2 m shelter height while the model predictions are derived from the node (middle point) of the first level in the MM5 model, 17.5 m. Thus, there is an unavoidable mis-match between the height of the measurement (2 m) and the height of the prediction (17.5 m). This introduces differences between the two that are quite independent of any error in the measurement or model prediction. In practice, this proves to be an inherent limitation of a rigorous performance appraisal of the meteorological model since no reliable method is presently available to transform the prediction and measurement to a common height.

Considering first the gross error estimates, Figure 8-3 shows that there is day-to-day variation in the mixing ratio errors on both nested grids. In general, the errors range from 1.5 gm/Kg to 2.2 gm/Kg. On most days, the MM5 performs better on the 36 km grid compared to the 12 km grid, but on a few days the performance is comparable. The episode average gross errors in mixing ratio on the 36 km and 12 km grids (1.48 gm/Kg and 1.92 gm/Kg, respectively) are shown in the two far right hand bars in Figure 8-3.

For the mixing ratio bias (Figure 8-4), the MM5 tends to underestimate the day-to-day values on the 36 km grid but overestimate bias on the 12 km grid for most days. In general the biases range from about –2 gm/Kg to 1.2 gm/Kg. For daily average bias, the MM5 performs systematically better on the 36 km grid compared to the 12 km grid. The episode average biases in mixing ratio on the 36 km and 12 km grids (-0.13 gm/Kg and 0.50 gm/Kg, respectively) are actually quite good given the 15.5 m vertical displacement between the measurement height and height of the layer 1 MM5 prediction.

8.1.2 Temperatures

Figures 8-5 and 8-6 depict the daily average gross error and bias in hourly near-surface temperatures on the two nested grids. Unlike the mixing ratios, the temperatures are estimated by MM5 (via a post-processor) for a height of 2 m, corresponding to the typical shelter height. For the temperature gross errors, Figure 8-5 shows typical day-to-day variation in the temperature errors on both grids. In general, the temperature gross errors range from 2 deg C to 3 deg C. On most days, the MM5 performs better on the 36 km grid compared to the 12 km grid. The episode average gross errors in temperature on the 36 km and 12 km grids are 2.13 deg C and 2.98 deg C, respectively.

For the daily average temperature bias (Figure 8-6), the MM5 tends to underestimate the day-to-day values on both grids for all days, a feature we attribute to the Pleim-Xiu (PX) surface module.[1] The daily average of the hourly temperature biases range between about –0.5 deg C to –2.7 deg C. For daily average temperature bias, the MM5 performs systematically better on the 36 km grid compared to the 12 km grid. The episode average biases in temperature on the 36 km and 12 km grids are -0.83 deg C and –1.87 deg C, respectively. Note that while temperature bias over the 12 km domain is, on average, roughly twice that on the 36 km grid, when averaged over the full episode. The implications of this systematic underestimation of hourly temperatures on the 36 km and 12 km grids is discussed in greater detail at the end of this chapter.

Because of the important hour-to-hour effect of temperature on biogenic emissions and evaporative emissions (motor vehicles, industrial solvents, etc), we also examined the average diurnal behavior of the MM5's predictions of near surface temperature on the 36 km and 12 km grids. Figure 8-7 presents these results. While it is somewhat difficult to tell from these spatial mean plots, the MM5 has a systematic tendency to underestimate the afternoon high temperatures but does a fairly reasonable job of reproducing the nighttime lows. This is more readily apparent in examining the temperature time series at individual weather station sites. (These data are on CD.)

The daily average peak temperature prediction accuracy over the 36 km and 12 km grids is shown in Figure 8-8. On the 36 km grid, the accuracies of peak temperature prediction range between roughly 8%-12% while for the 12 km domain the range is about 6%-9%. Of course, on some days the accuracy figures are larger or smaller than these ranges. These accuracy figures for daily maximum temperature prediction are typical of mesoscale model simulations, but this statistic reveals little about the systematic tendency to underestimate the hour-by-hour temperatures discussed above

8.1.3 Wind Speed and Direction

The height discrepancy issue that affects surface mixing ratios comparisons is diminished somewhat when examining the surface wind measurements since the latter are customarily taken at standard anemometer height, 10 m. Nonetheless, we expect a systematic overestimation in surface wind speed since speed typically increases with height and in the lower boundary layer a 7.5 m height differential introduce a perceptible bias.

[1] The current PX scheme does not warm up the ground temperature fast enough after initialization, typically producing a cool bias in the MM5 model. This has been observed not only in the Denver EAC study at 36/12 km scale, but also in the VISTAS MM5 modeling over the eastern U.S. at 36 km scale. This deficiency in the PX scheme has been avoided in the nested 4 km and 1.33 km simulations in the present Denver application by replacing it with the established Blackadar PBL scheme and no explicit land surface model.

Figures 8-9 through 8-14 present daily average wind performance statistics on the 36 km and 12 km grids. Beginning with Figure 8-9, the daily average modeled and observed surface wind speeds on the 12 km grid are plotted as a function of time during the episode. The mean predicted wind speeds are equal to the observations when averaged across all days in the Summer '02 episode. Specifically, over the 12 km grid, the average modeled and observed surface wind speeds are 1.88 m/s. For the 36 km winds (Figure 8-10), the average modeled and observed surface wind speeds are 1.06 m/s and 0.88 m/s, respectively. Note that the wind speeds on the 36 km grid are systematically slower (roughly one-half) than the 12 km grid.

Figure 8-11 depicts the daily average index of agreement on the 36 km and 12 km grids. This metric may be viewed as an overall 'correlation coefficient' for the surface wind speeds. On the 36 km grid, the index ranges between about 0.82 to 0.90 on some days. The 12 km grid results are just slightly poorer but still ranging from about 0.82 to 0.9. Across the entire episode, the index of agreement for the 36 km and 12 km grids are 0.88 and 0.85, respectively.

Daily average root mean square errors (RMSE) vary between about 2.1 m/s to 2.4 m/s on the 36 km grid (Figure 8-12) while they are somewhat poorer on the 4 km grid, ranging between about 2.1 m/s to about 2.9 m/s. Across the entire episode, the RMSE errors the 36 km and 12 km grids are 2.21 m/s and 2.34 m/s, respectively.

Figures 8-13 and 8-14 present comparisons between the daily average predicted and observed wind direction for the two MM5 grids. On the 12 km grid (Figure 8-13), the wind directions match quite well on virtually every day although there is obviously large day-to-day variability in the daily average direction. The same good agreement is seen in the 36 km winds (Figure 8-14). Across the entire episode, the average wind direction discrepancies on the 12 km grid is 27 deg, while for the 36 km grid the mean difference is 19 deg. Given the very wide range of synoptic and geophysical conditions encompassed by both the 36 km and 12 km grid scales, the wind direction agreement is good.

8.2 Aloft Comparisons

The aloft meteorological data available from the National Weather Service for MM5 performance testing included twice daily rawinsonde soundings at various airports. These soundings record specific humidity, pressure, temperature, wind speed and wind direction at all mandatory reporting levels.

A brief examination of the MM5 model's performance aloft included a qualitative evaluation of the planetary boundary layer height (PBL) fields for each hour of the Summer '02 episode. This evaluation could only be qualitative since pbl height is not routinely measured in the western U.S. and there were no radar profiler or aircraft data available for this purpose. Figure 8-23 presents a typical pbl height field at 1400 MDT on 14 July 2002 over the 12 km domain. (This day was selected since it was the warmest day of the Summer '02 episode on the 12 km grid). This plot reveals very high afternoon boundary layer heights (2400m and more) over the Great Basin and the Great Plains states and the lower depths over mountainous terrain. The full set of fields (contained on the CD archive) was examined for consistency. No unusual or erratic behavior was observed.

We also examined the MM5's performance aloft by comparing the modeled and observed upper level horizontal winds, temperatures, and mixing ratios for at two times each day across the 36/12 km grid domains. Due to data limitations (i.e., the monitoring data were limited to twice-daily balloon soundings), this evaluation was constrained to a comparison of daily-averaged winds and temperatures.

To place the balloon observations on a comparable basis with the model's grid layer predictions, we utilized Alpine's Flying Data Grabber routines to generate vertically-integrated meteorological variables for each model layer. Examination, of the vertically-integrated horizontal winds and temperatures revealed generally quite good agreement aloft. Part of this good agreement results because the aloft temperature and wind measurements from the NWS rawinsondes were employed in the FDDA nudging scheme. However, the MM5 weighting coefficients used in the nudging were fairly small so that the aloft fields were not under a heavy constraint to match the observations locally. Thus, this good agreement in the estimation of aloft temperatures, wind speeds, and wind directions is encouraging and gives some confidence that the modeled wind patterns are a reasonable approximation to the conditions that actually occurred. However, this evaluation is insufficient by itself to 'validate' the reasonableness of the model predictions aloft.

Further insight into the aloft model performance was developed for each day by constructing and examining so-called 'skew-T' plots of the modeled and observed wind and thermodynamic profiles. These plots were developed from the MM5 output on all three grid scales for each rawinsonde sounding in the domain. (The full set of plots is contained on the CD archive). Figures 8-24 and 8-25 are examples of a skew-T plot on 4 July at 1585 MDT at Albuquerque and Salt Lake City, respectively. The solid blue line represents the MM5 upper air temperature profile while the solid red line corresponds to rawinsonde observations. The thin blue and red lines denote the mixing ratio predictions and observations, respectively. Modeled and observed horizontal winds are shown in the stick plots to the right. Perusal of the results from the Albuquerque and Salt Lake City sites from these soundings reveals that there is fairly good agreement between modeled and observed temperatures through the depth of the lower troposphere. The MM5 model exhibits a warm bias at both sites throughout the entire vertical region. The model does a fair job of simulating the depth of the low level boundary layer, as evidenced by the mixing ratio discontinuity between 650-700 mb at Albuquerque and 550-600 mb over Salt Lake City. Systematic inter-comparison of the modeled vertical structures within the lowest thousand meters would be very interesting, but was outside the scope of this project. These data sets have been archived and are available to parties interested in conducting further analyses.

8.3 Comparisons with Other Studies

Figures 8-15 through 8-23 and Table 8-1 summarize the episode composite temperature, wind speed, wind direction and mixing ratio statistics for the 6 June–25 July Summer '02 MM5 episode in the context of nearly fifty recent regional scale MM5 and RAMS model performance evaluations performed by Alpine Geophysics since 1995. Virtually all of the studies covered in these figures and Table 8-1 were regulatory applications of prognostic models, either for ozone or PM. These results derive from model applications grid scales ranging from 4 km to 12 km since these are the ones most commonly reported. Included in the Table 8-1 are the bias and error in hourly ground level air temperature, bias and error in mixing ratio, surface wind speed error (i.e., difference between predicted and observed wind speed), wind speed root-mean-square-error (RMSE), Index of Agreement, and the average difference between predicted and observed hourly wind direction at each monitor. These statistics are based on hourly comparisons between predictions and observations in the first model layer near the ground. These lowest layers typically had node heights at 10 to 20 meters. Recall that in the Denver EAC application, the MM5 node height was 17.5 meters. The episode composite statistics for the 6 June–25 July 2002 episode, reported in Table 8-1, are derived from the daily statistics on the 12 km MM5 grid for each episode day which are themselves averaged over the full 24-hour diurnal cycle.

Below we put the current summer '02 MM5 results into this broader context of historical mesoscale modeling experience. We compare the episode composite statistical performance measures on the 12 km grid for the Summer '02 episode and with the mean of the nearly fifty other RAMS and MM5 simulations shown in Figures 8-15 through 8-22. These 'global' comparisons are be helpful in making *general assessments* of the current MM5 modeling results; however, it is clear that the calculation of episode mean statistics can conceal important day-to-day and/or hour-to-hour variations that may be quite important in terms of affecting an emissions or air quality model simulation.

8.3.1 Bias and Error in Mean Temperatures

The MM5 average bias in hourly ground level temperatures over the 12 km grid is –1.9 deg C. The average across all studies is –0.2 deg C (Figure 8-15). Thus, the current simulation tends to have a systematic cool bias relative to the other studies. The source of this bias is the PX land use module. We note that this problem has been rectified in the 4 km and 1.33 km MM5 simulations (discussed in Chapters 9 and 10 by reverting back to the Blackadar pbl scheme). The current 36/12 km MM5 simulation was not rerun, however, because this cool bias was not believed to be consequential in terms of adversely affecting the regional transport fields relative to the inner 4/1.33 km nests. The episode average error in hourly ground level temperatures for the 12 km MM5 simulation is 3.0 deg C. The mean temperature error over all studies is 2.0 deg C (Figure 8-16). Thus, the current simulation tends to have somewhat higher gross errors in ground level hourly temperature predictions compared to other studies. Again, the source of these errors is understood, has been rectified in the 4/1.33 km simulations and not believed to adversely impact the overall ozone regulatory modeling.

8.3.2 Bias and Error in Mean Mixing Ratios

The MM5 average bias in hourly ground level mixing ratios is 0.5 gm/Kg. The average across all studies is 0.0 gm/Kg (Figure 8-17). This slight positive moisture bias relative to the other studies is inconsequential. The episode average error in hourly ground level mixing ratio for the current MM5 run is 1.9 gm/Kg. The mean error over all studies is 1.8 gm/Kg (Figure 8-18). As with the temperatures, the Denver 12 km MM5 simulation tends to have comparable errors in ground level mixing ratios compared to other studies.

8.3.3 Error in Mean Wind Speed

The episode average error in hourly ground level wind speed is 0%. The average across all studies is 32% (Figure 8-19). Thus, the present simulation tends to have much lower wind speed errors on the 12 km grid, on average, compared to other studies. Part of this may be due to the fact that many of the historical studies identified in Table 8-1 derive from locations in the country where land-lake breeze circulations are present.

8.3.4 RMSE in Surface Wind Speeds

The episode average root mean square error (RMSE) in hourly ground level wind speed prediction is 2.3 m/s. The average RMSE error over all studies is 2.00 m/s (Figure 8-20). Unlike the surface wind speed errors, the Denver EAC 12 km grid simulation tend to have slightly higher RMSE wind speed errors, on average, compared to other studies.

8.3.5 Index of Agreement

The episode average IOA for hourly ground level wind speed prediction is 0.85. The average index of agreement over all studies is 0.71 (Figure 8-21). Here again, the Denver EAC 12 km MM5 simulation tends to have a better a statistical score compared to other studies.

8.3.6 Error in Mean Wind Direction

The episode average discrepancy in hourly wind direction for the Denver EAC 12 km simulation is 27 deg C which may be compared with an average wind direction error of 24 deg C from the other studies (Figure 8-22). Thus, on the 12 km grid at least, the current MM5 simulation tends to have significantly lower wind direction errors compared to other studies.

In summary, the statistical results for temperature, mixing ratio, and winds for the 6 June–25 July Summer '02 episode are quite consistent with the range of typical past mesoscale modeling performance in ozone and PM regulatory studies. The episode composite means for surface wind speed error and index of agreement are better than the typical performance levels in nearly 50 other MM5/RAMS modeling studies while the performance for wind direction and mixing ratio statistics are roughly comparable to the nationwide averages. As discussed above, the bias and error statistics for surface temperature in the Denver 36 km and 12 km simulations are poorer than the national average but the causes for these departures are understood and will not affect significantly the application of the models to the 4 km and 1.33 km Denver high resolution ozone modeling domains.

8.4 Assessment of Model Reliability and Suitability

A key question is whether the newly-created 36/12 km nested MM5 meteorological fields are adequate for their intended use in supporting the ozone modeling for the Denver EAC 8-hr ozone modeling study. We believe that the evaluation results presented in this chapter suggest that the 36/12 km MM5 simulation is indeed suitable for use in generating inputs to the nested 4/1.33 km MM5 domains and in providing general inputs to the CAMx model over the 36/12 km regional grid. Of course, there is no simple way to answer definitively this question. First, there are no commonly accepted performance benchmarks for prognostic meteorological models that, if passed, would allow one to declare the MM5 fields appropriate for use. For complex atmospheric modeling problems like the ones being addressed in this project, it is quite doubtful that any set of quantitative performance criteria will ever be completely sufficient. Benchmarks are needed and useful, but they do not provide the whole answer. Additional performance evaluation procedures in the form of 'weight of evidence' analyses are also required to supplement these simplified statistical measures. We discuss these 'weight of evidence' procedures, a necessary compliment to the summary statistics, in Chapter 11.

The question of meteorological data set adequacy depends, at a minimum, upon the specific host emissions and air quality models (EPS2x, CAMx in this instance) and the nature of the modeling episodes being used. Meteorological fields that might be adequate for use in an ozone model over a simple urban setting, for example, may be quite deficient in a seasonal PM episode over the great lakes region since the specific needs of the air quality model and the particular chemical and physical processes that must be simulated are different. Thus, quantitative statistical and graphical performance criteria, though helpful, are inherently insufficient in aiding modelers and decision-makers in deciding whether meteorological fields are adequate for air quality modeling. Other considerations must be brought to bear.

To formally judge the adequacy of the MM5 fields for use in the CAMx modeling based on currently available information we adopted the process used in several recently-completed air quality modeling studies (e.g., SAMI, Peninsular Florida 8-hr ozone, We Energies ozone/PM studies.) These studies utilized a formalism employing meteorological model 'performance benchmarks' proposed by Emery et al., (2001) which draw on earlier work by Roth, Tesche and Reynolds (1998) and Tesche et al., (2000, 2003c,d).

Three recent studies (Tesche et al., 2000; 2003b; Emery et al., 2001) formulate a set meteorological model performance benchmarks based on the most recent prognostic meteorological model evaluation literature. The purpose of these benchmarks is not to assign a passing or failing grade to a particular meteorological model application, but rather to put its results into a useful decision-making context. These benchmarks have proven to be helpful to decision-makers in understanding how poor or good their results are relative to the range of other model applications in other areas of the U.S. Certainly an important limitation of the EPA guidance for 1-hr ozone performance statistics is that they are relied upon much too heavily to establish an acceptable (to the EPA) model simulation of a given area and episode. Often lost in routine statistical ozone model evaluations is the need to critically evaluate all aspects of the model via the diagnostic and process-oriented approaches. The same must stressed for the meteorological performance evaluation. Thus, the appropriateness and adequacy of the current meteorological performance benchmarks should be carefully considered based upon the results of the specific meteorological model application being examined.

Since the mid 1990s, the study team has performed over 50 MM5 and RAMS model performance evaluations over grid scales ranging from 1.33 km to 36 km, (Tesche et al., 2001, 2002). The results of these varied model evaluations provide a solid foundation against which to compare the current MM5 modeling results for the We Energies analysis. Using this database as a guide, we consider the meteorological model performance benchmarks suggested by Emery et al, (2001):

Parameter	*Measure*	*Benchmark*
Wind Speed	**RMSE:**	**≤ 2 m/s**
	Bias:	**≤ ±0.5 m/s**
	IOA:	**≥ 0.6**
Wind Direction	**Gross Error:**	**≤ 30 deg**
	Bias:	**≤ ±10 deg**
Temperature	**Gross Error:**	**≤ 2 K**
	Bias:	**≤ ± 0.5 K**
	IOA	**≥ 0.8**
Humidity	**Gross Error:**	**≤ 2 g/kg**
	Bias:	**≤ ±1 g/kg**
	IOA:	**≥ 0.6**

Table 8-2 presents the results of comparing the MM5 episode average 36 km and 12 km statistical results (for those statistics that were produced in this study) with the proposed meteorological modeling benchmarks. Shaded cells in the table correspond to those meteorological variables that fall just outside of the benchmark ranges. On the 12 km grid, the bias and error in mixing ratio estimation and the wind speed index of agreement and wind direction errors all fall within the benchmarks. For the reasons noted earlier, the temperature bias and errors are outside the benchmarks as is the RMSE error in surface wind speed. When considering the full body of surface and aloft performance results on the 36/12 km grids, particularly in light of substantial experience with evaluations with this model in other regulatory applications, we conclude that the MM5 meteorological fields for the 6 June-25 July Summer '02 episode may be used, with reasonable caution, as input to: (a) nested 4/1.33 km MM5 simulations and (b) the regional emissions and photochemical/aerosol models for air quality impacts assessments for the Denver EAC study. In the aggregate, the nested 36/12 grid MM5 fields developed in this application are adequate and are in many cases better than the typical meteorology used in regulatory applications.

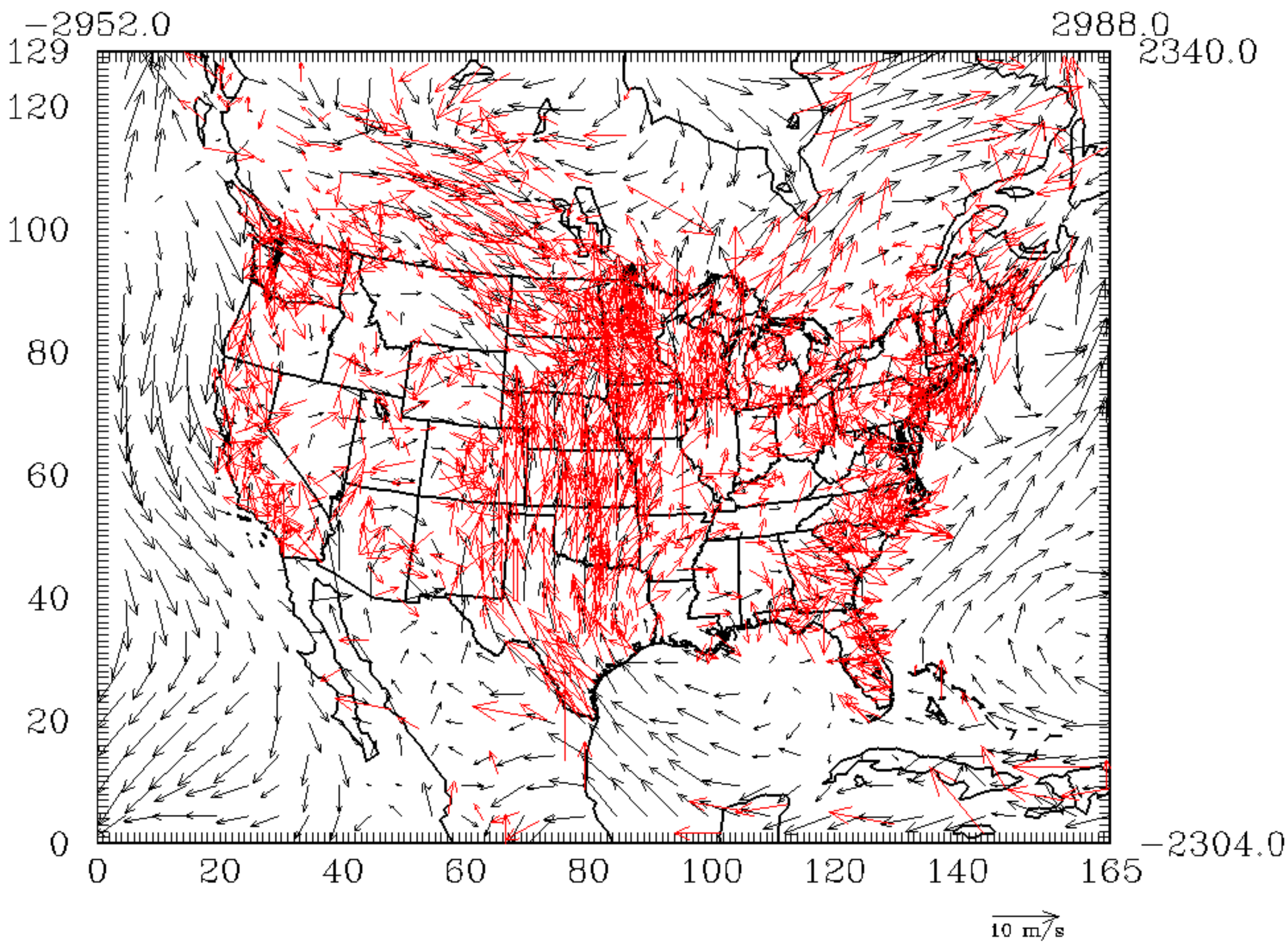


Figure 8-1. MM5 Surface Wind Fields at 1200 MDT on 20 July 2002 Over the 36 km Domain. (Predicted Vectors Plotted Every Fourth Grid Cell).

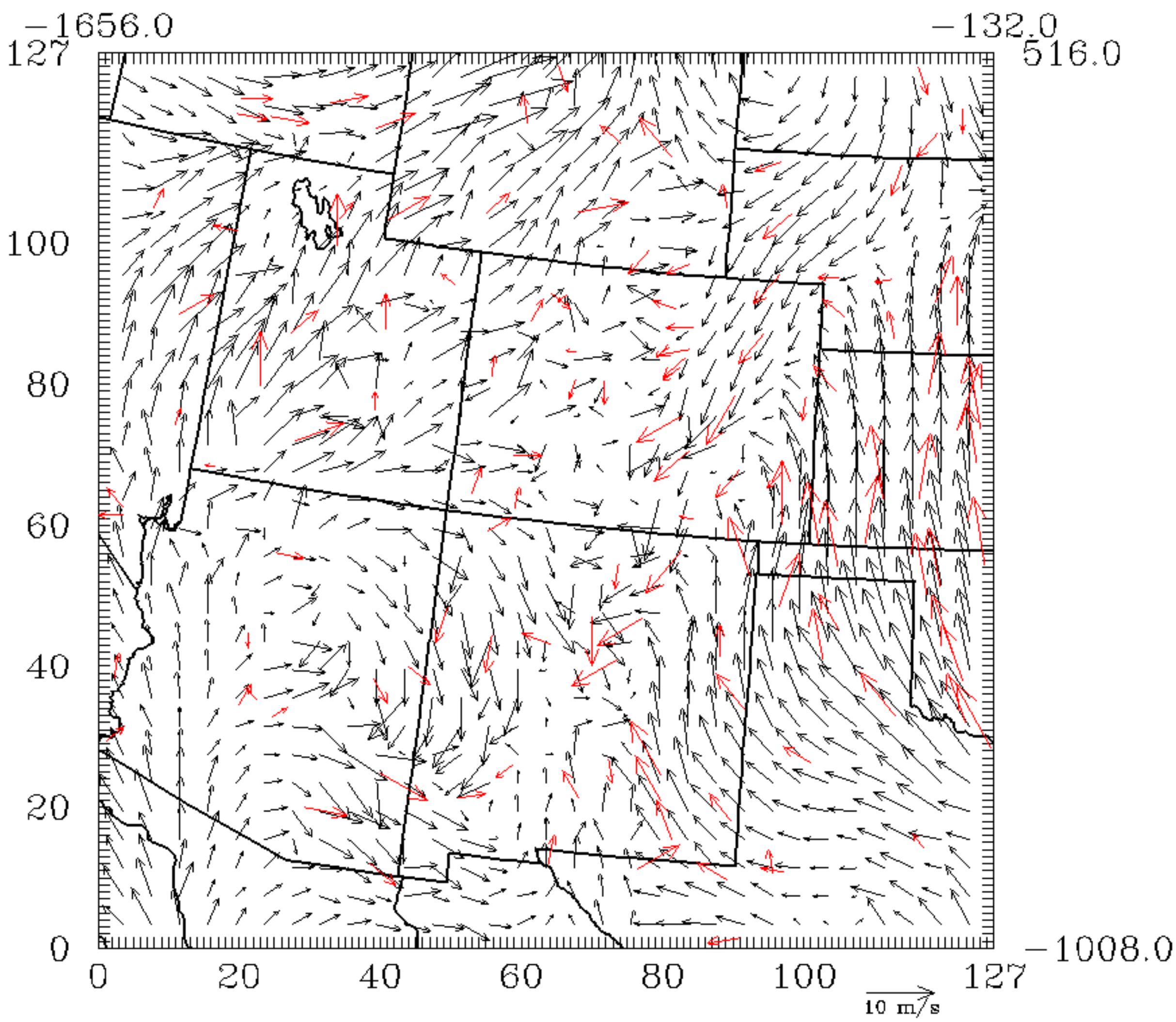


Figure 8-2. MM5 Surface Wind Fields at 1200 MDT on 30 June 2002 Over the 12 km Domain. (Predicted Vectors Plotted Every Fourth Grid Cell).

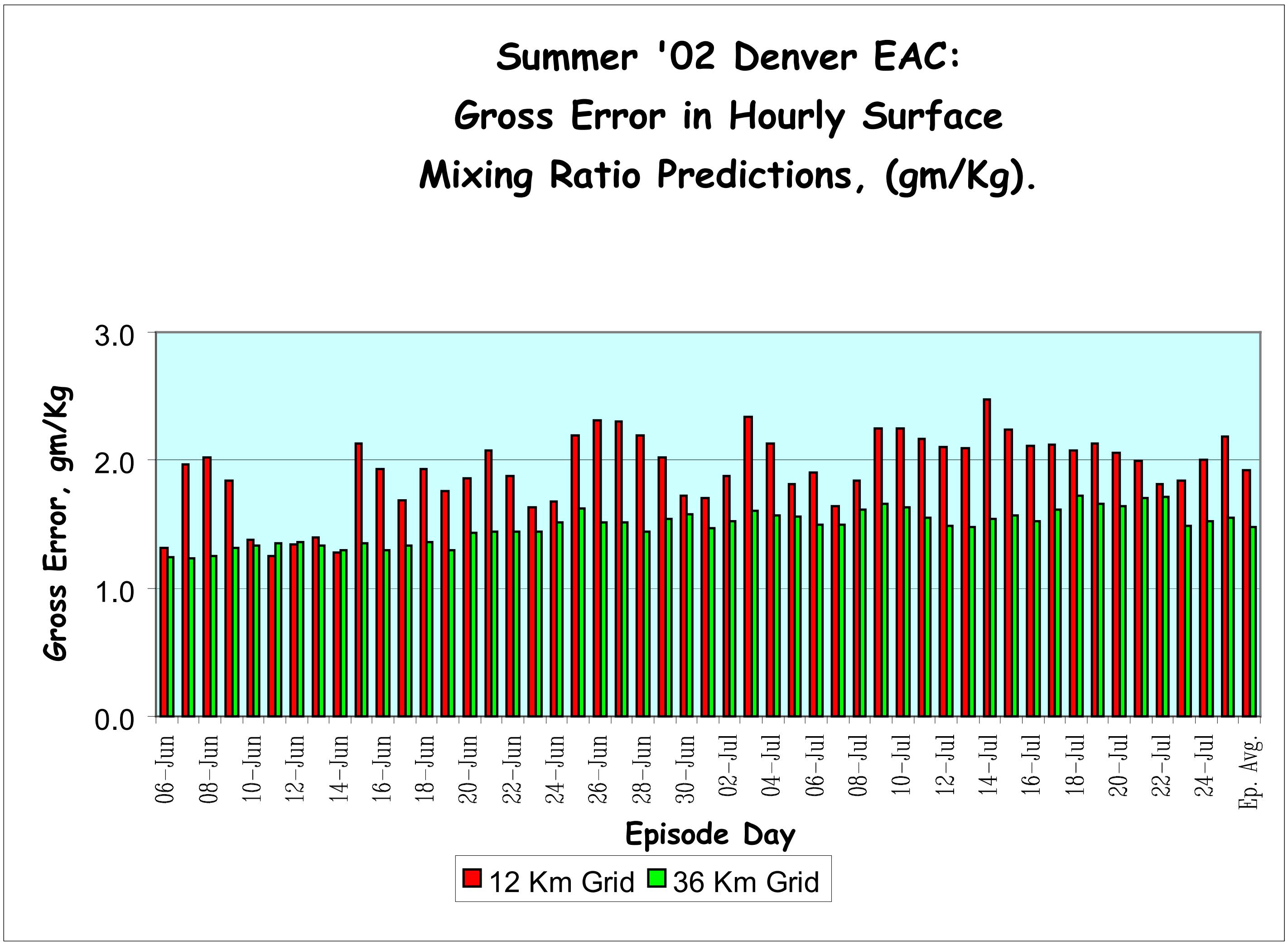


Figure 8-3. **Gross Error In MM5 Hourly Surface Mixing Ratio (gm/Kg) for the 6 June to 25 July 2002 Summer '02 Ozone Episode.**

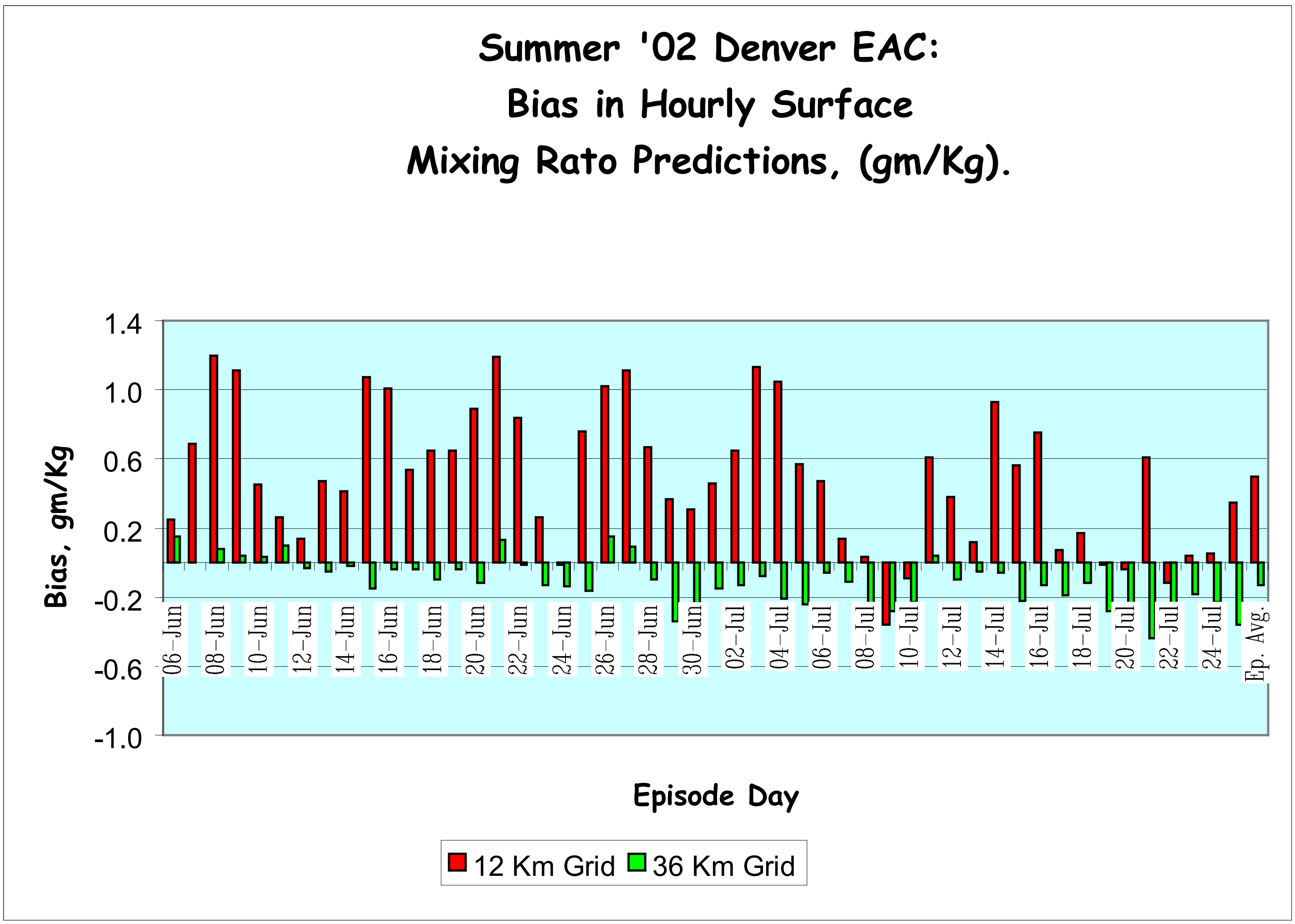


Figure 8-4. Bias In MM5 Hourly Surface Mixing Ratio (gm/Kg) for the 6 June to 25 July 2002 Summer '02 Ozone Episode.

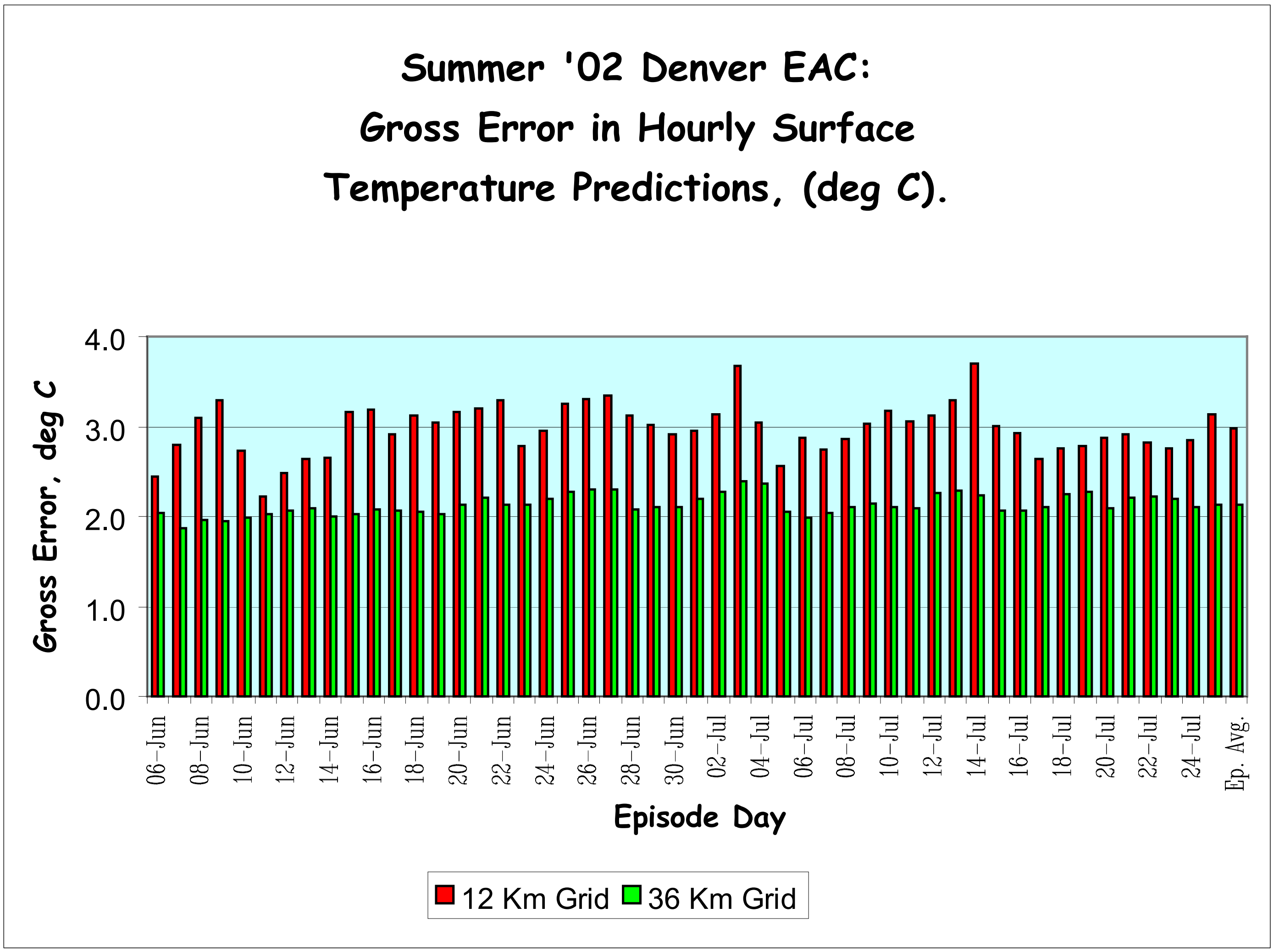


Figure 8-5. Gross Error In MM5 Hourly Surface Temperatures (deg C) for the 6 June to 25 July 2002Summer '02 Ozone Episode.

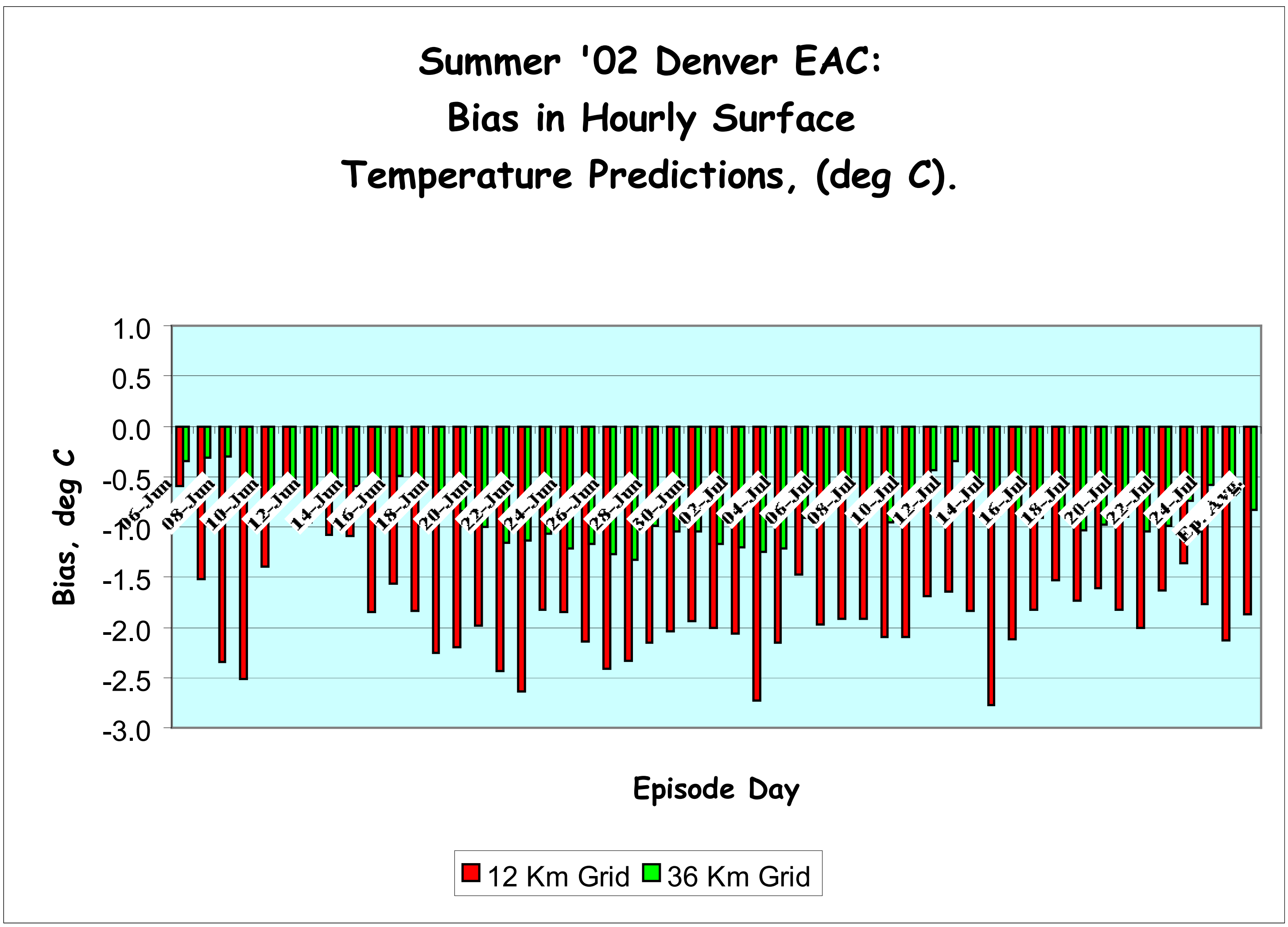
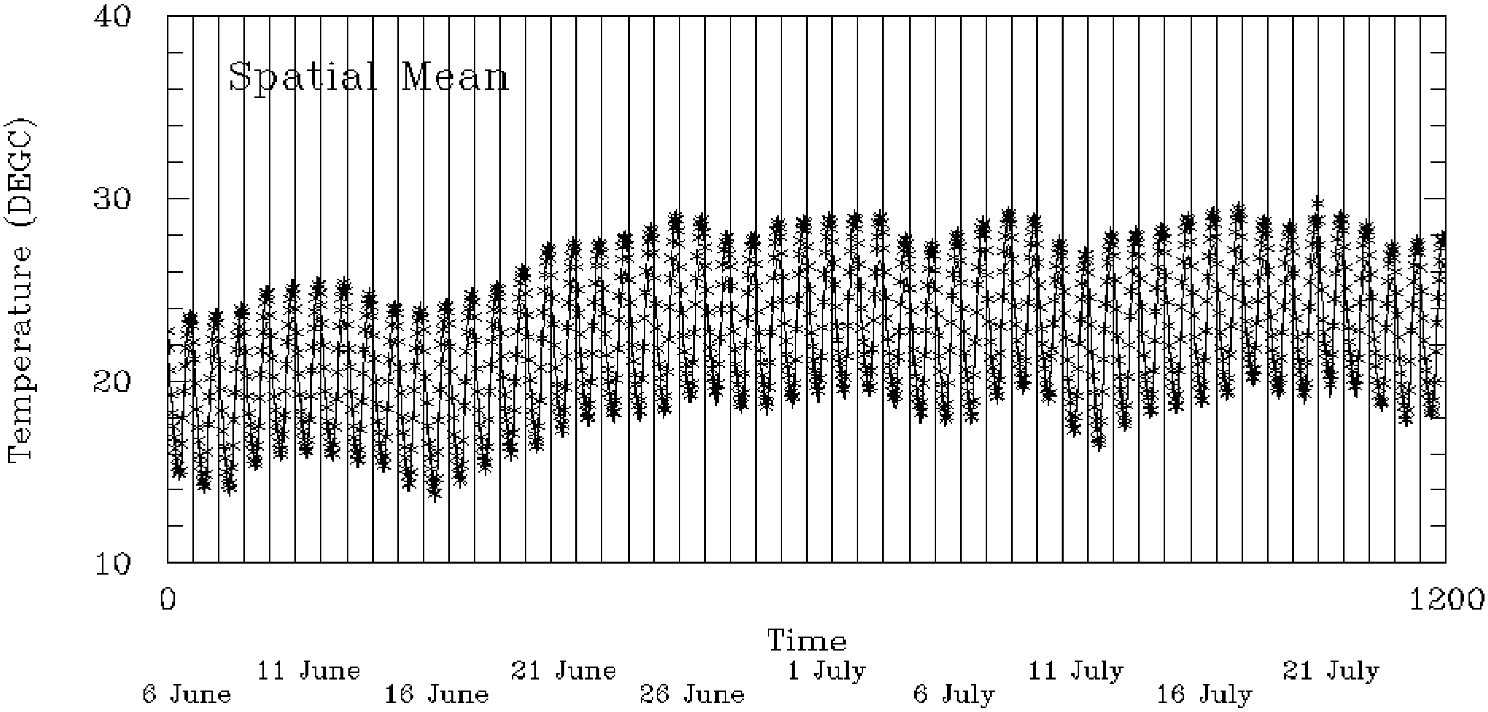
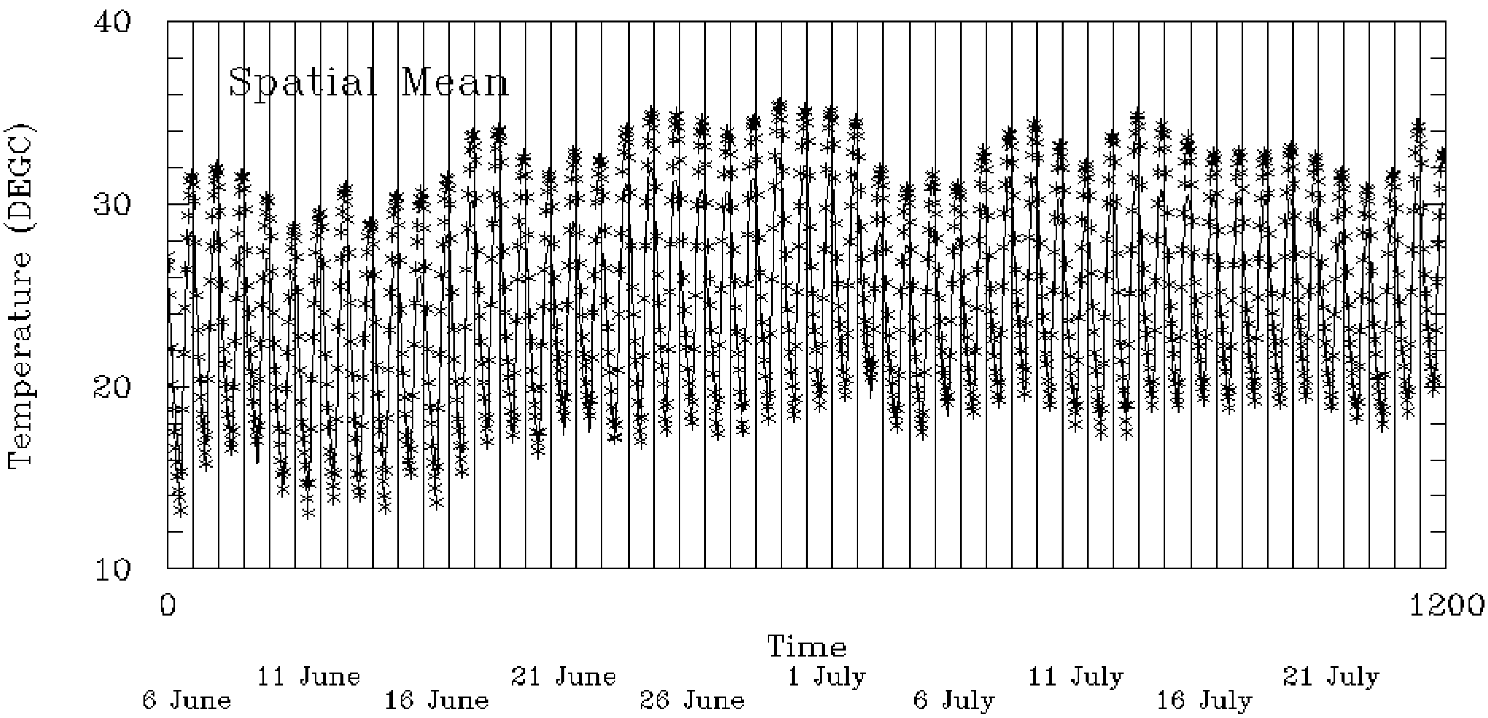


Figure 8-6. Bias In MM5 Hourly Surface Temperatures (deg C) for the 6 June to 25 July 2002 Summer '02 Ozone Episode.



(a) 36 Km Grid



(b) 12 Km Grid

Figure 8-7. Diurnal Variation in Spatial Mean Surface Temperatures for the 6 June–25 July 2002 Summer '02 Ozone Episode.

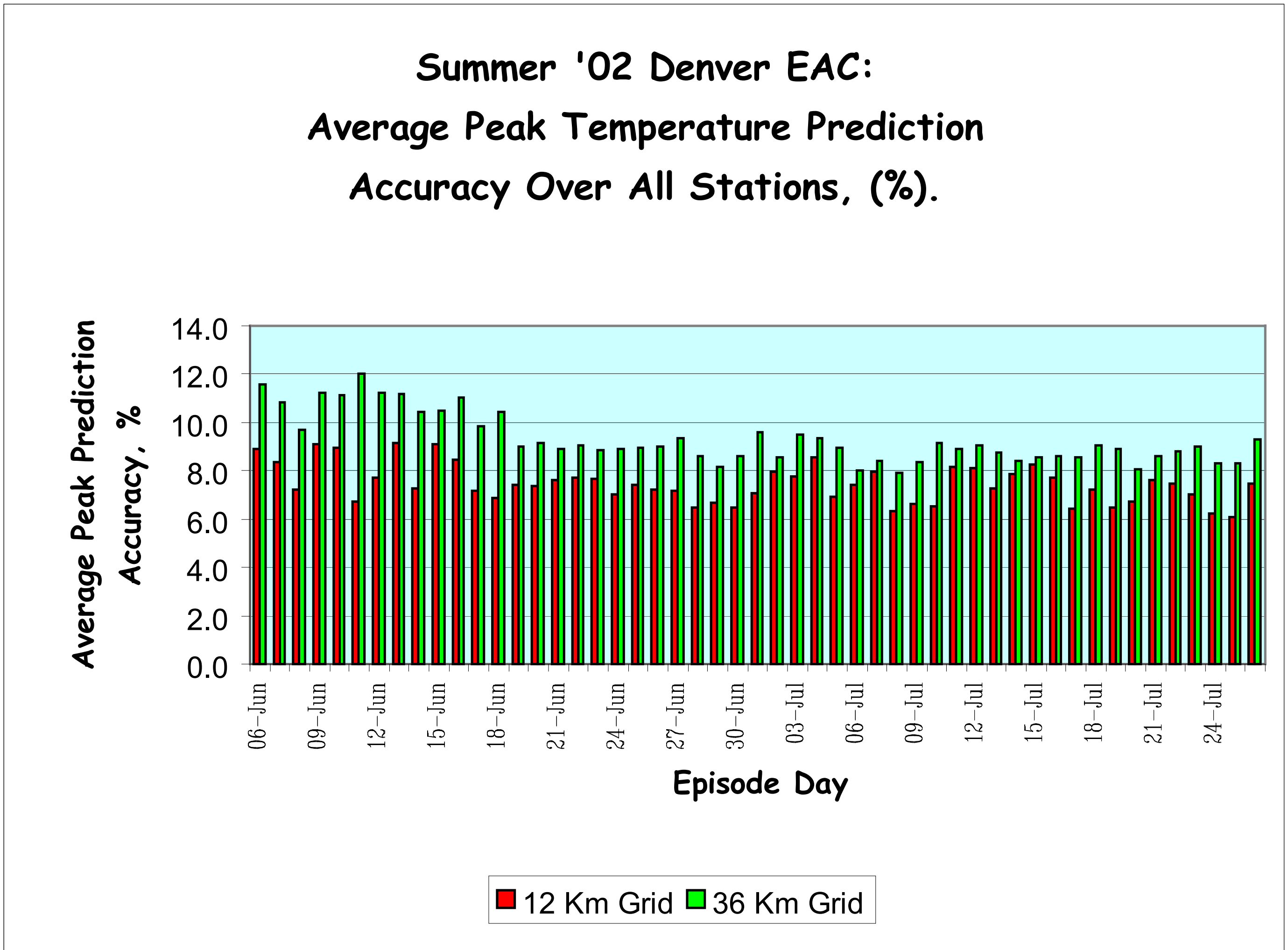


Figure 8-8. Average Peak Prediction Accuracy Over All Monitors for MM5 Hourly Surface Temperatures (deg C) for the 6 June-25 July 2002 Summer '02 Ozone Episode.

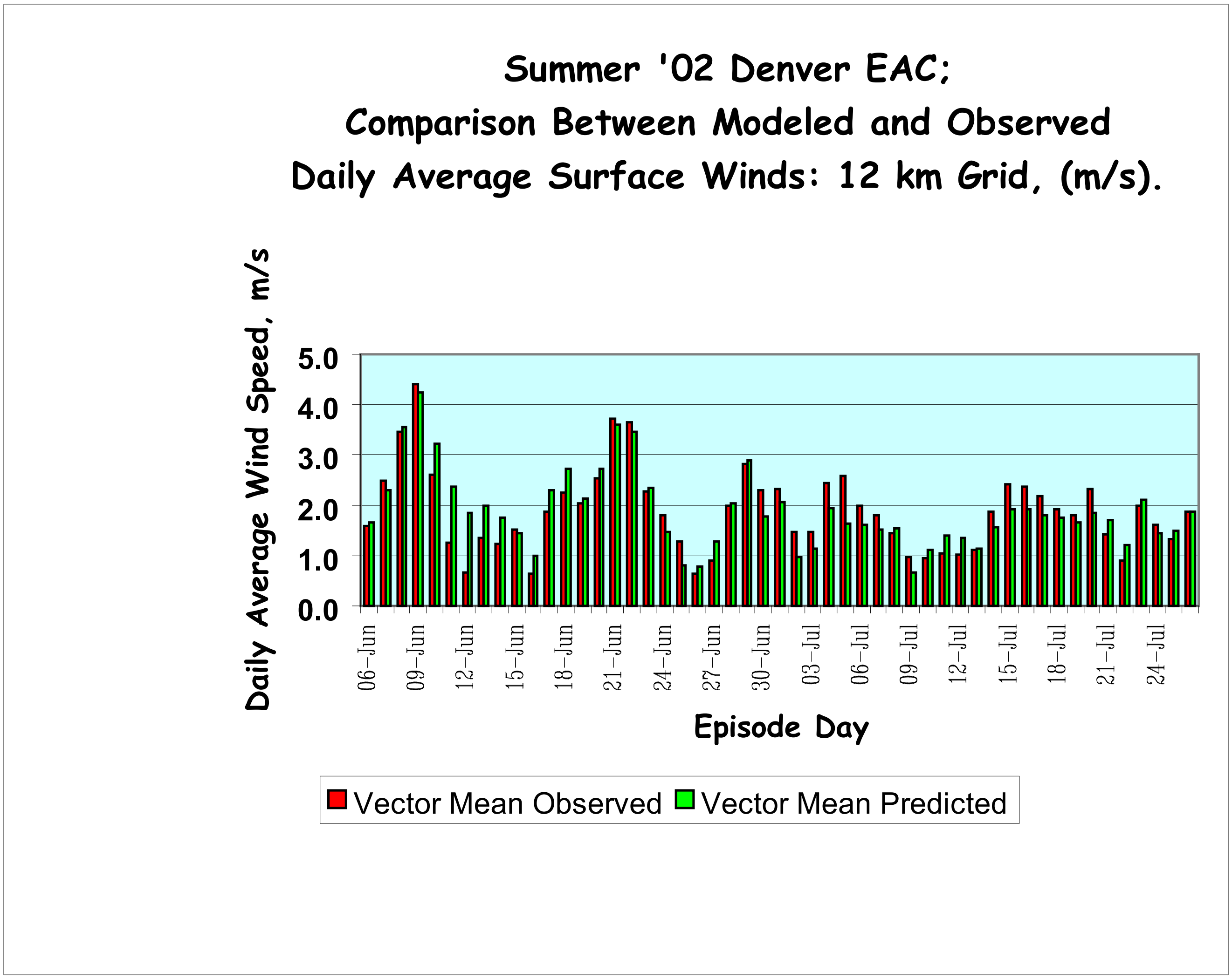


Figure 8-9. Daily Average Modeled and Observed Surface Winds (m/s) on the 12 km Grid for the 6 June –25 July 2002 Summer '02 Ozone Episode.

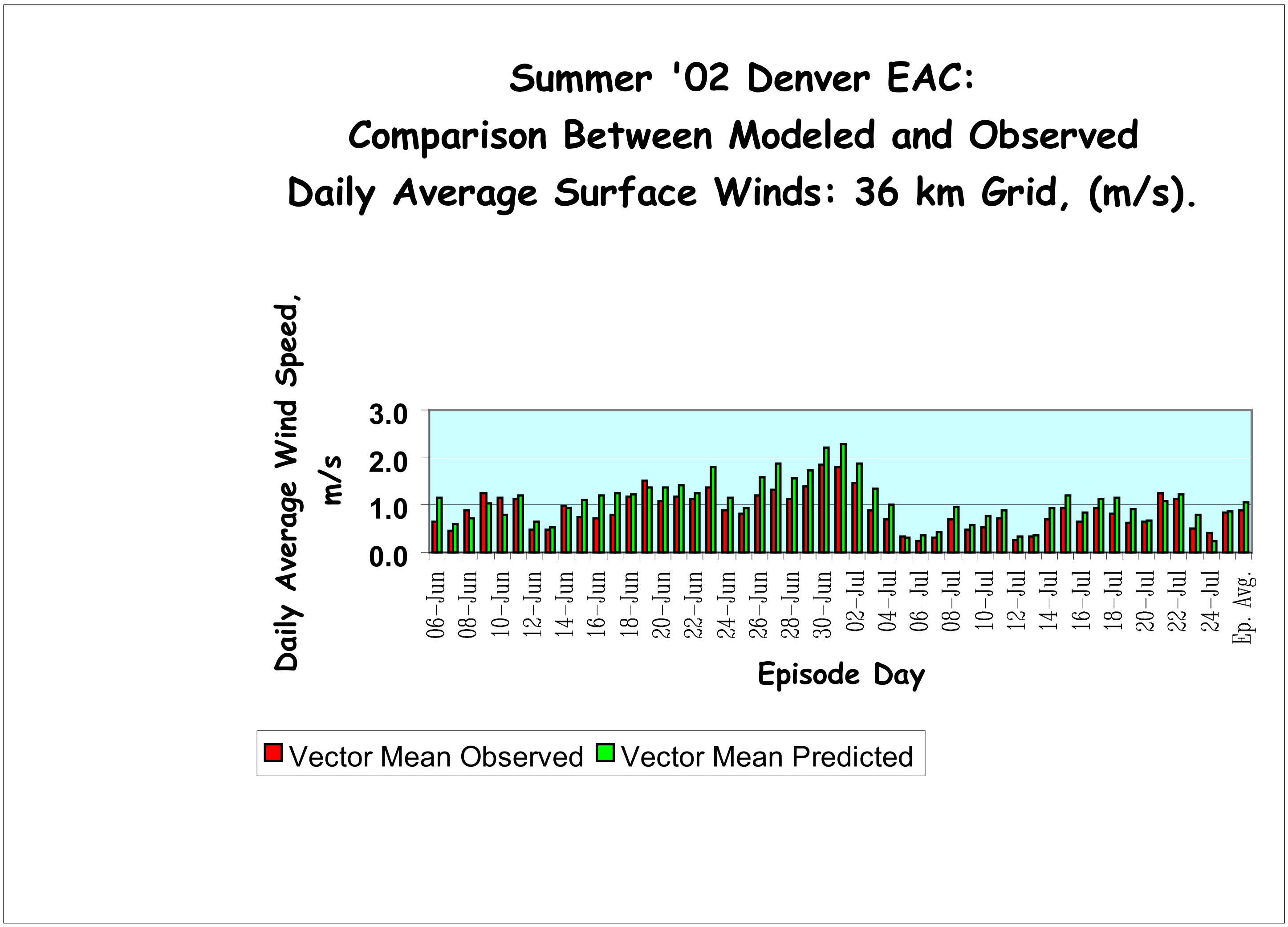


Figure 8-10. Daily Average Modeled and Observed Surface Winds (m/s) on the 36 km Grid for the 6 June-25 July 2002 Summer '02 Ozone Episode.

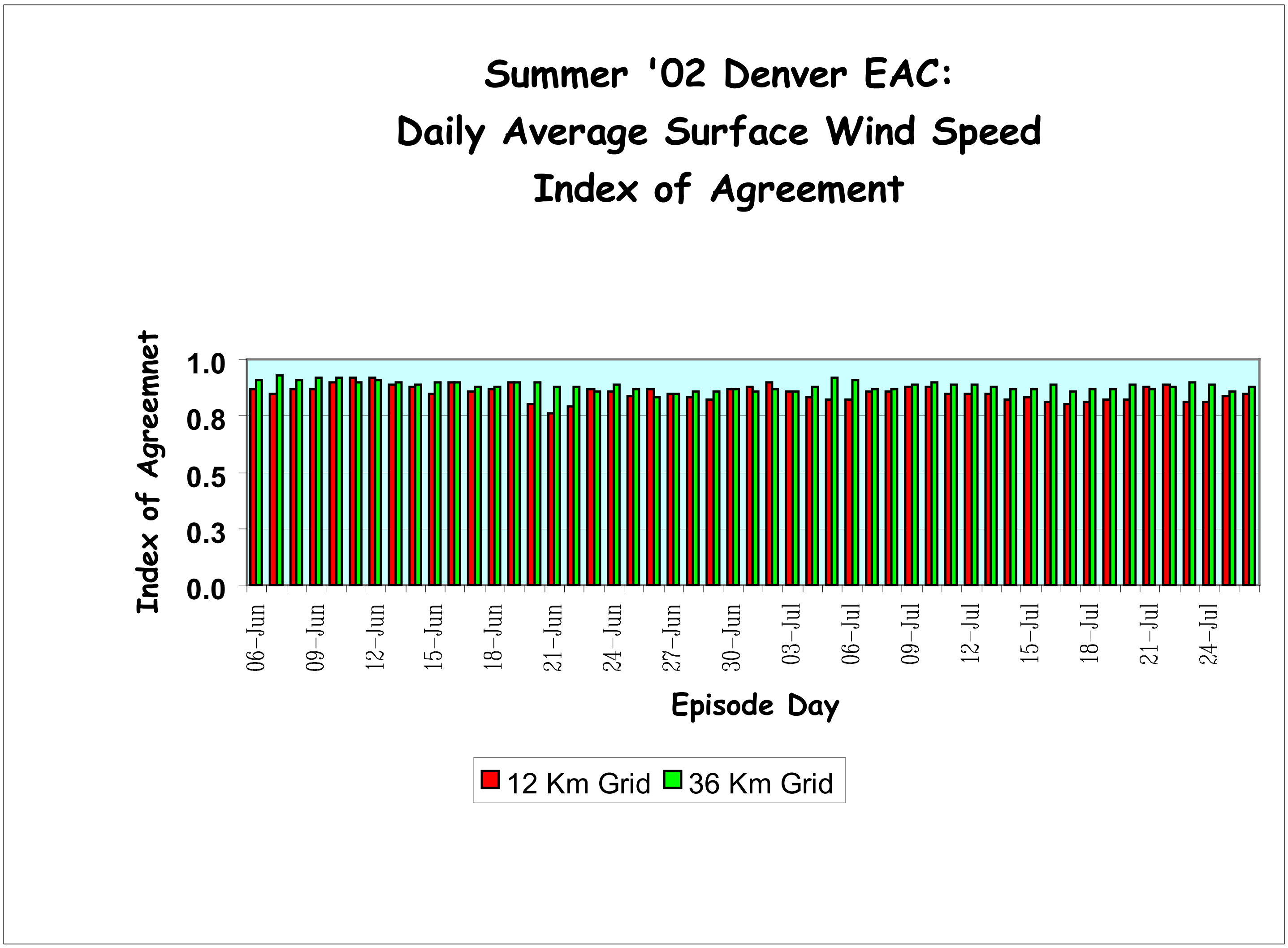


Figure 8-11. Daily Average Surface Wind Index of Agreement for the 6 June-25 July 2002 Summer '02 Ozone Episode.

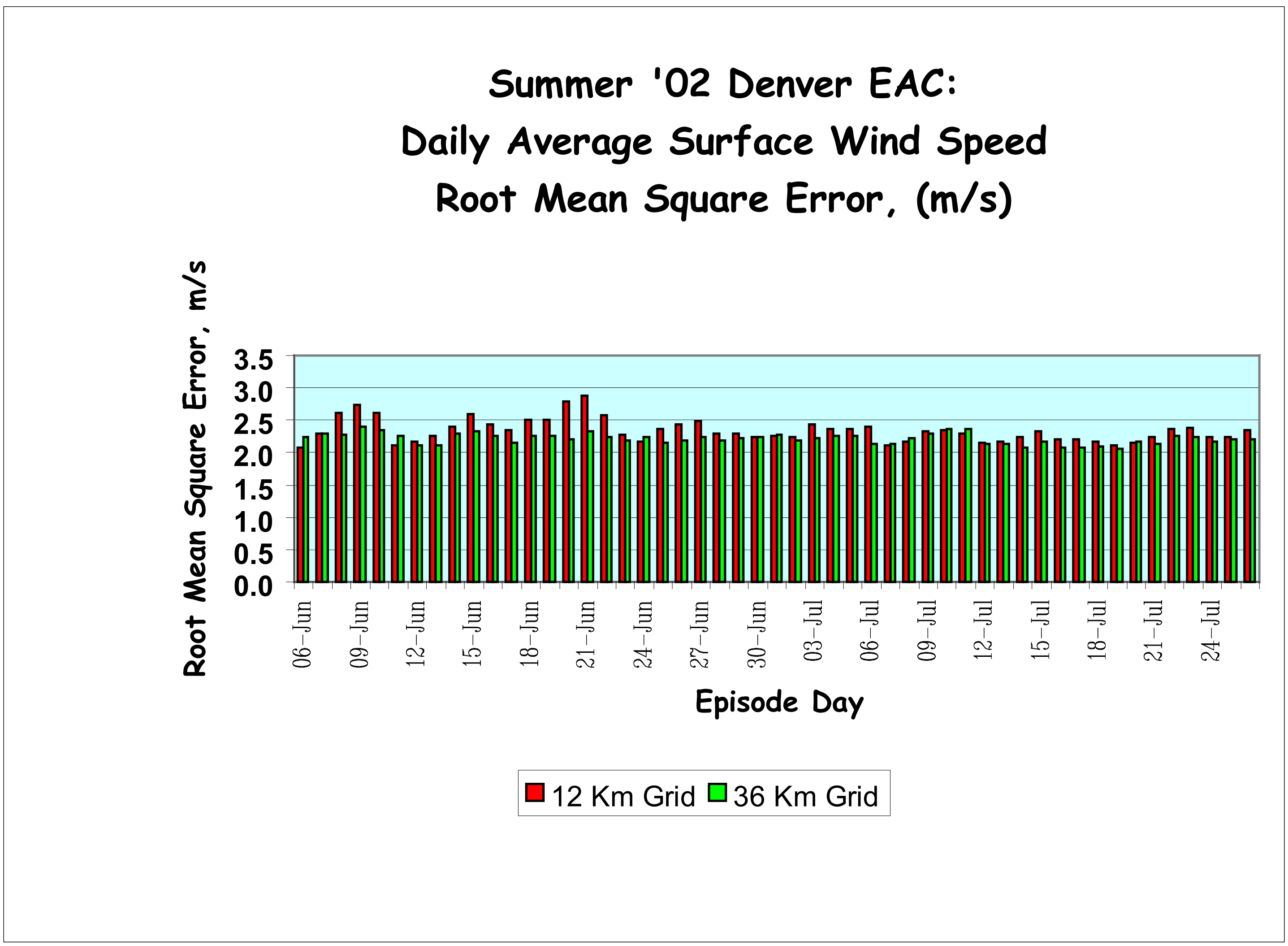


Figure 8-12. Daily Average Surface Wind Speed Root Mean Square Error (m/s) for the 6 June-25 July 2002 Summer '02 Ozone Episode.

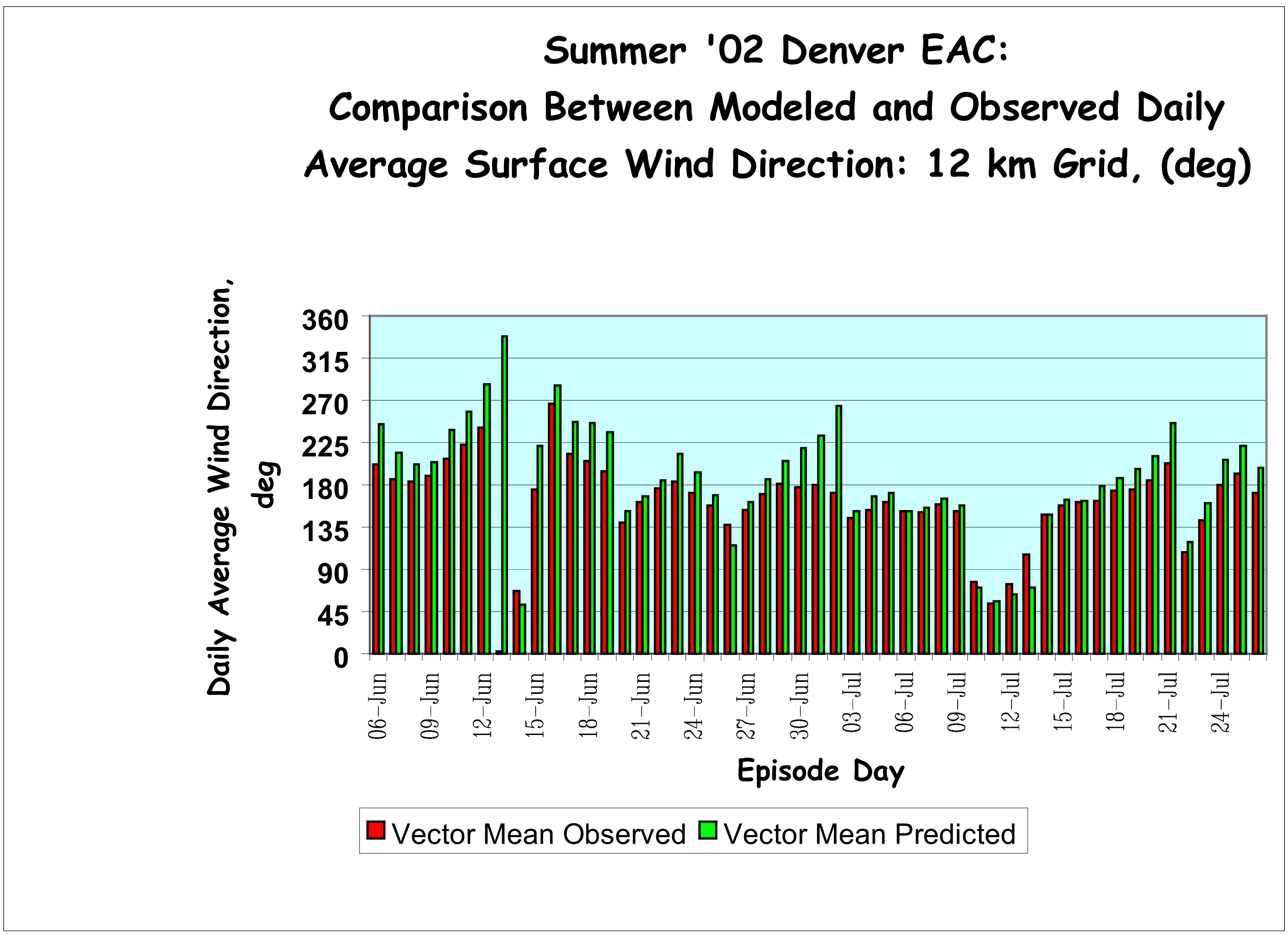


Figure 8-13. Daily Average Modeled and Observed Surface Wind Direction (deg) on the 12 Km Grid for the 6 June-25 July 2002 Summer '02 Ozone Episode.

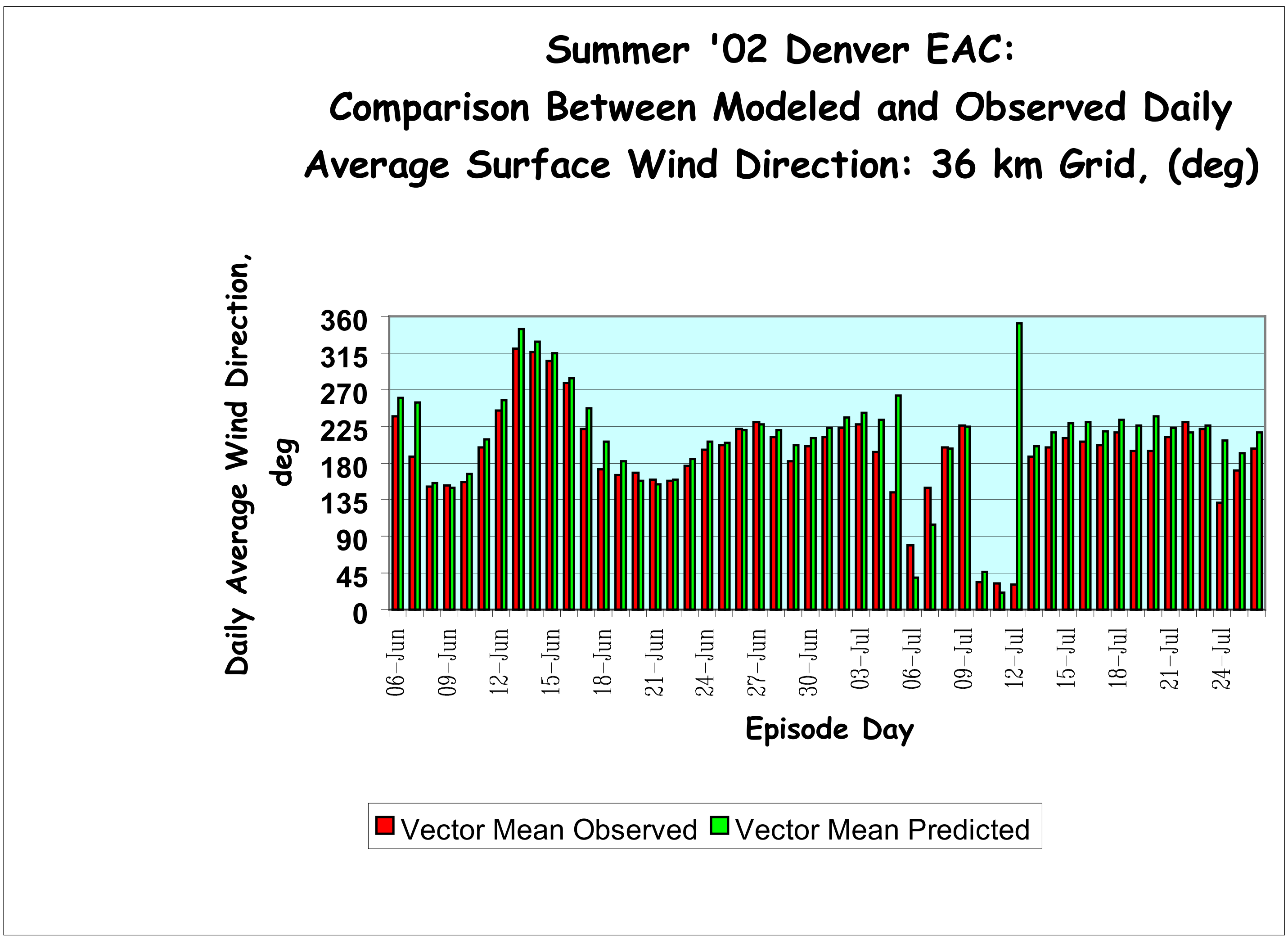


Figure 8-14. Daily Average Modeled and Observed Surface Wind Direction (deg) on the 36 Km Grid for the 6 June-25 July 2002 Summer '02 Ozone Episode.

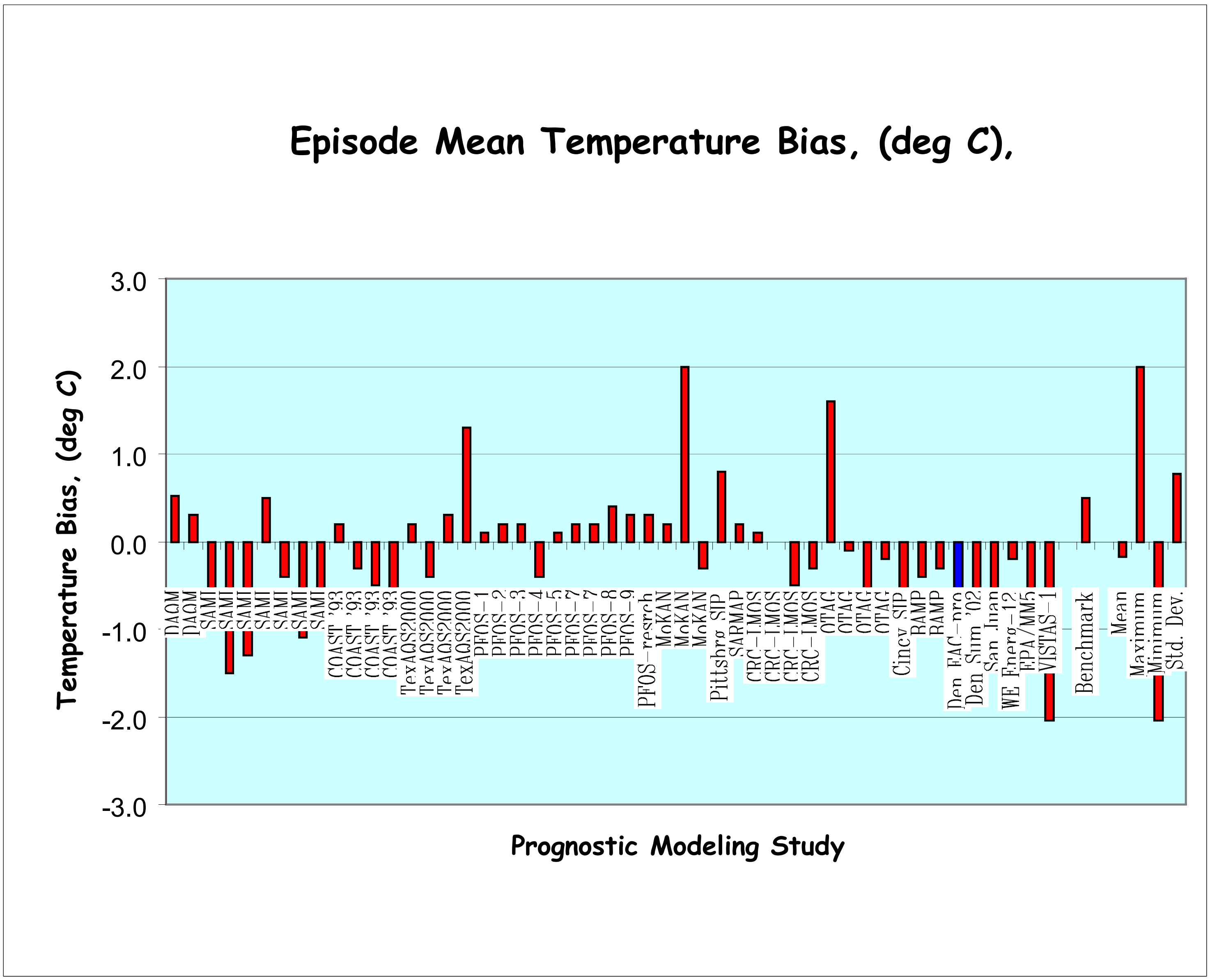


Figure 8-15. Episode Mean Temperature Bias From 50 Prognostic Model Evaluations in the U.S. Since 1995.

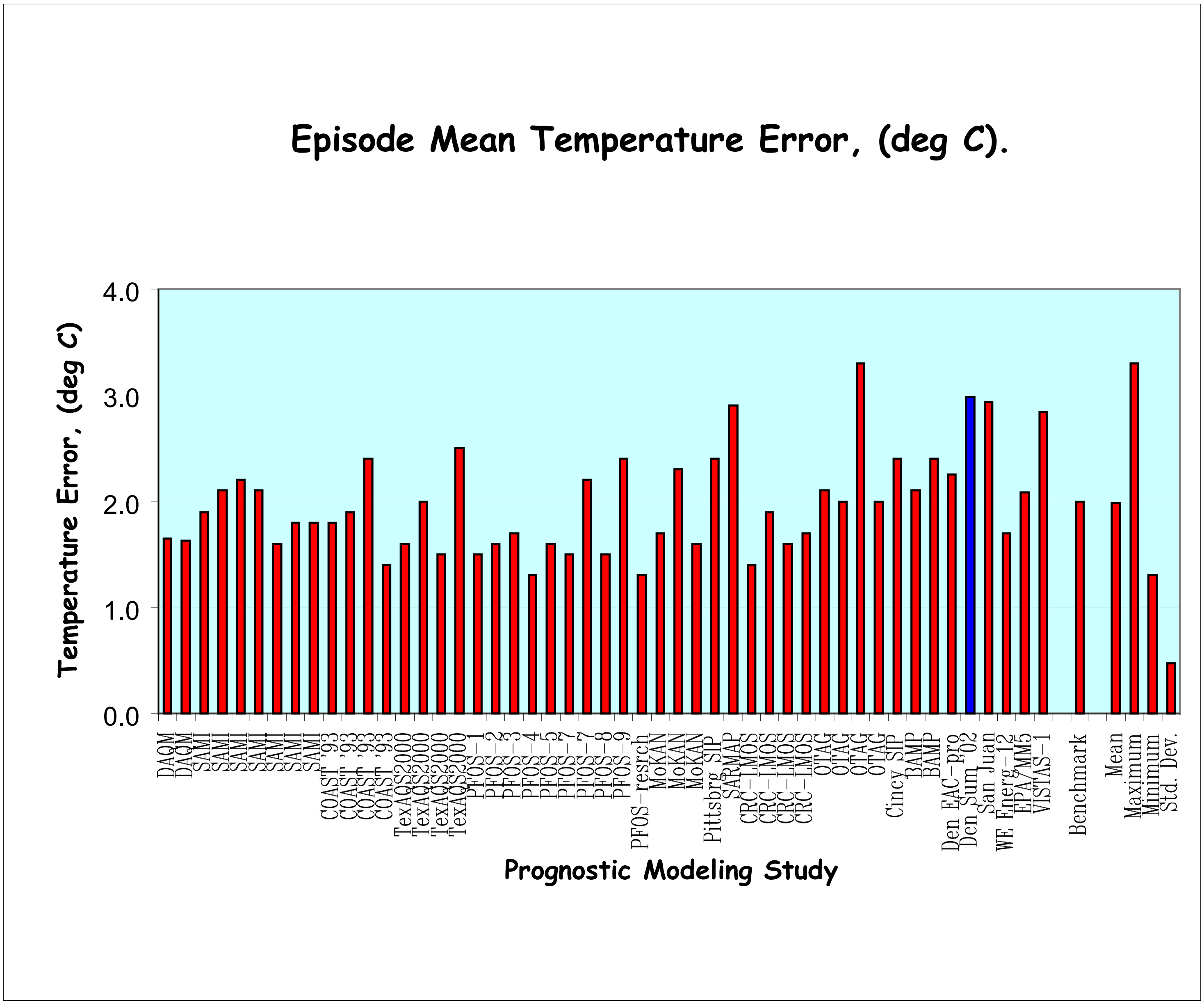


Figure 8-16. Episode Mean Temperature Gross Errors From 50 Prognostic Model Evaluations in the U.S. Since 1995.

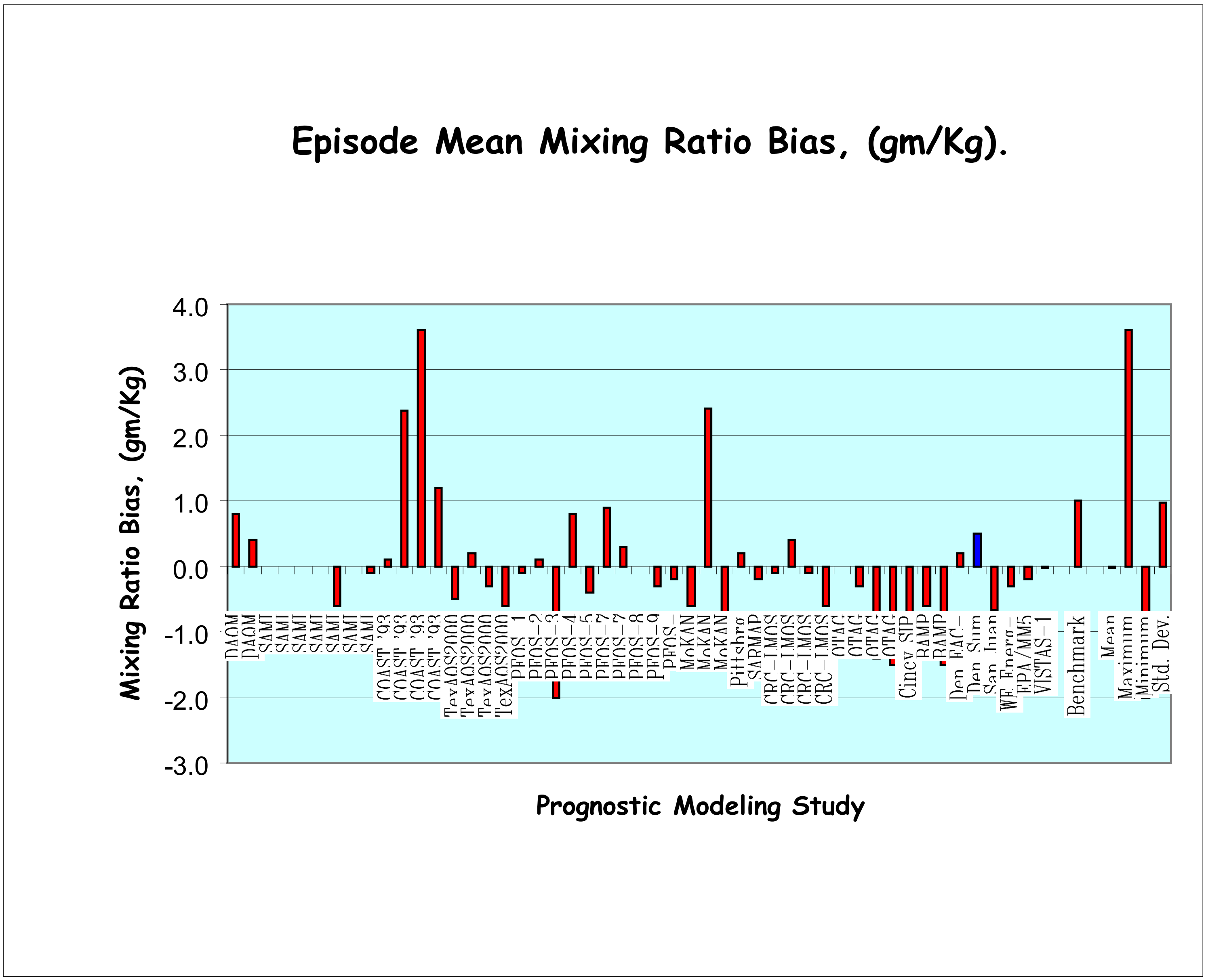


Figure 8-17. Episode Mean Mixing Ratio Bias From 50 Prognostic Model Evaluations in the U.S. Since 1995.

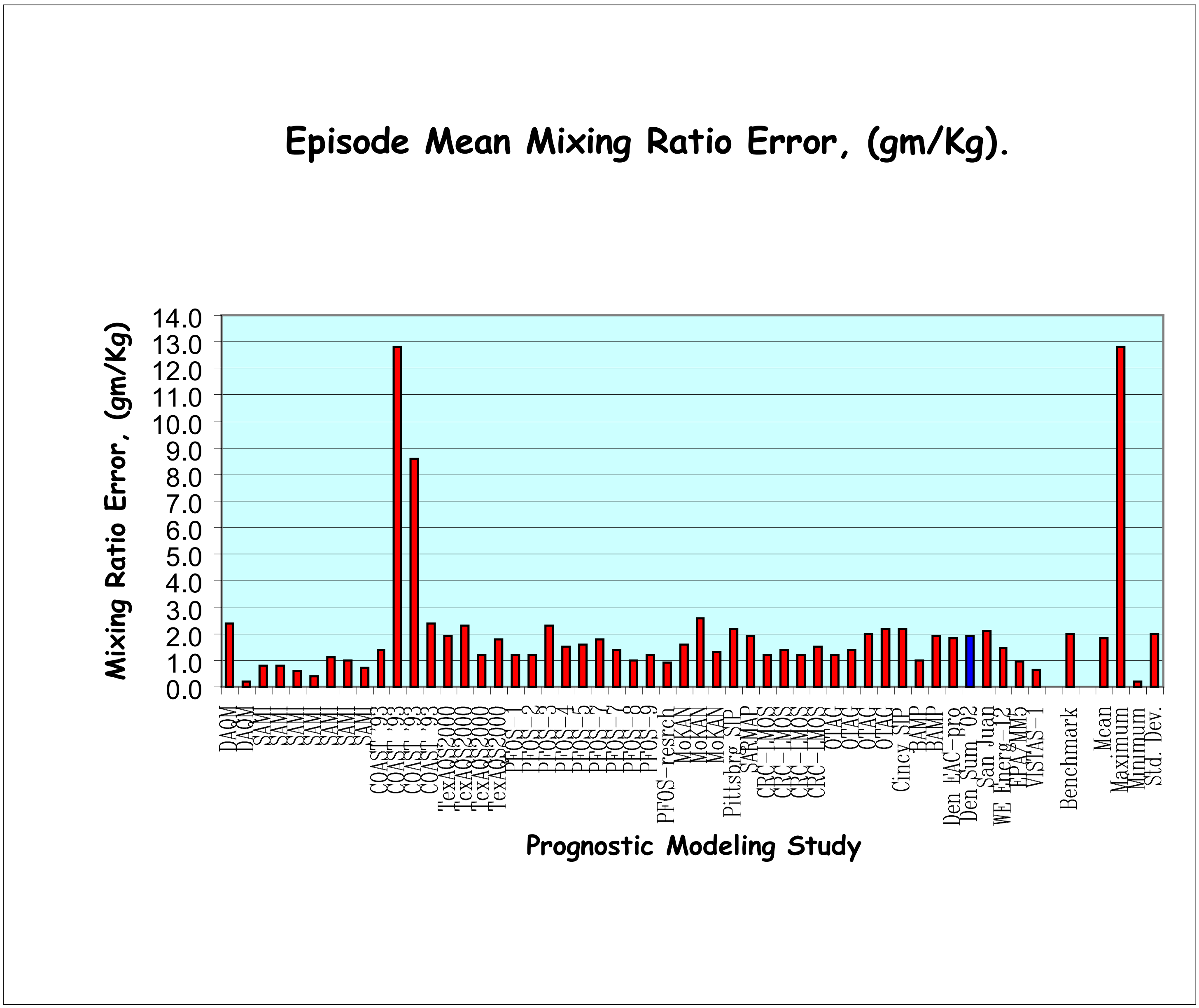


Figure 8-18. Episode Mean Mixing Ratio Gross Error From 50 Prognostic Model Evaluations in the U.S. Since 1995.

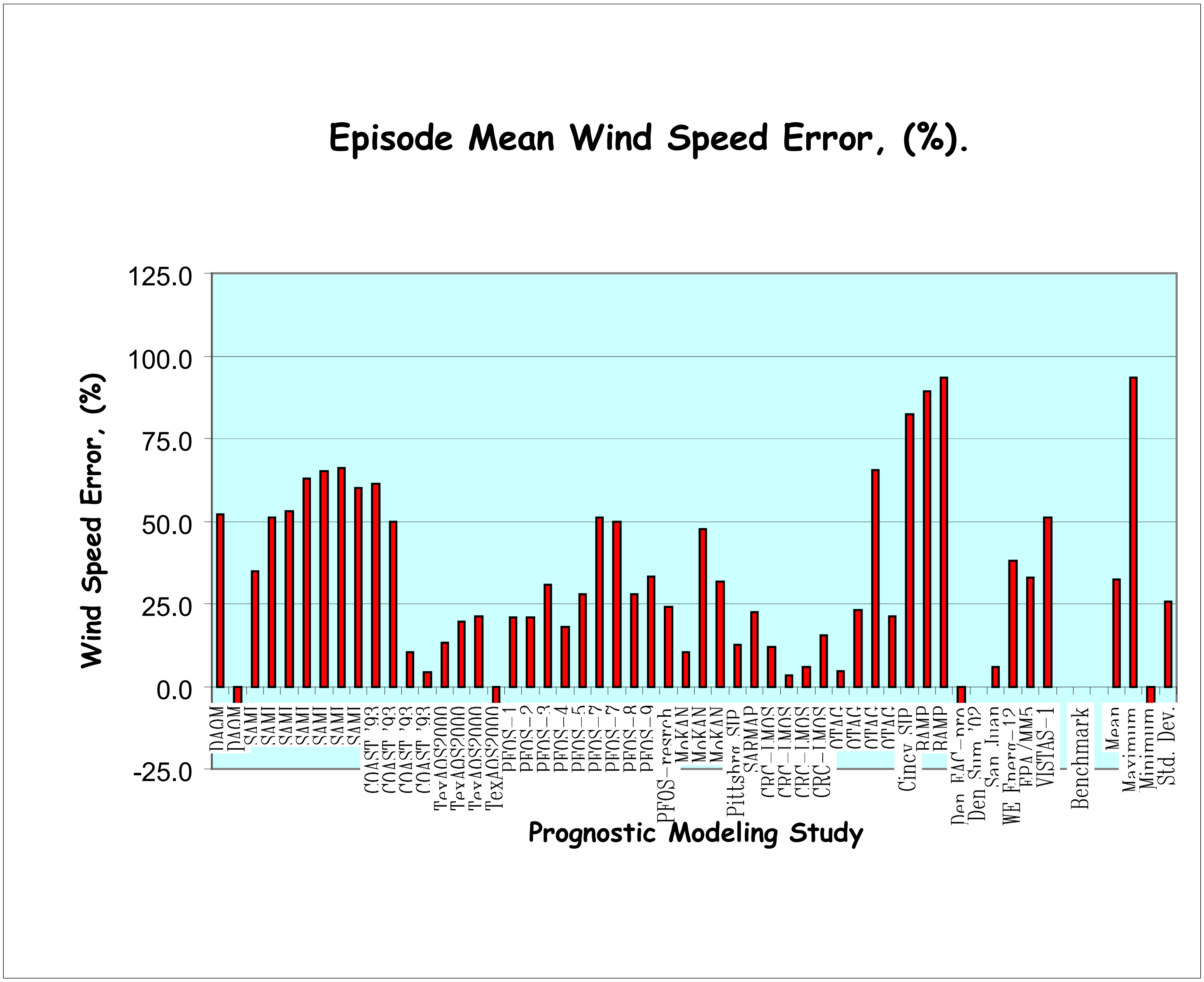


Figure 8-19. Episode Mean Error in Surface Wind Speed (%) From 50 Prognostic Model Evaluations in the U.S. Since 1995.

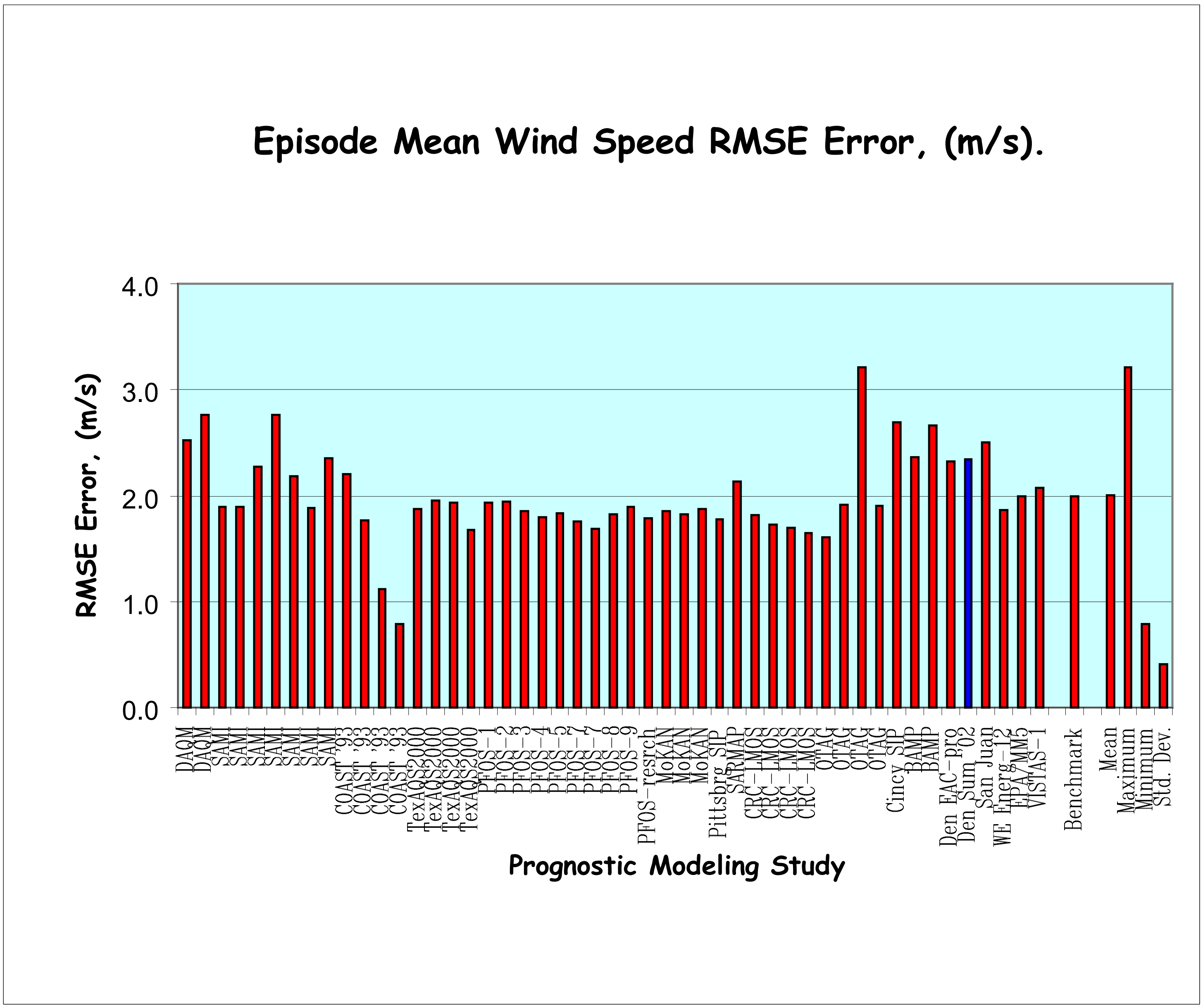


Figure 8-20. Episode Mean Root Mean Square Error in Surface Wind Speed (m/s)From 50 Prognostic Model Evaluations in the U.S. Since 1995.

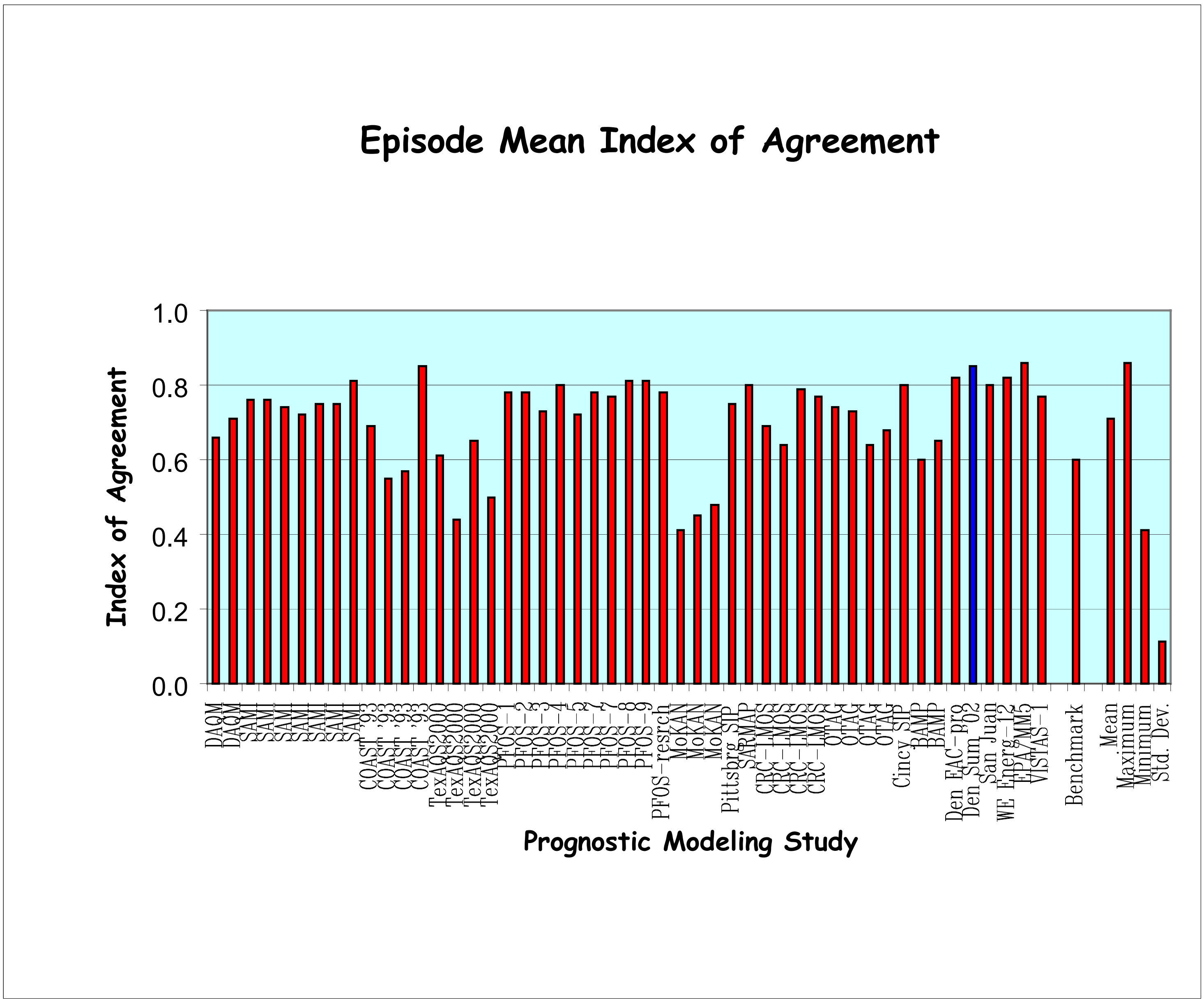


Figure 8-21. Episode Mean Index of Agreement in Surface Wind Speed From 50 Prognostic Model Evaluations in the U.S. Since 1995.

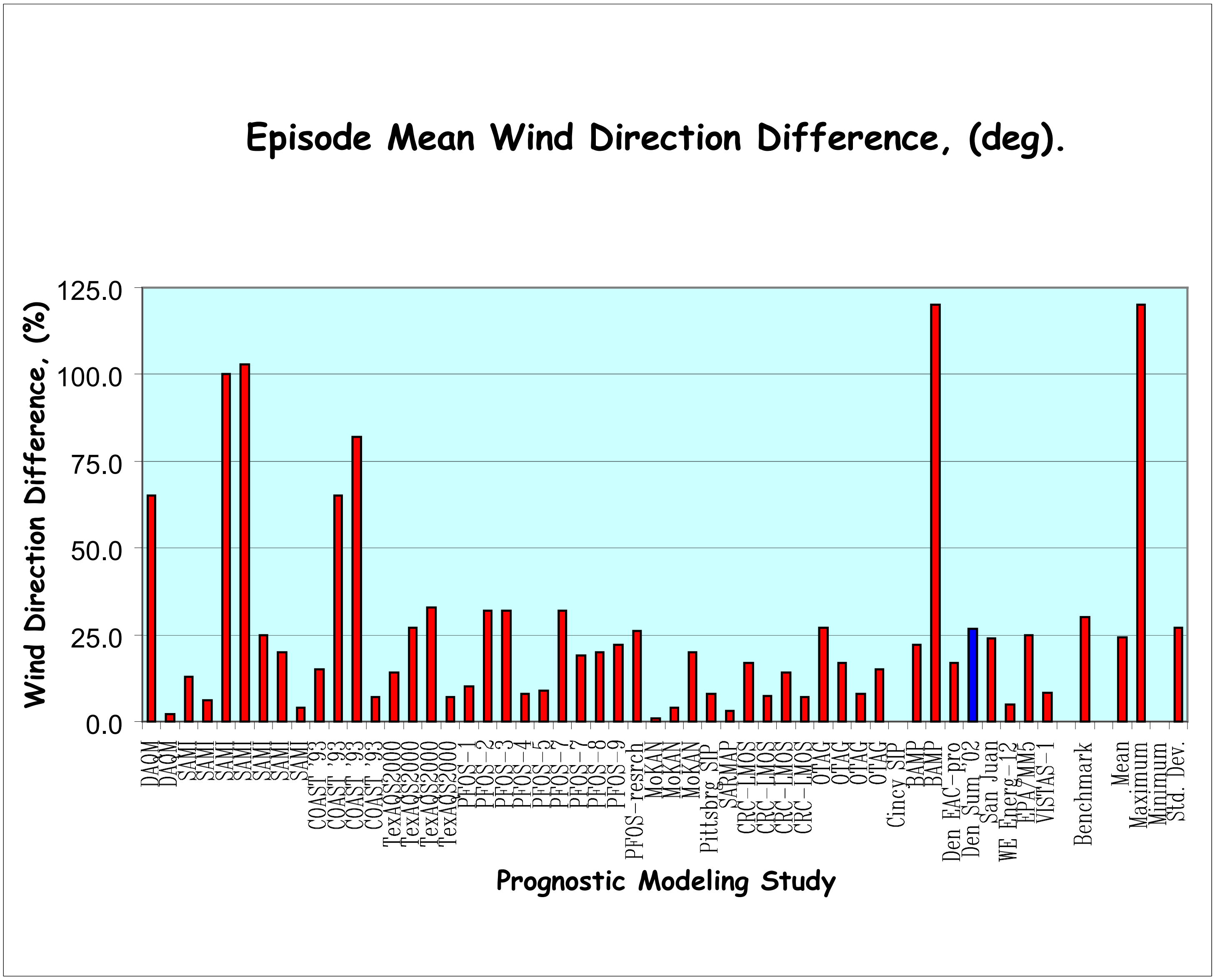


Figure 8-22. Episode Mean Difference in Surface Wind Directions (deg) From 50 Prognostic Model Evaluations in the U.S. Since 1995.

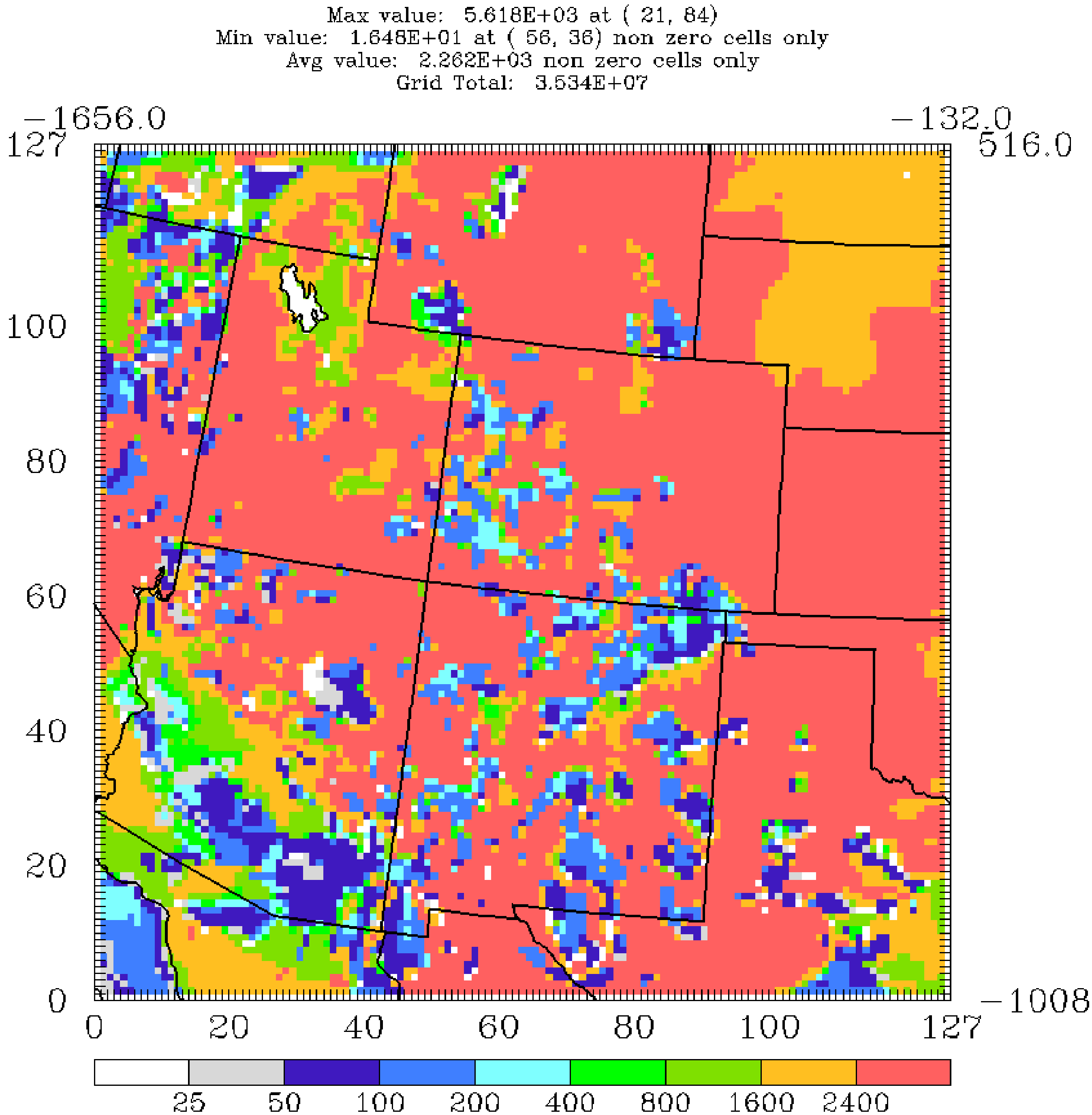


Figure 8-23. Planetary Boundary Layer Heights (m) at 1400 MDT on 14 July 2002 Over the 12 km Grid.

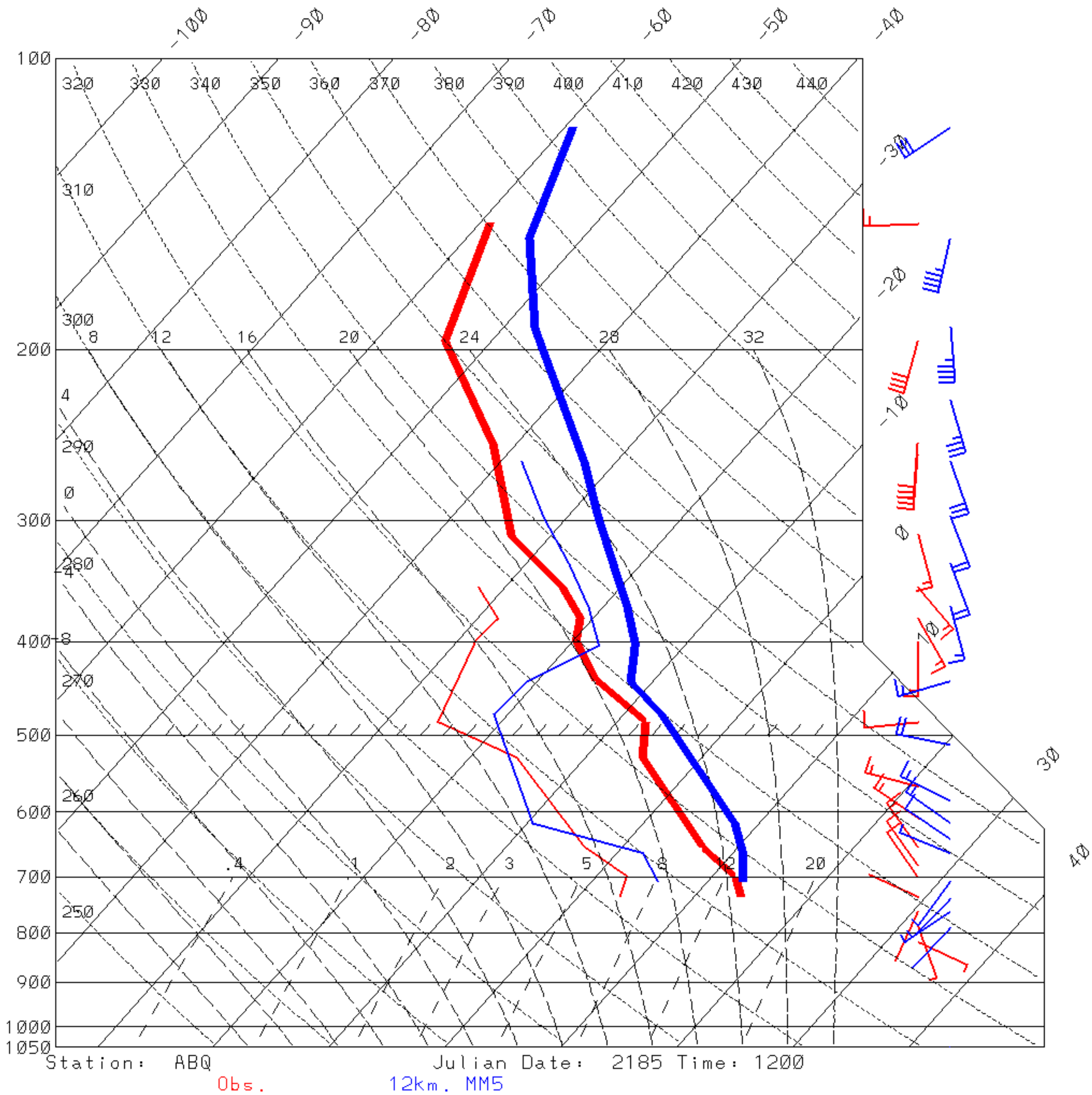


Figure 8-24. Skew-T Plot of Modeled and Observed Aloft Winds, Temperatures and Mixing Ratios at Albuquerque at 1585 MDT on 4 July 2002--12 km Grid.

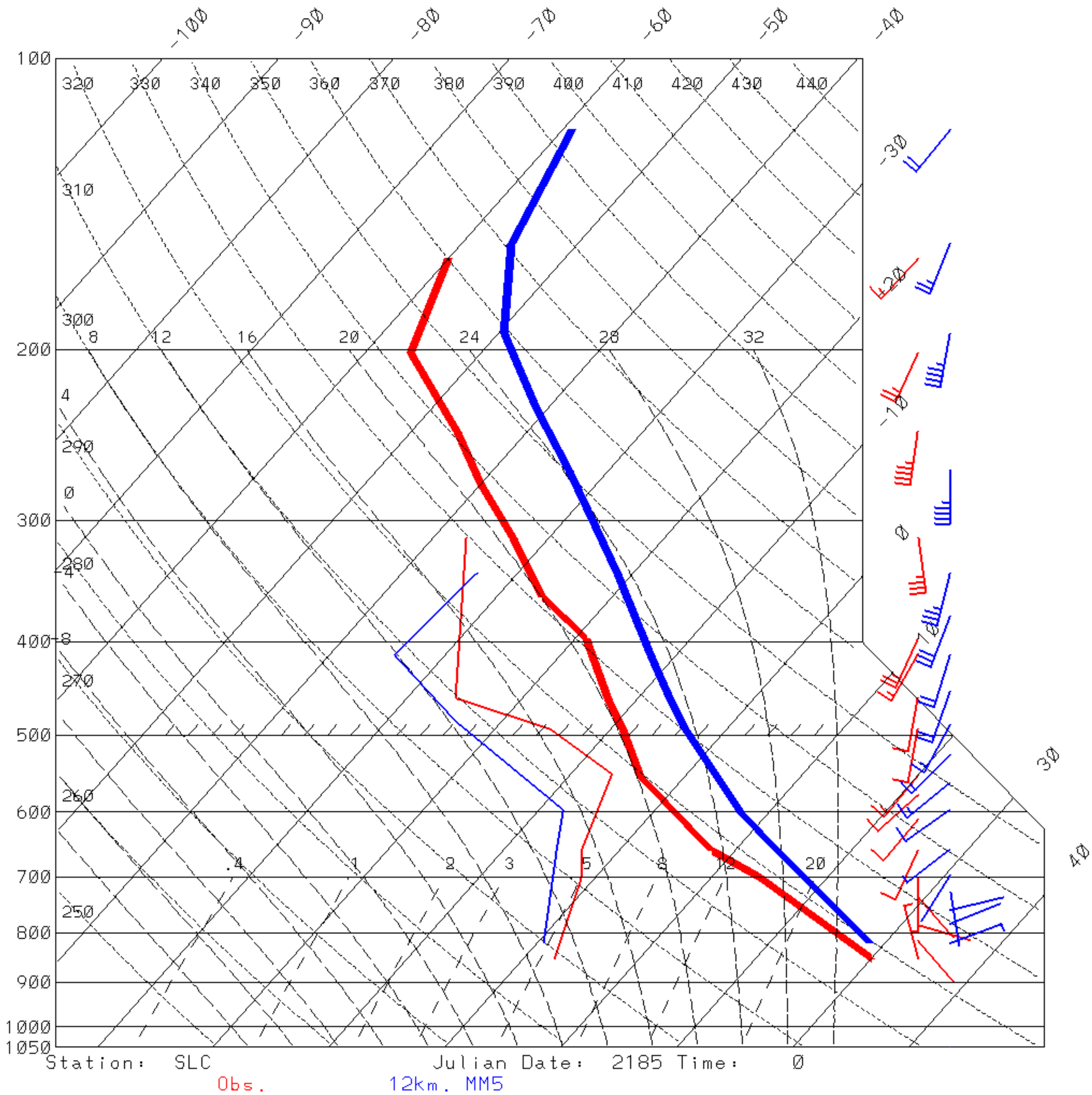


Figure 8-25. Skew-T Plot of Modeled and Observed Aloft Winds, Temperatures and Mixing Ratios at Salt Lake City at 1585 MDT on 4 July 2002--12 km Grid.

Table 8-1. Summary of Prognostic Meteorological Model Evaluations by Alpine Geophysics Since 1995.

No	Study	Domain	Model	Ref	Episode	Temp, (deg C)		MixR, (gm/Kg)		Surface Winds (m/s)			
						Bias	Error	Bias	Error	Error	RMSE	Indx A	WDir Dif
1	DAQM	Rocky Mtns	MM5	13	12-20 Jan '97	0.5	1.7	0.8	2.4	52.2	2.5	0.66	65
2	DAQM	Rocky Mtns	MM5	13	28-30 Dec '87	0.3	1.6	0.4	0.2	-5.2	2.8	0.71	2
3	SAMI	SE U.S.	RAMS	7	24-29 May '95	-1.0	1.9	0.0	0.8	35.0	1.9	0.76	13
4	SAMI	SE U.S.	RAMS	7	11-17 May '93	-1.5	2.1	0.0	0.8	51.0	1.9	0.76	6
5	SAMI	SE U.S.	RAMS	7	23-31 Mar '93	-1.3	2.2	0.0	0.6	53.0	2.3	0.74	100
6	SAMI	SE U.S.	RAMS	7	8-13 Feb '94	0.5	2.1	0.0	0.4	63.0	2.8	0.72	103
7	SAMI	SE U.S.	RAMS	7	3-12 Aug '93	-0.4	1.6	-0.6	1.1	65.0	2.2	0.75	25
8	SAMI	SE U.S.	RAMS	7	22-29 Jun '92	-1.1	1.8	0.0	1.0	66.0	1.9	0.75	20
9	SAMI	SE U.S.	RAMS	7	24Ap-3My '91	-0.8	1.8	-0.1	0.7	60.0	2.4	0.81	4
10	COAST '93	Cent. U.S.	MM5	11	4-11 Sept '93	0.2	1.8	0.1	1.4	61.4	2.2	0.69	15
11	COAST '93	Cent. U.S.	MM5	12	6-11 Sept '93	-0.3	1.9	2.4	12.8	50.0	1.8	0.55	65
12	COAST '93	Cent. U.S.	RAMS	12	6-11 Sept '93	-0.5	2.4	3.6	8.6	10.2	1.1	0.57	82
13	COAST '93	Cent. U.S.	SAIMM	12	6-11 Sept '93	-0.6	1.4	1.2	2.4	4.2	0.8	0.85	7
14	TexAQS2000	Cent. U. S.	MM5-T	12	25Aug-1 Sep '00	0.2	1.6	-0.5	1.9	13.2	1.9	0.61	14
15	TexAQS2000	Cent. U. S.	MM5-M	12	25Aug-1 Sep '00	-0.4	2.0	0.2	2.3	19.5	2.0	0.44	27
16	TexAQS2000	Cent. U. S.	MM5-NG	14	25Aug-1 Sep '00	0.3	1.5	-0.3	1.2	21.2	1.9	0.65	33
17	TexAQS2000	Cent. U. S.	RAMS-PNL	14	28ug-1 Sep '00	1.3	2.5	-0.6	1.8	-6.0	1.7	0.50	7
18	PFOS-1	SE U.S.	MM5	10	16-24 Apr '99	0.1	1.5	-0.1	1.2	20.9	1.9	0.78	10
19	PFOS-2	SE U.S.	MM5	10	2-10 May '97	0.2	1.6	0.1	1.2	21.0	2.0	0.78	32
20	PFOS-3	SE U.S.	MM5	10	25-30 Aug '97	0.2	1.7	-2.0	2.3	30.6	1.9	0.73	32
21	PFOS-4	SE U.S.	MM5	10	4-10 Apr '99	-0.4	1.3	0.8	1.5	18.1	1.8	0.80	8
22	PFOS-5	SE U.S.	MM5	10	17-23 Sep '97	0.1	1.6	-0.4	1.6	27.9	1.8	0.72	9
23	PFOS-7	SE U.S.	MM5	10	25-28 Aug '98	0.2	1.5	0.9	1.8	51.2	1.8	0.78	32
24	PFOS-7	SE U.S.	MM5	10	8-10 May '99	0.2	2.2	0.3	1.4	49.8	1.7	0.77	19
25	PFOS-8	SE U.S.	MM5	10	20-28 Apr '98	0.4	1.5	0.0	1.0	27.9	1.8	0.81	20
26	PFOS-9	SE U.S.	MM5	10	26Jul-1Aug '99	0.3	2.4	-0.3	1.2	33.2	1.9	0.81	22
27	PFOS-resrch	SE U.S.	MM5	10	18-24 Apr '98	0.3	1.3	-0.2	0.9	24.0	1.8	0.78	26
28	MoKAN	Midwest U.S.	MM5	8	8-15 Jul '95	0.2	1.7	-0.6	1.6	10.3	1.9	0.41	1
29	MoKAN	Midwest U.S.	MM5	8	14-21 Aug '98	2.0	2.3	2.4	2.6	47.5	1.8	0.45	4
30	MoKAN	Midwest U.S.	MM5	8	11-24 Jun '95	-0.3	1.6	-0.9	1.3	31.6	1.9	0.48	20
31	Pittsbrg SIP	East U.S.	MM5	1	31Jul-2 Aug '95	0.8	2.4	0.2	2.2	12.6	1.8	0.75	8
32	SARMAP	West U.S.	MM5	4	3-6 Aug '90	0.2	2.9	-0.2	1.9	22.6	2.1	0.80	3

33	CRC-LMOS	Midwest U.S.	RAMS	6	26-28 June '91	**0.1**	**1.4**	**-0.1**	**1.2**	**11.9**	**1.8**	**0.69**	**17**
34	CRC-LMOS	Midwest U.S.	RAMS	6	17-19 Jul '91	**0.0**	**1.9**	**0.4**	**1.4**	**3.5**	**1.7**	**0.64**	**7**
35	CRC-LMOS	Midwest U.S.	MM5	6	26-28 Jul '91	**-0.5**	**1.6**	**-0.1**	**1.2**	**5.8**	**1.7**	**0.79**	**14**
36	CRC-LMOS	Midwest U.S.	MM5	6	17-19 Jun '91	**-0.3**	**1.7**	**-0.6**	**1.5**	**15.6**	**1.7**	**0.77**	**7**
37	OTAG	East U.S.	RAMS	3	13-21 Jul '91	**1.6**	**2.1**	**0.0**	**1.2**	**4.6**	**1.6**	**0.74**	**27**
38	OTAG	East U.S.	MM5	3	13-21 Jul '91	**-0.1**	**2.0**	**-0.3**	**1.4**	**23.0**	**1.9**	**0.73**	**17**
39	OTAG	East U.S.	MM5	2	1-11 Jul '88	**-0.6**	**3.3**	**-1.4**	**2.0**	**65.6**	**3.2**	**0.64**	**8**
40	OTAG	East U.S.	MM5	1	12-15 Jul '95	**-0.2**	**2.0**	**-1.5**	**2.2**	**21.2**	**1.9**	**0.68**	**15**
41	Cincy SIP	Midwest U.S.	MM5	5	18-22 Jun '94	**-0.7**	**2.4**	**-1.6**	**2.2**	**82.4**	**2.7**	**0.80**	**0**
42	BAMP	SE U.S.	MM5	9	6-11 Sept '93	**-0.4**	**2.1**	**-0.6**	**1.0**	**89.4**	**2.4**	**0.60**	**22**
43	BAMP	SE U.S.	MM5	9	15-19 Aug '93	**-0.3**	**2.4**	**-1.5**	**1.9**	**93.6**	**2.7**	**0.65**	**120**
44	Den EAC-pro	Western U.S.	MM5	15	15-21 Jul '02	**-1.1**	**2.3**	**0.2**	**1.8**	**-10.4**	**2.3**	**0.82**	**17**
45	Den Sum '02	Western U.S.	MM5-12 km	18	6 Jun-25 Jul '02	**-1.9**	**3.0**	**0.5**	**1.9**	**0.0**	**2.3**	**0.85**	**27**
46	San Juan pro	Western U.S.	MM5	15	30 Jul-5 Aug '00	**-1.0**	**2.9**	**-0.7**	**2.1**	**6.0**	**2.5**	**0.80**	**24**
47	WE Energ-12	Midwest U.S.	MM5	17	16 Jun-14 Aug '01	**-0.2**	**1.7**	**-0.3**	**1.5**	**37.9**	**1.9**	**0.82**	**5**
48	EPA/MM5	Entire U.S.	MM5	16	1 Jan -31 Dec '01	**-0.6**	**2.1**	**-0.2**	**1.0**	**33.0**	**2.0**	**0.86**	**25**
49	VISTAS-1	Southeast US	MM5	19	2-20 Jan '02	**-2.0**	**2.8**	**0.0**	**0.6**	**51.3**	**2.1**	**0.77**	**8**
	Benchmark					**0.5**	**2.0**	**1.0**	**2.0**		**2.0**	**0.60**	**30**
	Mean					**-0.2**	**2.0**	**0.0**	**1.8**	**32.2**	**2.0**	**0.71**	**24**
	Maximum					**2.0**	**3.3**	**3.6**	**12.8**	**93.6**	**3.2**	**0.86**	**120**
	Minimum					**-2.0**	**1.3**	**-2.0**	**0.2**	**-10.4**	**0.8**	**0.41**	**0**
	Std. Dev.					**0.8**	**0.5**	**1.0**	**2.0**	**25.5**	**0.4**	**0.11**	**27**

Table 8-2. Summary Results for the 6 June–25 July Summer '02 MM5 Simulation on the 36/12 km Regional Grids Compared with the Ad Hoc Performance Benchmarks and Fifty Recent Prognostic Model Performance Evaluations Throughout the U.S.

Episode Grid Resolution	Temperature, deg C		Mixing Ratio, kg/KG		Surface Winds, (m/s)		
	Bias	Error	Bias	Error	RMSE	I	WD diff
12 km	-1.9	3.0	0.5	1.9	2.34	0.85	27
36 km	-0.8	2.1	-0.1	1.5	2.21	0.88	19
Benchmark	≤ ± 0.5	≤ 2.0	≤ ± 1.0	≤ 2.0	≤ 2.00	≥ 0.60	≤ 30
U.S. Average	-0.2	2.0	0.0	1.8	2.00	0.71	24

9.0 MM5 EVALUATION FOR THE 16-22 JULY '02 EPISODE: 4/1.33 KM GRIDS

This chapter presents results of the operational evaluation of the MM5 model for the 16-22 June 2002 ozone episode over the Denver-Northern Front Range Region. In this evaluation we focus on the 4 km and 1.33 km grid results. As indicated by the results presented in this chapter, the problem with the 36/12 km surface temperature under-predictions has been quite well resolved on the 4/1.3 km grids (to be used in the CAMx air quality modeling) by eliminating the biases present in the Pleim-Xiu land surface module scheme. Tables 9-1 thorough 9-3 present summary mixing ratio, temperature and wind performance statistics on a daily basis for the MM5 results on both the 4 km and 1.33 km grids.

The same set of surface meteorological data used previously for the 36/12 km MM5 performance testing discussed in Chapter 8 was used here and in Chapter 10. These measurements included mixing ratio, temperature, wind speed and direction.

Figures 9-1 and 9-2 present the predicted surface winds on 20 July 2002 over the 4 km (1200 MDT) and 1.33 km (0600 MDT) grid regions, respectively. The generally good correspondence between the measurements (black vectors) and observations (red vectors) is seen not only in these example plots but in the full set of hourly 4/1.33 km grid comparisons contained on the CD. Review of the 4/1.33 km surface winds for the 16-22 June 2002 episode did not reveal any unusual or obviously flawed wind predictions, behavior symptomatic of spurious model results.

9.1 Mixing Ratio

Figures 9-3 and 9-4 present the daily average gross error and bias in hourly near-surface mixing ratio for the 16-22 July 2002 episode. The figures present the 1.33 km grid results in red and the 4 km grid results in green. Recall that there is an unavoidable mis-match between the height of the mixing ratio measurement (2 m) and the height of the MM5's first grid level prediction (17.5 m) which introduces differences between the two that are quite independent of any error in the measurement or model prediction. Considering first the mixing ratio gross error estimates, Figure 9-3 shows that there is day-to-day variation in the mixing ratio errors on both nested grids with maxima occurring on 19 July 2002. In general, the errors range from 1.2 gm/Kg to 1.9 gm/Kg. On most days, the MM5 performs slightly better on the 1.33 km grid compared to the 4 km grid. The episode average gross errors in mixing ratio on the 4 km and 1.33 km grids (1.57 gm/Kg and 1.47 gm/Kg, respectively) are shown in the two far right hand bars in Figure 9-3 and in Table 9-2.

For the mixing ratio bias (Figure 9-4), the MM5 tends to underestimate the day-to-day values on both grids for most all days. The biases range from about –1.6 gm/Kg to 0.4 gm/Kg. For daily average bias, the MM5 performs about the same on both grids. The episode average biases in mixing ratio on the 1.33 km and 4 km grids (-0.59 gm/Kg and -0.66 gm/Kg, respectively) are actually quite good given the 15.5 m vertical displacement between the measurement height and height of the layer 1 MM5 prediction.

9.2 Temperatures

Figures 9-5 and 9-6 depict the daily average gross error and bias in hourly near-surface temperatures on the two nested grids. Table 9-1 summarizes the statistical results. Unlike the mixing ratios, the temperatures are by a MM5 post-processor for a height of 2 m, corresponding to the typical shelter height. For the temperature gross errors, Figure 9-5 shows typical day-to-day variation in the temperature errors on both grids with the errors ranging from 0.9 deg C to 2.7 deg C. On all days, the MM5 performs better on the 1.33 km grid compared to the 4 km grid. The episode average gross errors in temperature on the 1.33 km and 4 km grids are 1.6 deg C and 2.3 deg C, respectively. These statistical measures are notably improved over the 36/12 km temperature gross errors reported in the previous chapter.

For the daily average temperature bias (Figure 9-6), the MM5 using the Blackadar pbl scheme tends to overestimate the day-to-day values on both grids for all days except 16 July. Opposite the gross error results above, for temperature bias the MM5 appears to perform better on the 4 km grid compared to the 1.33 km grid. The daily average of the hourly temperature biases range between about –0.4 deg C to 1.6 deg C. The episode average biases in temperature on the 4 km and 1.33 km grids are 0.45 deg C and 0.81 deg C, respectively. Note that the temperature bias over the 1.33 km domain is, on average, roughly twice that on the 4 km grid.

The average diurnal behavior of the MM5's predictions of near surface temperature on the 4 km and 1.33 km grids are shown in Figure 9-7. There are 40 surface temperature stations within the 4 km domain while only 8 stations are within the 1.33 km high resolution grid. From Figure 9-7 it is clear that the MM5's skill in predicting the diurnal temperature profiles on both the 4 km and 1.33 km grids has been improved over the 36/12 km grids reported previously. The diurnal match in quite good on the 4 km grid and is also acceptable on the 1.33 km grid where there are fewer reporting stations. The cool afternoon temperature bias and warm nighttime temperature bias, common to most mesoscale modeling simulations is evident at both grid scales.

The daily average peak temperature prediction accuracy over the 4 km and 1.33 km grids is shown in Figure 9-8. On the 4 km grid, the accuracies of peak temperature prediction range between roughly 4%-10% while for the 1.33 km domain the range is about 2%-7%. Overall, the average peak afternoon temperature prediction accuracy on the 4 k and 1.33 km grids are 5.7% and 3.3%, respectively. These accuracy figures are actually quite good for the fine spatial scales modeled here.

9.3 Wind Speed and Direction

Figures 9-9 through 9-14 present daily average wind performance statistics on the 4 km and 1.33 km grids. Beginning with Figure 9-9, the daily average modeled and observed surface wind speeds on the 4 km grid are plotted as a function of time during the episode. The mean predicted wind speeds (1.48 m/s) are nearly equal to the observations (1.57 m/s) when averaged across all days in the 16-22 July 2002 episode. For the 1.33 km winds (Figure 9-10), the average modeled and observed surface wind speeds are 2.63 m/s and 2.76 m/s, respectively. Note that the wind speeds on the 4 km grid are roughly 40% slower than the 1.33 km grid.

Figure 9-11 depicts the daily average index of agreement on the 4 km and 1.33 km grids. On the 4 km grid, the index ranges between about 0.68 to 0.83 across the 7 modeling days. The 1.33 km grid results are poorer but still ranging from about 0.51 to 0.63. Across the entire episode, the index of agreement for the 4 km and 1.33 km grids are 0.78 and 0.57, respectively. Given the fewer number of surface wind stations on the 1.33 km grid (8 compared to 40), this trend in poorer I scores with increasing finer grid meshes is expected; indeed it is proven in virtually all of the prognostic multi-scale model simulations we have evaluated (Tesche et al., 2003c,d).

Daily average root mean square errors (RMSE) vary between 2.45 m/s to 2.81 m/s on the 4 km grid (Figure 9-12) while they are somewhat better on the 1.33 km grid, ranging between 1.95 m/s to about 3.4 m/s (see Table 9-3). Across the entire episode, the RMSE errors the 4 km and 1.33 km grids are 2.61 m/s and 2.53 m/s, respectively.

Figures 9-13 and 9-14 present comparisons between the daily average predicted and observed wind direction for the two MM5 grids. On the 4 km grid (Figure 9-13), the wind directions match reasonably well on most every day although there is obviously larger variability on 20 July, a day when wind speeds were quite low (see Figure 9-9). Somewhat surprisingly, better agreement is seen in the 1.33 km winds (Figure 9-14). Across the entire episode, the average wind direction discrepancies on the 1.33 km grid is 37 deg, while for the 4 km grid the mean difference is 60 deg. Given the very wide range of topographic conditions encountered over the 4 km and 1.33 km grid scales, this wind direction agreement is considered good.

We briefly examined the MM5's performance aloft by reviewing the planetary boundary layer height (PBL) fields for each hour of the 16-22 July 2002 episode and by studying the skew-T plots for the sounding locations within the 4 km and 1.33 km domains. As before, the pbl evaluation was qualitative since boundary layer height measurements are not routinely made. Figure 9-15 presents a typical pbl height field at 1400 MDT on 20 July 2002 over the 4 km domain. This plot shows the high afternoon boundary layer heights (2400m or more above ground level) over eastern Colorado and Wyoming and southwestern Nebraska and lower depths (100m-800m) over the mountainous terrain of the Front Range and the San Juan Mountains to the south. The full set of fields contained on the CD archive were examined briefly for consistency. No unusual or erratic behavior was observed. We also examined the MM5's performance aloft by comparing the modeled and observed upper level skew-T plots containing horizontal winds, temperatures, and mixing ratios for at the standard 0000 and 1200 UTC sounding times each day at Denver for the 1.33 km grid and at Denver and Grand Junction on the 4 km grid domain. These data sets have been archived and are available to parties interested in conducting further analyses. The skew-T results were similar to those presented in Figures 8-24 and 8-25. These brief aloft performance reviews on the 4 km and 1.33 km grids did not reveal any causes for concern with the MM5 predictions aloft as they might adversely affect the photochemical model applications.

9.4 Comparisons with Other Studies

Tables 9-1 through 9-3 summarize the episode composite temperature, wind speed, wind direction and mixing ratio statistics for the 16-22 July 2002 episode. Below, we compare the MM5 results on the 4 km grid with the much broader set of prognostic model applications studies listed in Table 8-1. Note the statistics for most of the studies reported in Table 8-1 were derived from model applications at scales between 4 km and 12 km so the current MM5 results on the 4 km Denver grid are from the lower bound of this range.

The MM5 average bias in hourly ground level temperatures over the 4 km grid is 0.45 deg C. The average across all studies is –0.2 deg C (Figure 8-15). Thus, the current simulation tends to have a systematic warm bias relative to the other studies. The source of this bias is the use of the Blackadar pbl scheme which has replaced the PX land use module that was employed on the 36 km and 12 km grid meshes. The episode average gross error in hourly ground level temperatures for the 4 km MM5 simulation is 2.3 deg C. The mean temperature error over all studies is 2.0 deg C (Figure 8-16). Thus, the current simulation on the 4 km grid tends to have somewhat higher gross errors in ground level hourly temperature predictions compared to other studies.

The MM5 average bias in hourly ground level mixing ratios on the 4 km grid is –0.66 gm/Kg. The average across all studies is 0.0 gm/Kg (Figure 8-17). This negative moisture bias relative to the other studies is in all likelihood inconsequential. The episode average error in hourly ground level mixing ratio for the current 4 km MM5 run is 1.57 gm/Kg. The mean error over all studies is 1.8 gm/Kg (Figure 8-18). As with the temperatures, the Denver 4 km MM5 simulation tends to have comparable errors in ground level mixing ratios compared to other studies. The episode average error in hourly ground level wind speed on the 4 km domain is -6%. The average across all studies is 32% (Figure 8-19). Thus, the present simulation tends to have much lower wind speed errors on the 4 km grid, on average, compared to other studies. The episode average root mean square error (RMSE) in hourly ground level wind speed prediction is 2.61 m/s. The average RMSE error over all studies is 2.00 m/s (Figure 8-20). Unlike the surface wind speed errors, the Denver EAC 4 km grid simulation tend to have somewhat higher RMSE wind speed errors, on average, compared to other studies. The episode average IOA for hourly ground level wind speed prediction is 0.78. The average index of agreement over all studies is 0.71 (Figure 8-21). On the 4 km grid, the MM5 tends to yield a better a statistical score compared to other studies. The episode average discrepancy in hourly wind direction for the Denver EAC 4 km simulation is 60 deg C which is more than twice the size of the average wind direction error (24 deg C) from the other studies (Figure 8-22). Thus, on the 4 km grid, the current MM5 simulation tends to have significantly larger wind direction errors compared to other studies.

9.5 Assessment of the 16-22 July 2002 Episode

Table 9-4 compares the 4 km and 1.33 km MM5 results for the 16-22 July 2002 episode with the proposed meteorological modeling benchmarks. Shaded cells in the table correspond to those meteorological variables that fall just outside of the benchmark ranges. On the 4 km grid for the 16-22 July 2002 episode the gross error in surface temperature prediction, the RMSE error in surface wind speed prediction, and the mean wind direction prediction difference all fall outside the suggested model performance benchmarks and the average results for model applications at scales ranging from 4 km to 12 km. The remaining four statistical measures are within the suggested performance ranges. The results on the 1.33 km grid are only slightly poorer when compared with the benchmarks and this is attributed in part to the difference in scales. However, when considering the full set of model performance results on the 4/1.33 km grids, particularly in light of substantial challenges posed by simulating such fine scales over the Denver-Northern Front Range Region, we conclude that the 4/1.33 km MM5 meteorological fields for the 16-22 July 2002 episode may be used, with reasonable caution, as input to the regional emissions and photochemical models for air quality impacts assessments for the Denver EAC study.

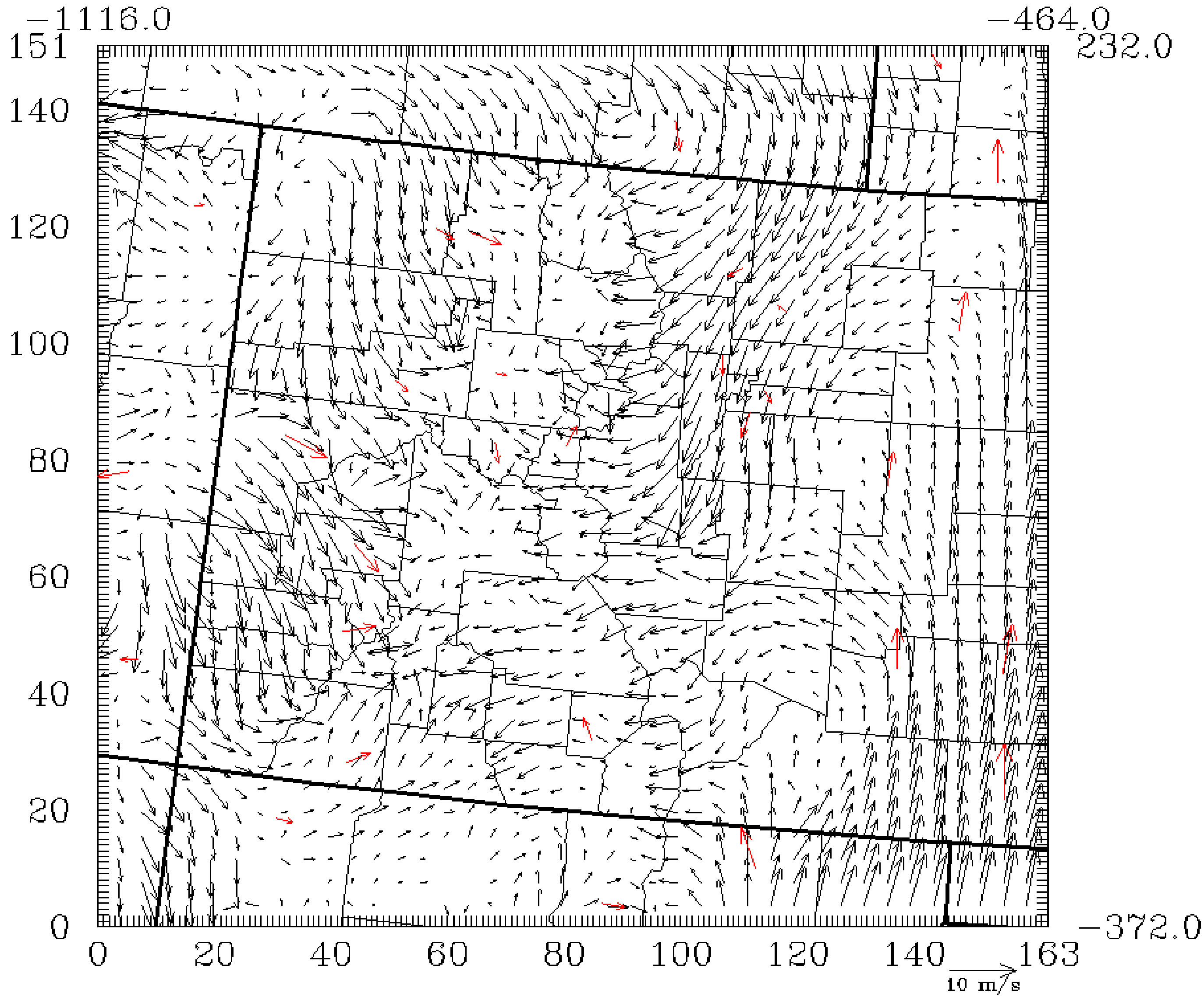


Figure 9-1. MM5 Surface Wind Fields at 1200 MDT on 20 July 2002 Over the 4 km Domain. (Predicted Vectors Plotted Every Fourth Grid Cell).

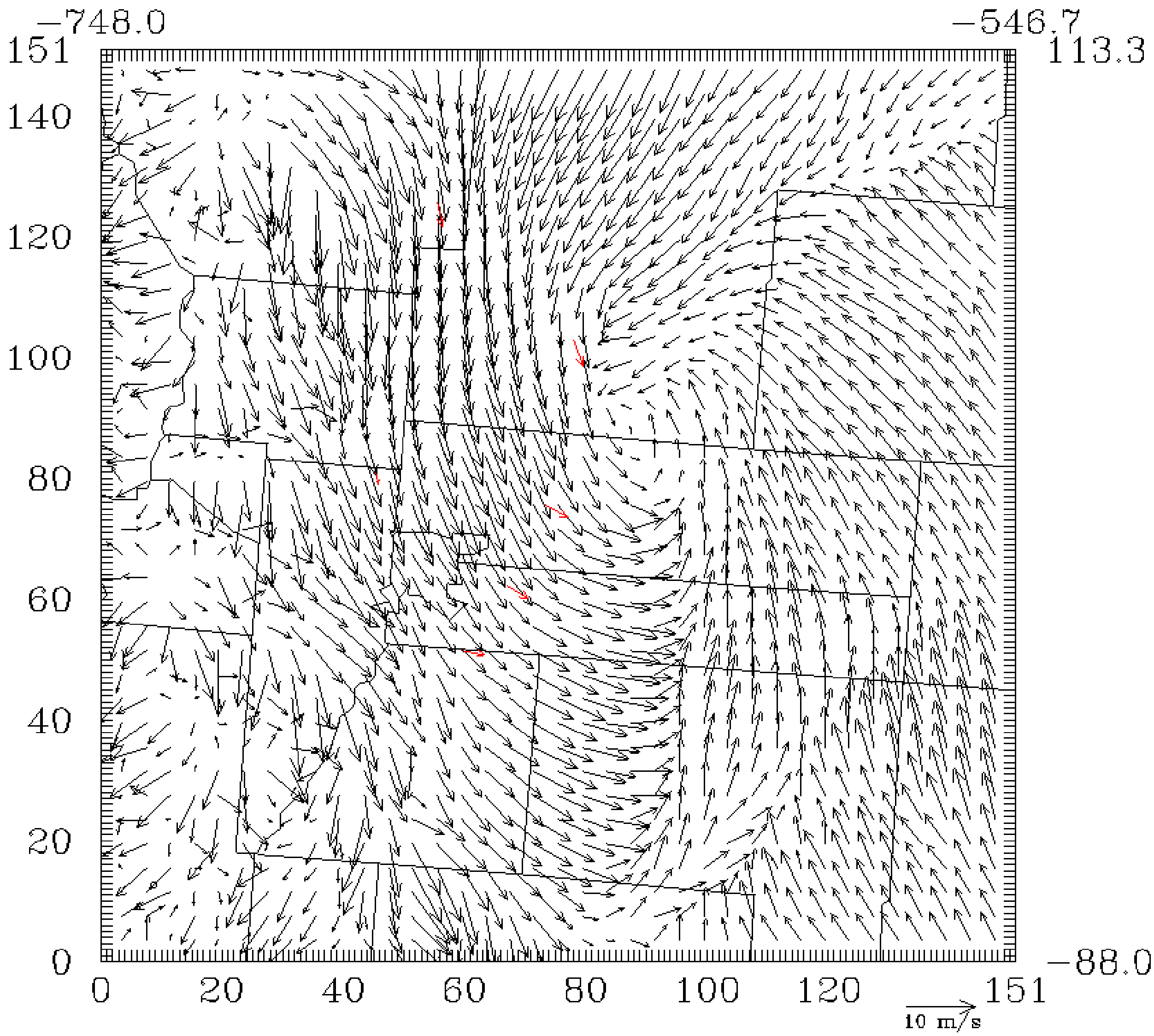


Figure 9-2. MM5 Surface Wind Fields at 0600 MDT on 20 July 2002 Over the 1.33 km Domain. (Predicted Vectors Plotted Every Fourth Grid Cell).

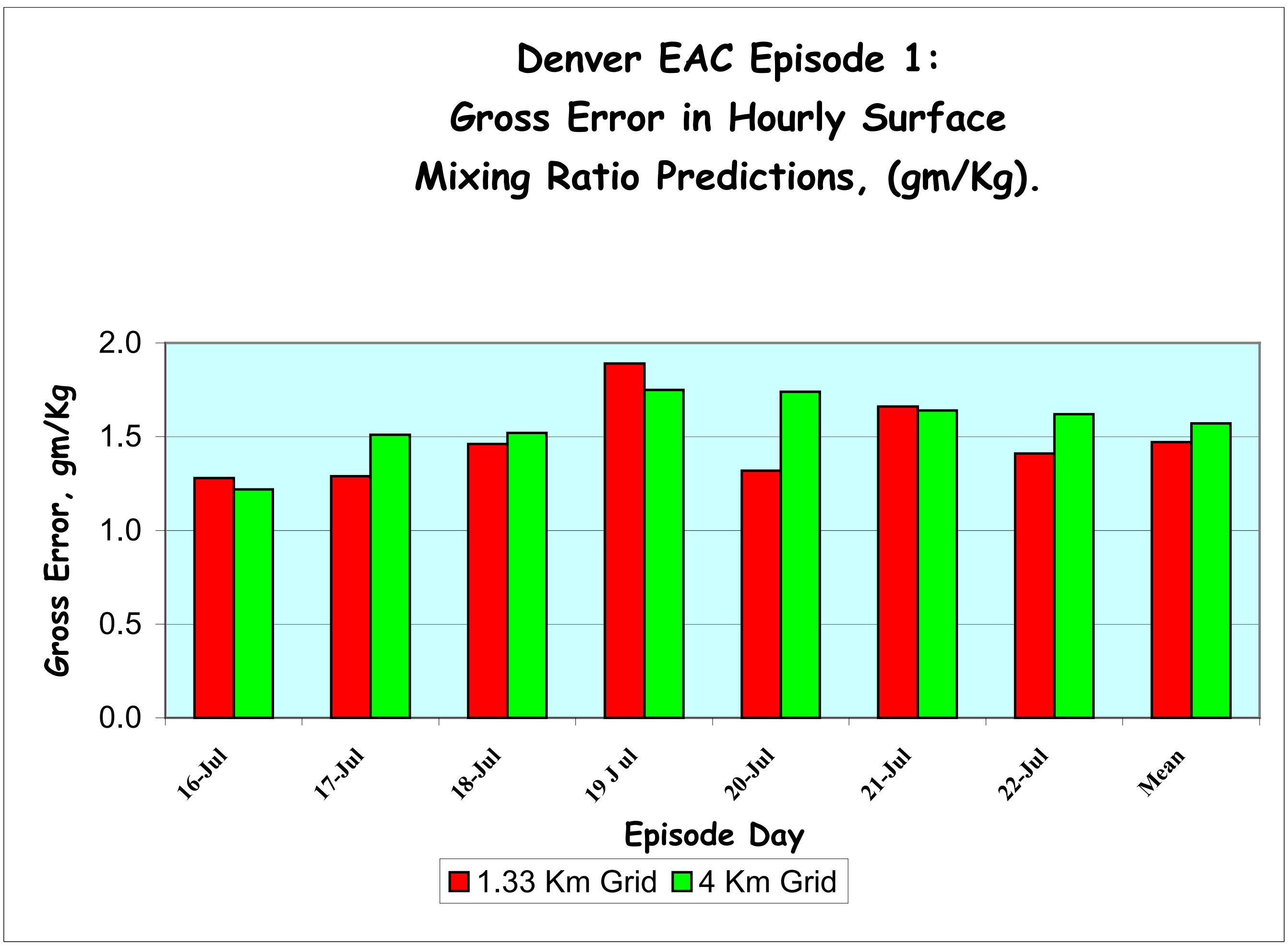


Figure 9-3. Gross Error In MM5 Hourly Surface Mixing Ratio (gm/Kg) for Episode 1.

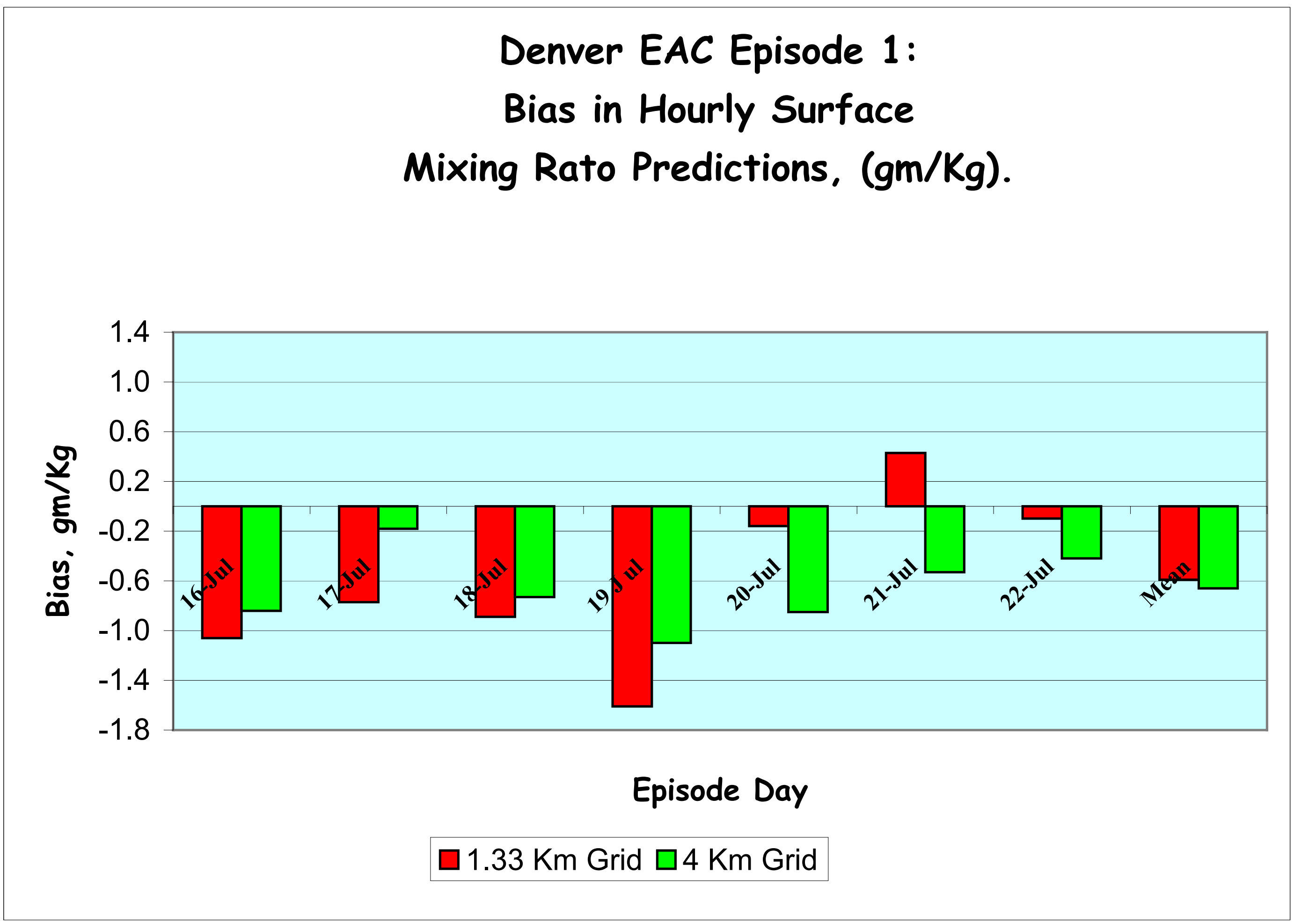


Figure 9-4. Bias In MM5 Hourly Surface Mixing Ratio (gm/Kg) for Episode 1

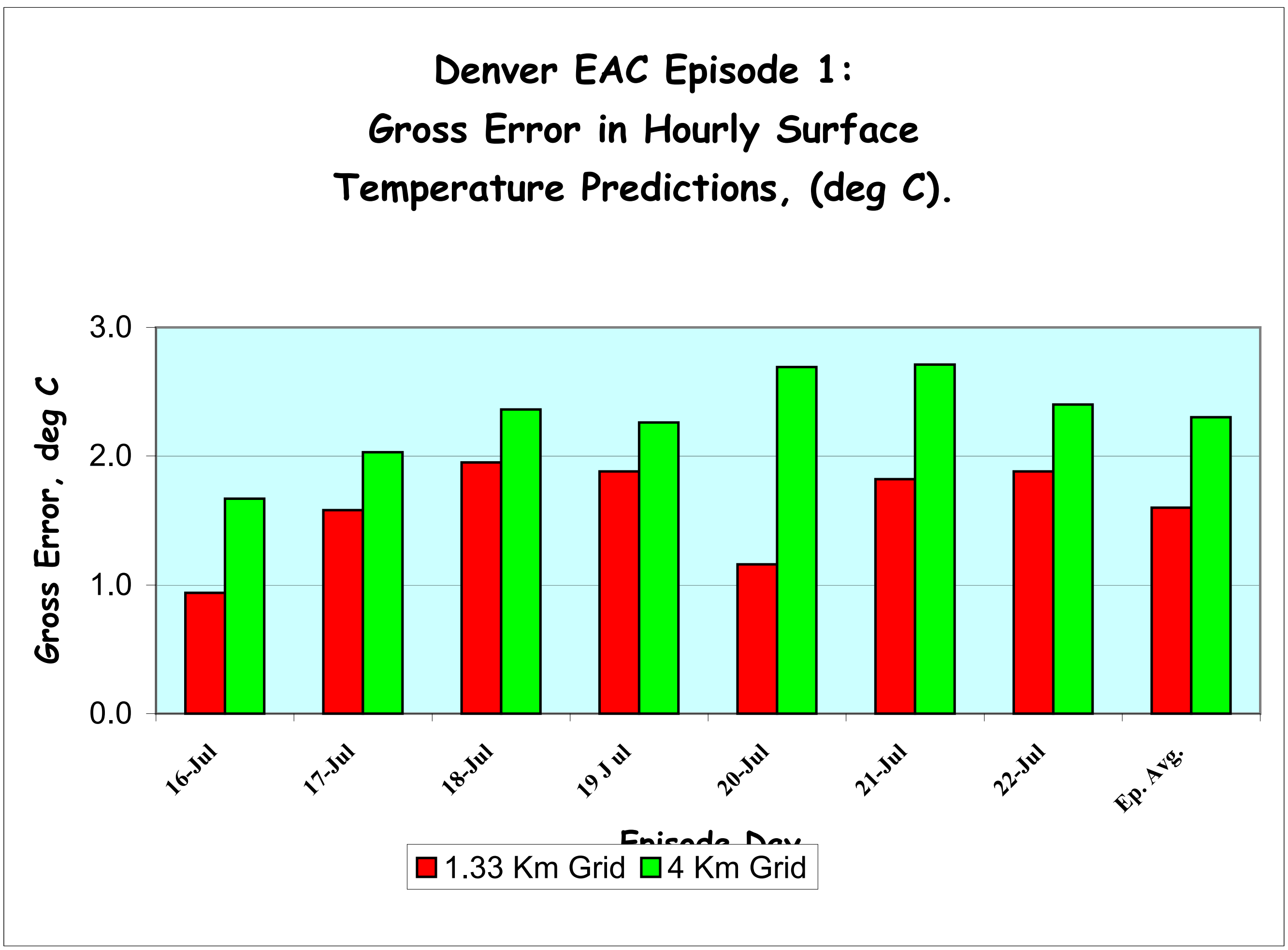


Figure 9-5. Gross Error In MM5 Hourly Surface Temperatures (deg C) for Episode 1.

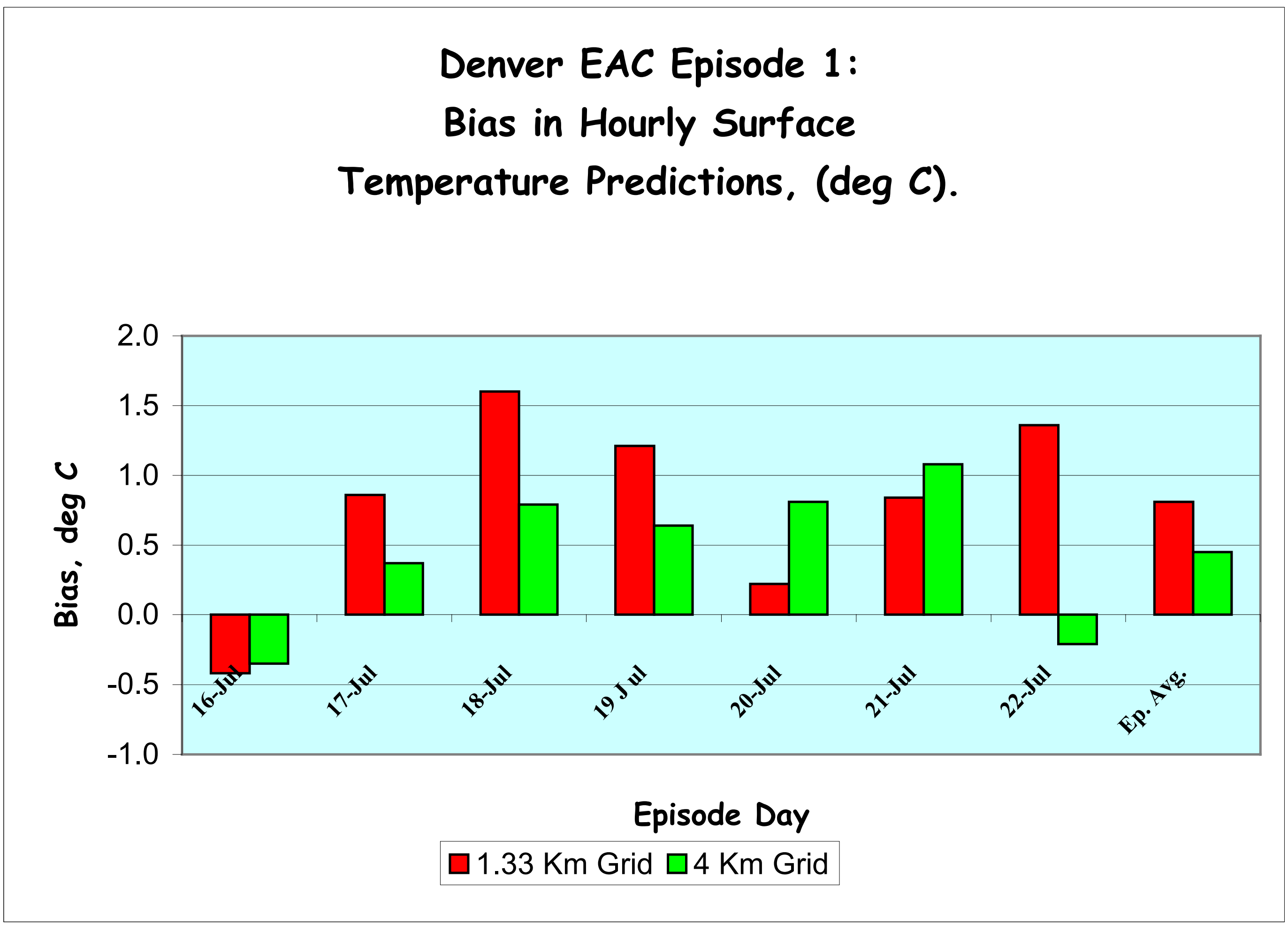


Figure 9-6. Bias In MM5 Hourly Surface Temperatures (deg C) for the Episode 1.

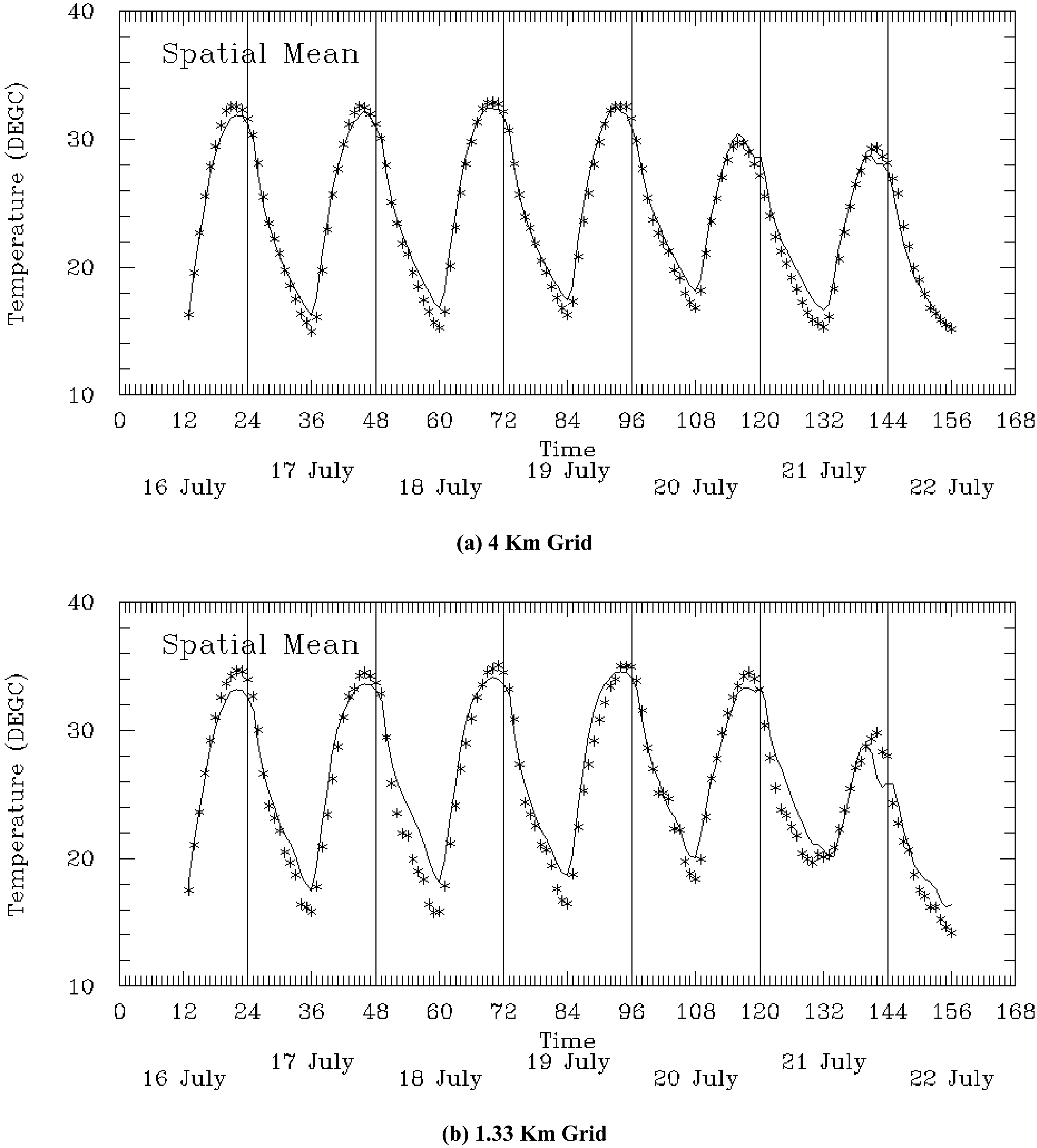


Figure 9-7. Diurnal Variation in Spatial Mean Surface Temperatures for Episode 1.

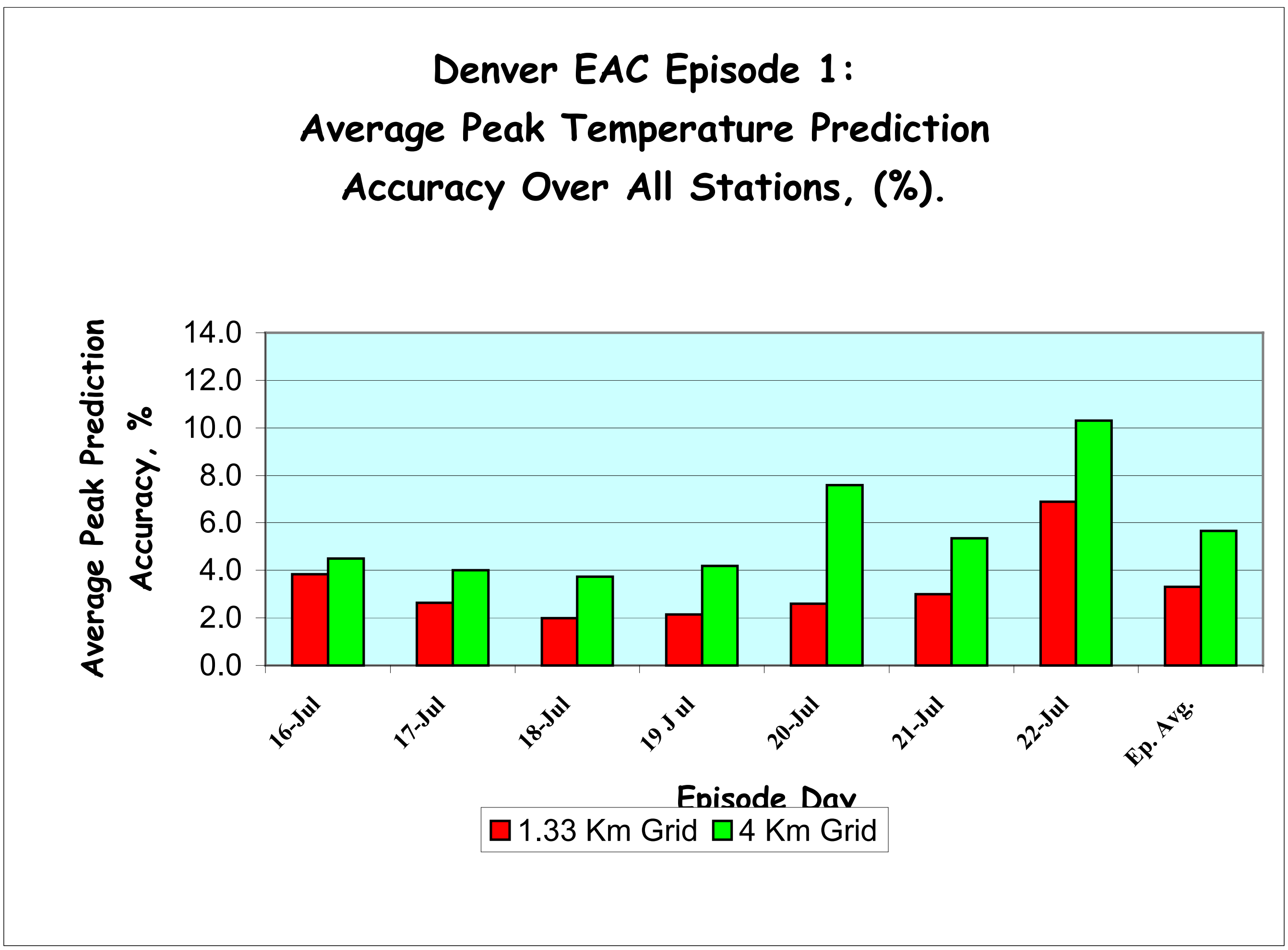


Figure 9-8. **Average Peak Prediction Accuracy Over All Monitors for MM5 Hourly Surface Temperatures (deg C) for Episode 1.**

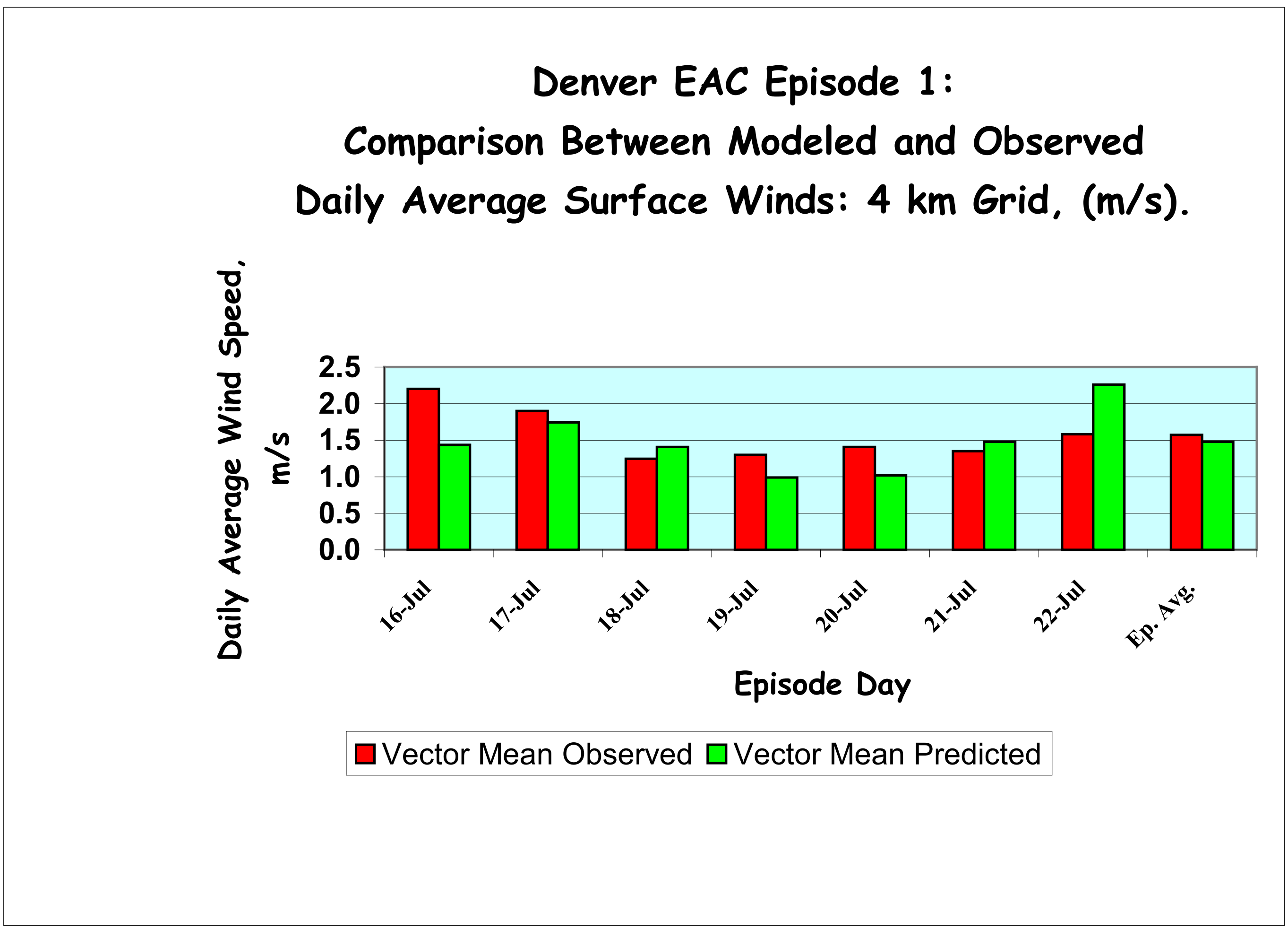


Figure 9-9. Daily Average Modeled and Observed Surface Winds (m/s) on the 4 km Grid for Episode 1.

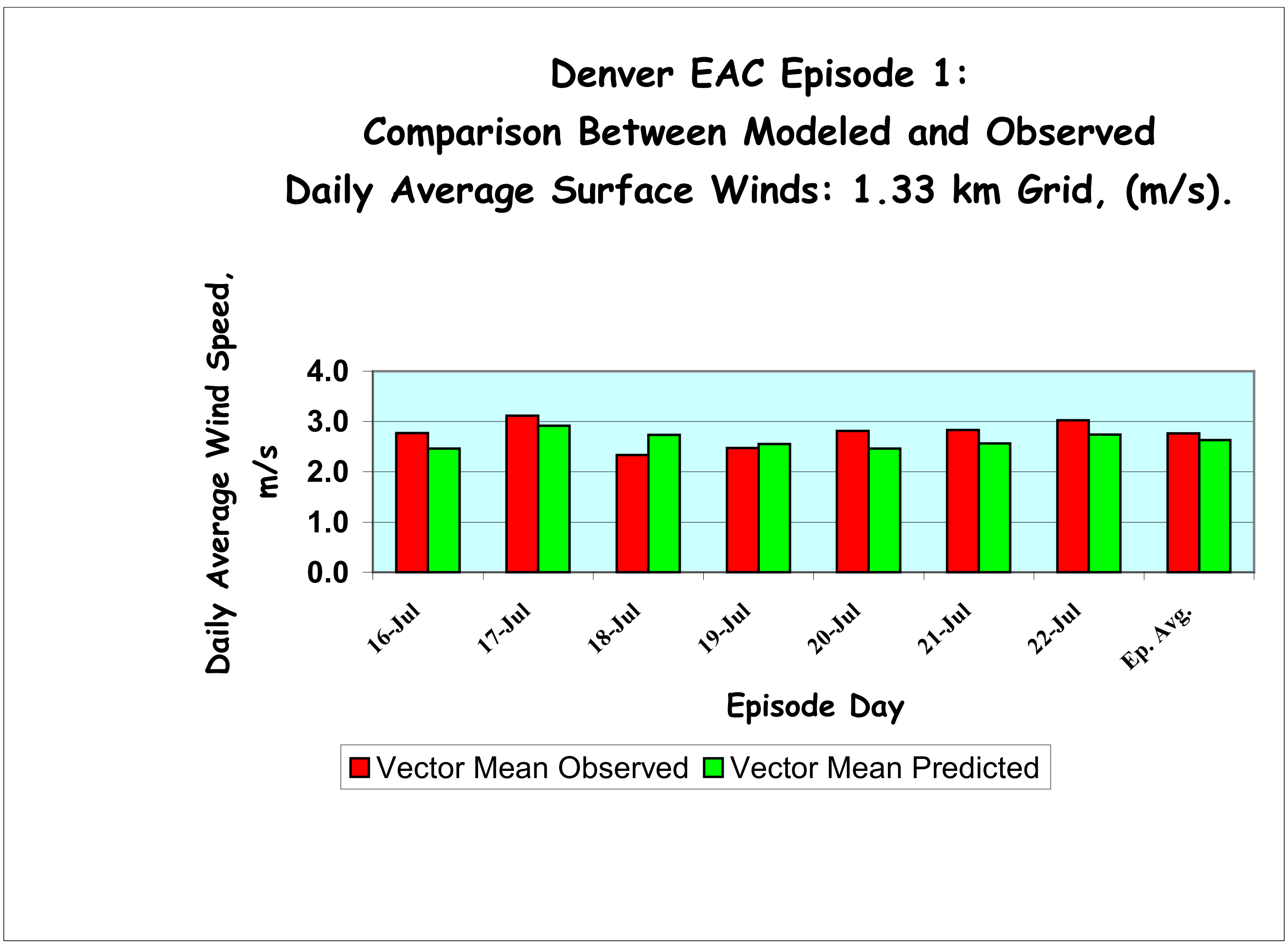


Figure 9-10. Daily Average Modeled and Observed Surface Winds (m/s) on the 1.33 km Grid for Episode 1.

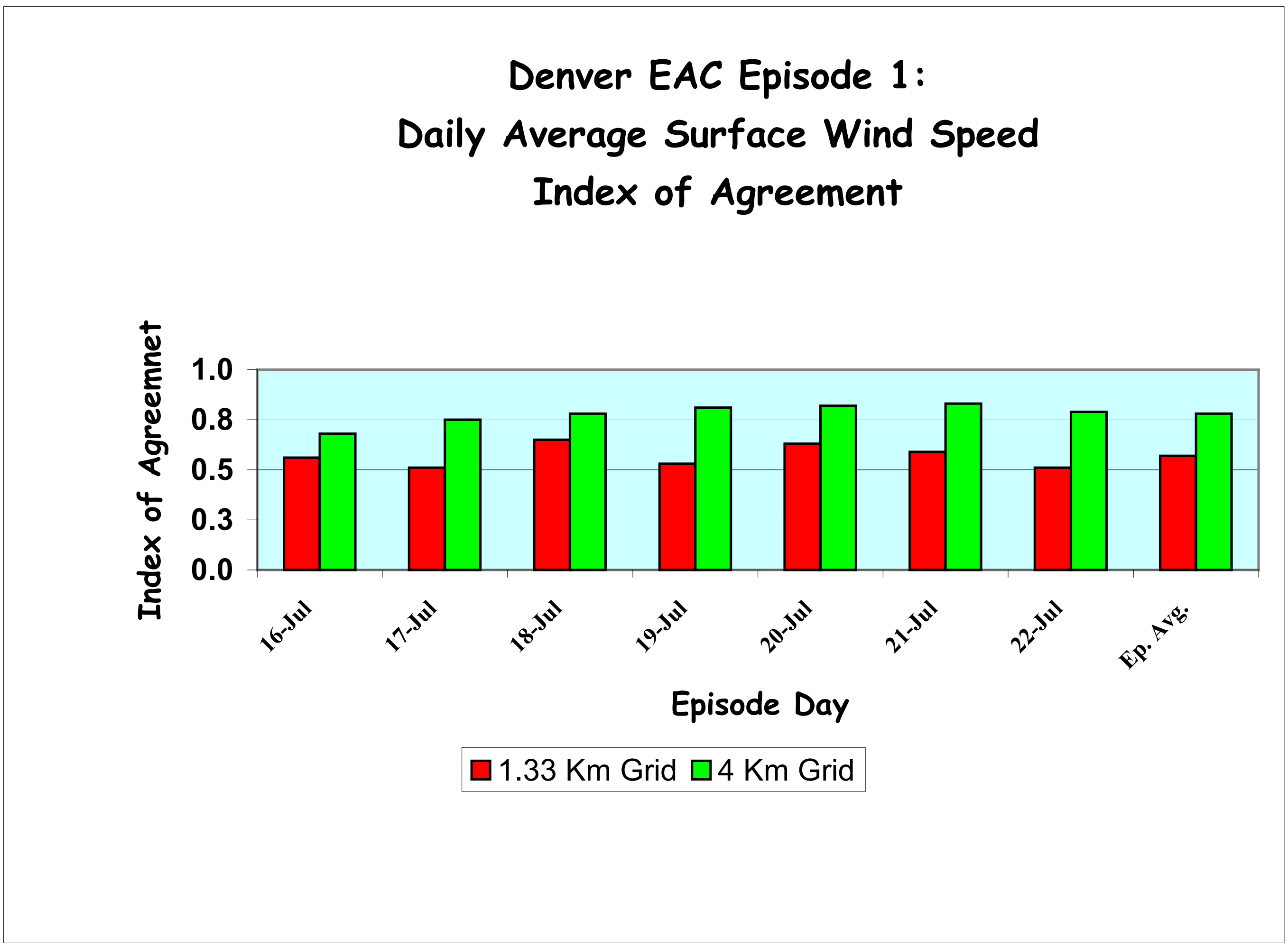


Figure 9-11. Daily Average Surface Wind Index of Agreement for Episode 1.

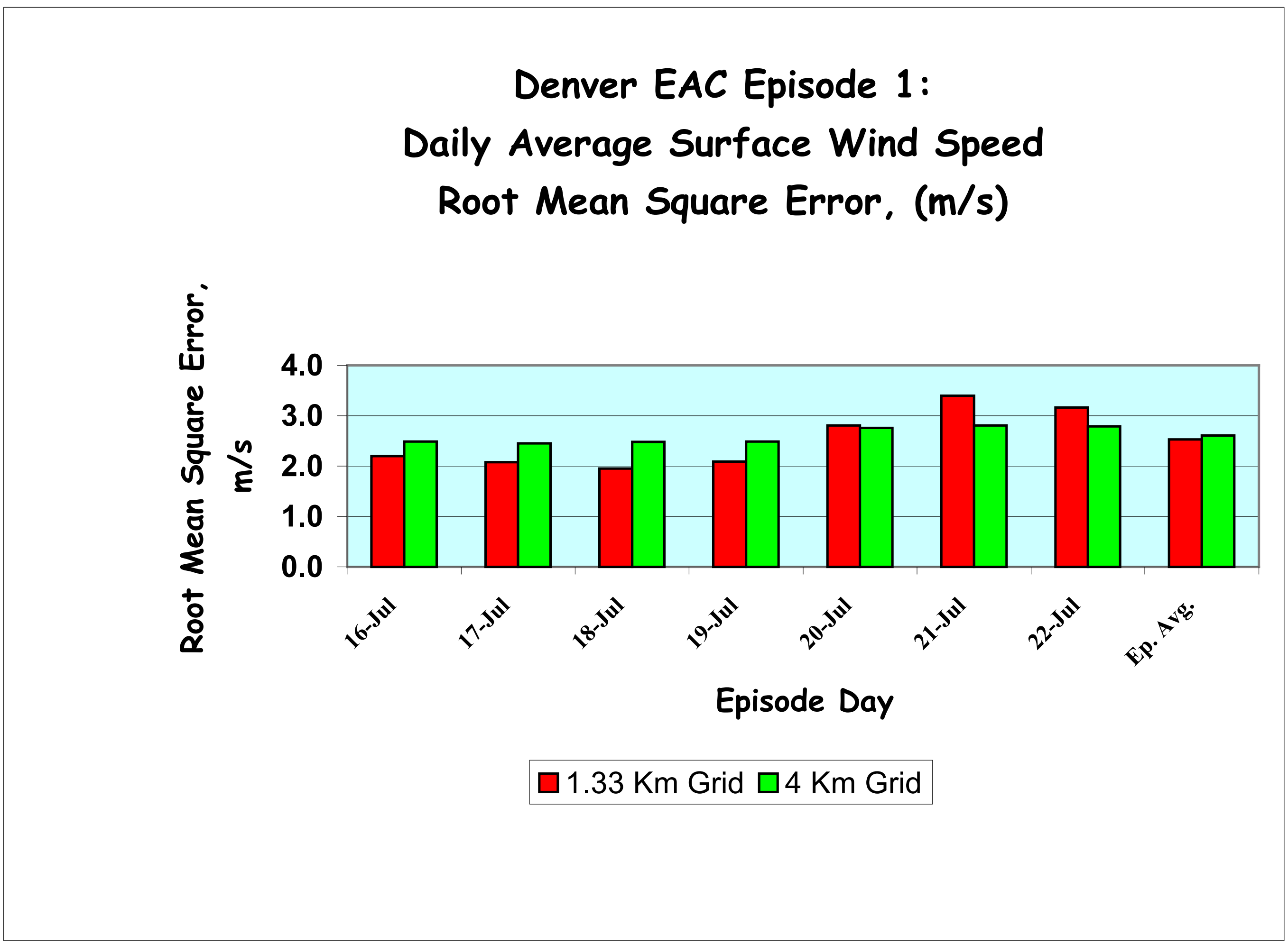


Figure 9-12. Daily Average Surface Wind Speed Root Mean Square Error (m/s) for Episode 1.

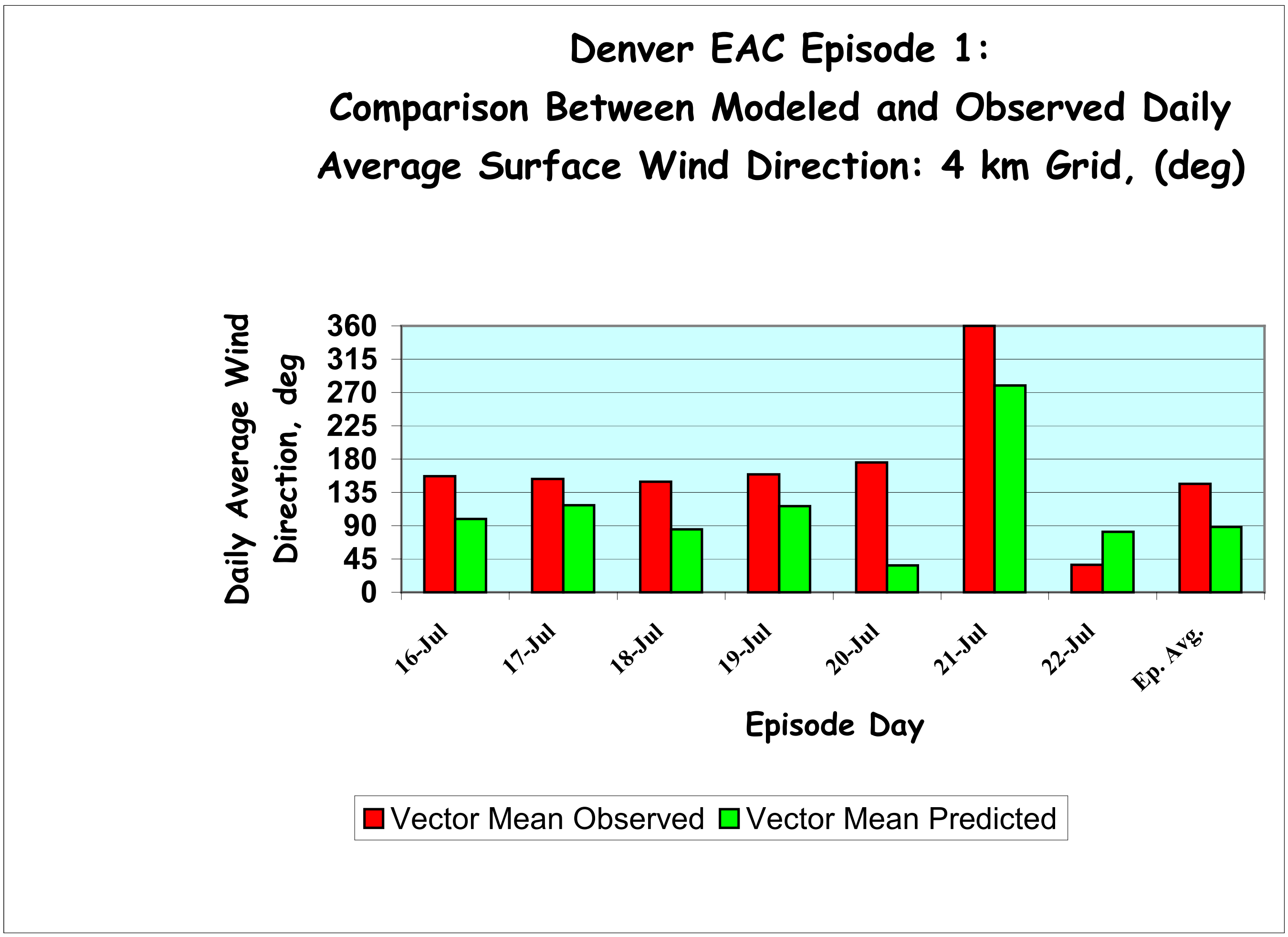


Figure 9-13. Daily Average Modeled and Observed Surface Wind Direction (deg) on the 4 Km Grid for Episode 1.

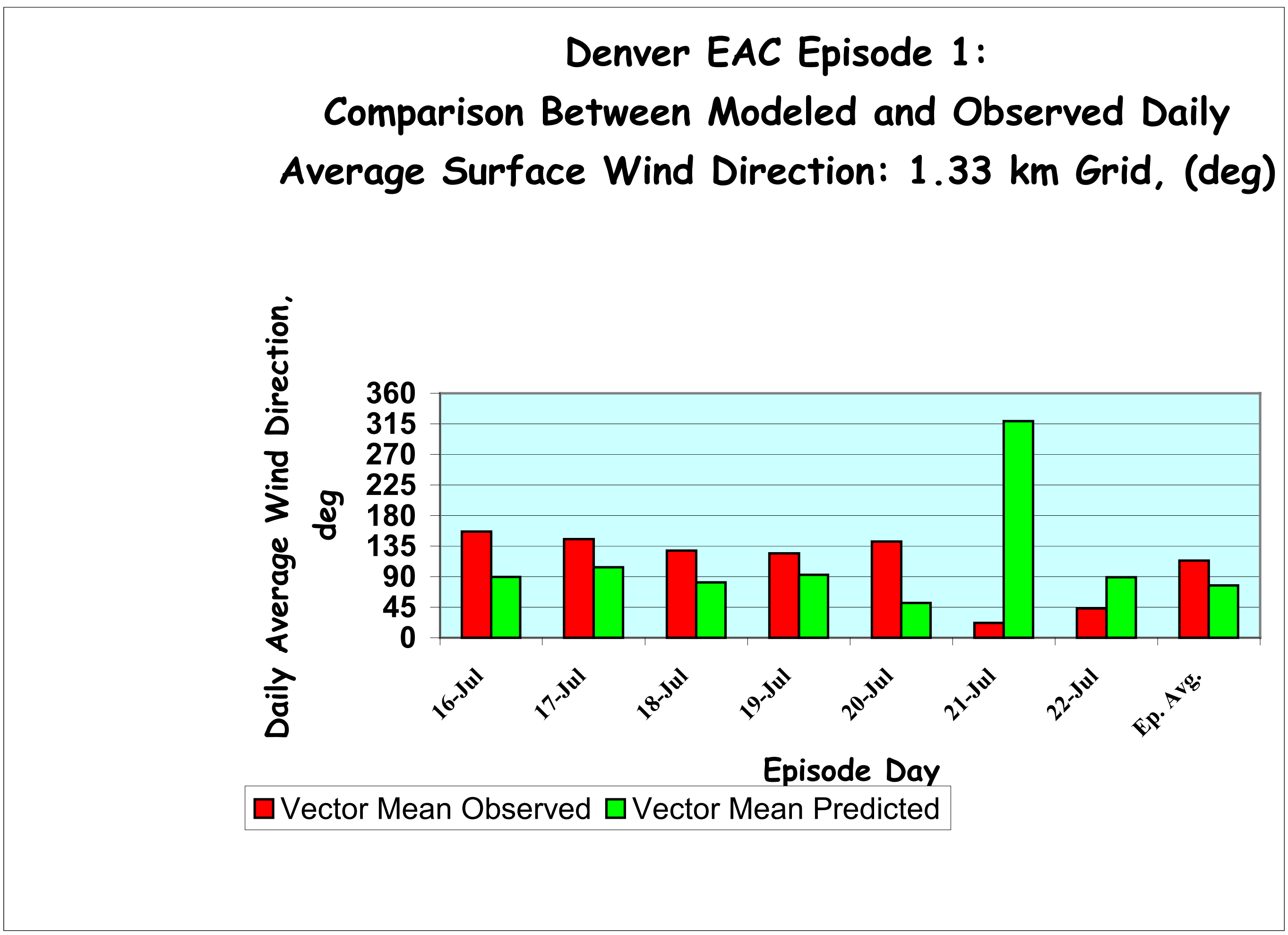


Figure 9-14. Daily Average Modeled and Observed Surface Wind Direction (deg) on the 1.33 Km Grid for Episode 1.

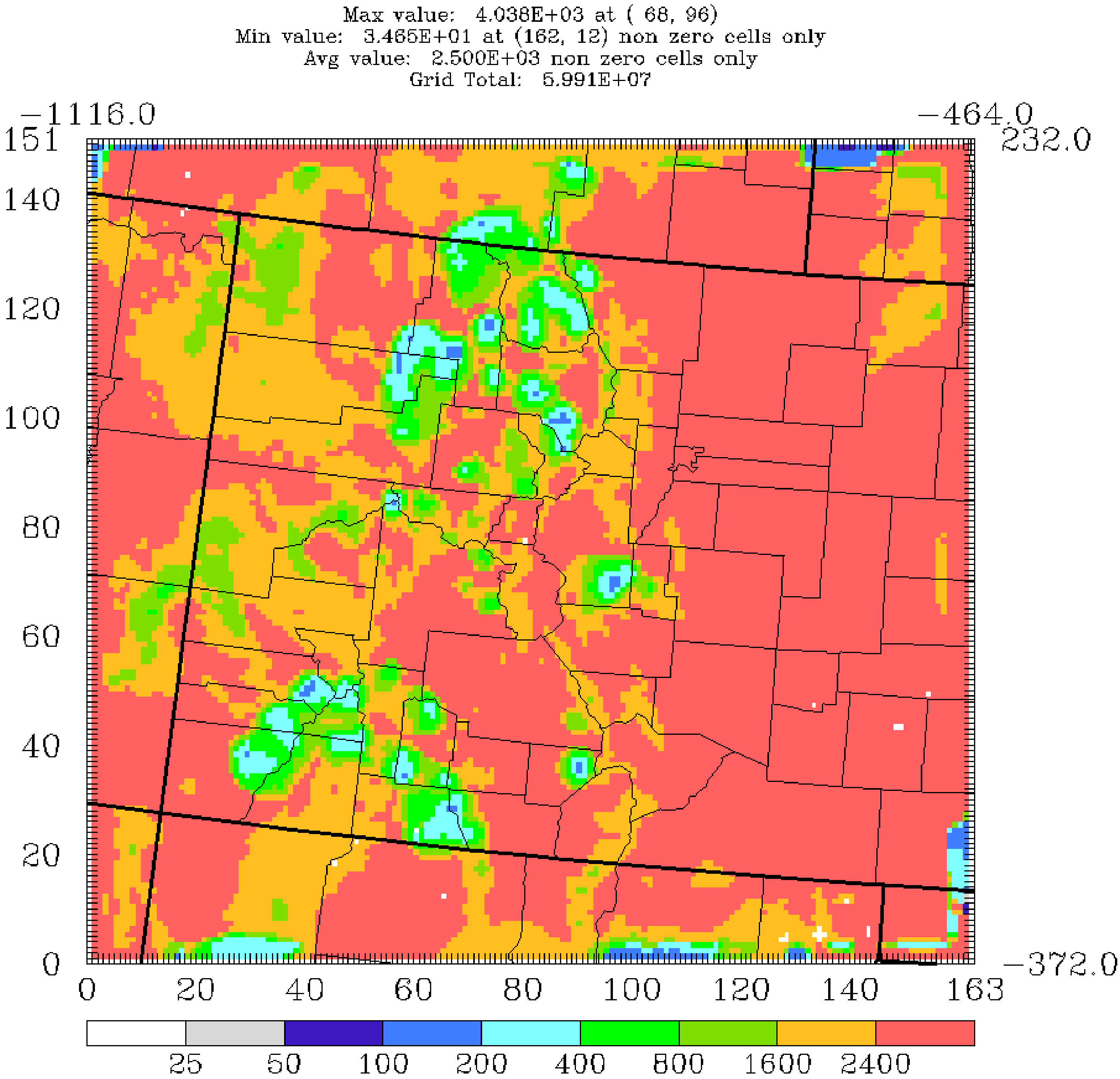


Figure 9-15. Planetary Boundary Layer Heights (m) at 1400 MDT on 20 July 2002 Over the 4 km Grid.

Table 9-1. MM5 Temperature MPE for the Denver EAC Ozone Episode 1: 4/1.3 km Grids.

Date	Day	ATS	AS	AT	AU	A-MEAN	N. Bias	Bias	N. Error	Error	Var	Max. O	Max. P
1.33 Km Grid Domain													
Date	Day	ATS	AS	AT	AU	A-MEAN	N. Bias	Bias	N. Error	Error	Var	Max. O	Max. P
16-Jul	197	-7.71	-6.84	-7.12	-2.26	3.83	-0.81	-0.42	3.47	0.94	1.24	37.22	36.38
17-Jul	198	-5.47	-5.47	-3.18	2.09	2.63	5.43	0.86	7.92	1.58	3.74	36.11	36.86
18-Jul	199	-5.45	-5.45	-3.34	3.01	1.99	9.05	1.60	10.14	1.95	4.47	36.11	37.20
19 J ul	200	-3.91	-3.91	-3.33	1.94	2.15	6.20	1.21	8.60	1.88	4.37	37.22	37.94
20-Jul	201	-6.60	-3.91	-6.16	0.27	2.59	1.93	0.22	5.01	1.16	2.51	37.22	37.32
21-Jul	202	-1.09	-1.09	-0.24	5.24	3.00	4.05	0.84	7.54	1.82	5.03	34.44	36.25
22-Jul	203	-18.53	1.43	-13.52	6.05	6.88	8.83	1.36	10.83	1.88	3.75	29.44	31.22
Ep. Avg.	999	-6.97	-3.61	-5.27	2.33	3.30	4.95	0.81	7.64	1.60	3.59	37.22	37.94
4 Km Grid Domain													
Date	Day	ATS	AS	AT	AU	A-MEAN	N. Bias	Bias	N. Error	Error	Var	Max. O	Max. P
16-Jul	197	-52.66	-52.02	1.93	2.73	4.50	0.38	-0.35	7.24	1.67	12.27	37.78	38.81
17-Jul	198	-53.80	-51.44	0.71	3.31	4.00	3.75	0.37	9.97	2.03	13.26	37.78	39.03
18-Jul	199	-2.20	-1.49	1.01	2.60	3.73	5.80	0.79	11.72	2.36	14.25	37.78	38.76
19 J ul	200	-52.62	-51.45	1.61	2.80	4.18	4.47	0.64	10.59	2.26	14.25	38.33	39.40
20-Jul	201	-2.94	-2.56	-0.39	2.12	7.58	6.08	0.81	12.85	2.69	21.29	38.33	39.14
21-Jul	202	-1.89	-0.80	-0.30	1.53	5.35	7.94	1.08	14.59	2.71	18.67	37.22	37.79
22-Jul	203	3.27	3.82	4.12	5.98	10.31	0.73	-0.21	13.69	2.40	14.49	35.00	37.09
Ep. Avg.	999	-23.26	-22.28	1.24	3.01	5.66	4.16	0.45	11.52	2.30	15.50	38.33	39.40

Table 9-2. MM5 Mixing Ratio MPE for the Denver EAC Ozone Episode 1: 4/1.3 km Grids.

1.33 Km Grid Domain														
Date	Day	ATS	AS	AT	AU	A-MEAN	N. Bias	Bias	N. Error	Error	Var	Max. O	Max. P
16-Jul	197	-68.19	-53.24	-66.40	-22.23	18.32	-12.01	-1.06	16.20	1.28	2.81	12.90	10.03
17-Jul	198	-62.01	-59.89	-61.32	-0.62	15.51	-9.67	-0.77	17.73	1.29	2.38	12.00	11.93
18-Jul	199	-60.08	-57.03	-59.63	-12.11	19.65	-10.67	-0.89	21.59	1.46	2.47	11.14	9.79
19 J ul	200	-32.98	-32.83	-29.69	-13.32	25.65	-19.19	-1.61	24.42	1.89	2.84	12.37	10.72
20-Jul	201	-50.03	-42.32	-49.49	7.23	22.42	-0.64	-0.16	17.20	1.32	2.78	12.03	12.90
21-Jul	202	1.86	1.86	3.88	23.97	12.12	5.49	0.43	18.77	1.66	3.93	11.69	14.49
22-Jul	203	-24.69	-11.24	-21.90	-3.10	6.56	0.47	-0.10	12.56	1.41	3.10	14.79	14.33
Mean	999	-42.30	-36.39	-40.65	-2.88	17.17	-6.60	-0.59	18.35	1.47	2.90	14.79	14.49
4 Km Grid Domain													
Date	Day	ATS	AS	AT	AU	A-MEAN	N. Bias	Bias	N. Error	Error	Var	Max. O	Max. P
16-Jul	197	-67.58	-53.61	-64.04	40.39	14.42	-11.75	-0.84	16.89	1.22	1.96	12.90	18.11
17-Jul	198	-32.66	-28.44	-31.40	49.92	19.68	-2.47	-0.18	20.83	1.51	3.59	13.00	19.49
18-Jul	199	-43.46	-33.36	-26.86	41.04	18.23	-8.64	-0.73	20.32	1.52	3.00	13.96	19.69
19 J ul	200	-39.34	-39.34	-35.32	43.22	22.81	-11.94	-1.10	21.26	1.75	3.39	12.85	18.40
20-Jul	201	-32.05	-21.29	-29.67	50.39	19.98	-7.96	-0.85	19.27	1.74	3.96	12.99	19.54
21-Jul	202	-48.42	-29.23	-46.49	42.93	14.65	-5.33	-0.53	17.67	1.64	3.67	14.72	21.04
22-Jul	203	-0.68	-0.55	1.22	40.55	16.01	-5.72	-0.42	19.31	1.62	3.62	15.21	21.38
Mean	999	-37.74	-29.40	-33.22	44.06	17.97	-7.68	-0.66	19.36	1.57	3.31	15.21	21.38

Table 9-3. MM5 Surface Wind MPE for the Denver EAC Ozone Episode 1: 4/1.3 km Grids.

1.33 Km Grid Domain							
Date	Day	VMOBS	VMEST	RMSE	IA	OBSDIR	ESTDIR
16-Jul	197	2.77	2.46	2.20	0.56	156	90
17-Jul	198	3.11	2.91	2.08	0.51	145	104
18-Jul	199	2.33	2.73	1.95	0.65	128	82
19-Jul	200	2.47	2.55	2.09	0.53	124	93
20-Jul	201	2.81	2.46	2.81	0.63	142	51
21-Jul	202	2.83	2.56	3.40	0.59	22	319
22-Jul	203	3.02	2.74	3.16	0.51	43	89
Ep. Avg.		2.76	2.63	2.53	0.57	114	77
4 Km Grid Domain							
Date	Day	VMOBS	VMEST	RMSE	IA	OBSDIR	ESTDIR
16-Jul	197	2.20	1.44	2.49	0.68	157	99
17-Jul	198	1.90	1.74	2.45	0.75	153	118
18-Jul	199	1.25	1.41	2.48	0.78	149	85
19-Jul	200	1.30	0.99	2.49	0.81	159	116
20-Jul	201	1.41	1.02	2.76	0.82	175	36
21-Jul	202	1.35	1.48	2.81	0.83	360	280
22-Jul	203	1.58	2.26	2.79	0.79	37	82
Ep. Avg.		1.57	1.48	2.61	0.78	146	88

Table 9-4. Summary Results for the 16-22 June 2002 MM5 Simulation on the 4/1.33 km High Resolution Grids Compared with the Ad Hoc Performance Benchmarks and Fifty Recent Prognostic Model Performance Evaluations Throughout the U.S.

Episode Grid Resolution	Temperature, deg C		Mixing Ratio, kg/KG		Surface Winds, (m/s)		
	Bias	Error	Bias	Error	RMSE	I	WD diff
4 km	0.45	2.30	-0.66	1.57	2.61	0.78	60
1.33 km	0.81	1.60	-0.59	1.47	2.53	0.57	37
Benchmark	≤ ± 0.5	≤ 2.0	≤ ± 1.0	≤ 2.0	≤ 2.00	≥ 0.60	≤ 30
U.S. Average	-0.2	2.0	0.0	1.8	2.00	0.71	24

10.0 MM5 EVALUATION FOR THE 24 JUNE-2 JULY '02 EPISODE: 4/1.33 KM GRIDS

This chapter presents results of the operational evaluation of the MM5 model for the 24 June-2 July 2002 ozone episode over the Denver-Northern Front Range Region. In this evaluation we focus on the 4 km and 1.33 km grid results.

Figures 10-1 and 10-2 present the predicted surface winds on 28 June 2002 over the 4 km (1200 MDT) and 1.33 km (0600 MDT) grid regions, respectively. The good correspondence between the measurements and observations is particularly encouraging since the surface winds at the 4 km and 1.33 km scale were not subjected to nudging. In reviewing the full set of surface winds for this episode, we did not encounter any obviously flawed wind predictions or indications of spurious model behavior.

10.1 Mixing Ratio

Figures 10-3 and 10-4 show the daily average gross error and bias in hourly near-surface mixing ratio for the 24 June-2 July 2002 episode. For the mixing ratio gross error, Figure 10-3 shows day-to-day variation in the mixing ratio errors on both nested grids with maxima typically occurring toward the end of the episode. The errors range from 1.2 gm/Kg to 2.2 gm/Kg. On most days, the MM5 performs slightly better on the 4 km grid compared to the 1.33 km grid (just the opposite of Episode 1). The episode average gross errors in mixing ratio on the 4 km and 1.33 km grids (1.59 gm/Kg and 1.68 gm/Kg, respectively) are shown in the two far right hand bars in Figure 10-3 and in Table 10-2.

For the mixing ratio bias (Figure 10-4), the MM5 does well for the first five days of the episode before tending to underestimate between 29 June-1 July on the 1.33 km grid. The biases range from about –1.85 gm/Kg to 0.31 gm/Kg. For daily average bias, the MM5 performs systematically better on the 4 km grid compared to the 1.33 km grid. The episode average biases in mixing ratio on the 1.33 km and 4 km grids (-0.70 gm/Kg and -0.17 gm/Kg, respectively). These results are good given the 15.5 m vertical displacement between measurement and prediction height.

10.2 Temperatures

Figures 10-5 and 10-6 depict the daily average gross error and bias in hourly near-surface temperatures on the two nested grids. Table 10-1 summarizes the statistical results. For the temperature gross errors, Figure 10-5 shows day-to-day variation in the temperature errors on both grids with the errors ranging from 1.84 deg C to 3.01 deg C. On all but 26 June MM5 performs better on the 1.33 km grid compared to the 4 km grid. The episode average gross errors in temperature on the 1.33 km and 4 km grids are 2.35 deg C and 2.75deg C, respectively. These error statistics are somewhat poorer than Episode 1.

As with Episode 1, the daily average temperature bias (Figure 10-6) using the Blackadar pbl scheme leads to systematic over-estimation on both grids for all days. For temperature bias, MM5 appears to perform much better on the 4 km grid compared to the 1.33 km grid. The daily average of the hourly temperature biases range between about 0.12 deg C to 2.85 deg C. The episode average biases in temperature on the 4 km and 1.33 km grids are 0.65 deg C and 1.78 deg C, respectively. Note that the temperature bias over the 1.33 km domain is, on average, more than twice that on the 4 km grid.

The average diurnal behavior of the MM5's predictions of near surface temperature on the 4 km and 1.33 km grids are shown in Figure 10-7. Recall that there are 40 surface temperature stations within the 4 km domain while only 8 stations are within the 1.33 km high resolution grid. From Figure 10-7 the MM5's skill in predicting the diurnal temperature profiles on the 4 km and 1.33 km grids for Episode 2 is roughly the same as for Episode 1 except that in Figure 10-7b (the 1.33 km grid) there is larger warm bias at night for Episode 2. Except for this nighttime warm bias on both grids, the models diurnal temperature predictions appear adequate based on comparisons with other studies. The cool afternoon temperature bias and warm nighttime temperature bias, common to most mesoscale modeling simulations is again evident at both grid scales for Episode 2.

The daily average peak temperature prediction accuracy over the 4 km and 1.33 km grids is shown in Figure 10-8. On the 4 km grid, the accuracies of peak temperature prediction range between roughly –0.3% to 3% while for the 1.33 km domain the range is about –0.3% to 7%. Overall, the average peak afternoon temperature prediction accuracy on the 4 k and 1.33 km grids are 1.6% and 2.82%, respectively. These accuracy figures, even better than for Episode 1, are quite good for the fine spatial scales modeled here.

10.3 Wind Speed and Direction

Figures 10-9 through 10-14 present daily average wind performance statistics on the 4 km and 1.33 km grids. Beginning with Figure 10-9, the daily average modeled and observed surface wind speeds on the 4 km grid are plotted as a function of time during the episode. The mean predicted wind speeds (1.64 m/s) are about 33% greater than observations (1.23 m/s) when averaged across all days in the 24 June-2 July 2002 episode. For the 1.33 km winds (Figure 10-10), the average modeled and observed surface wind speeds are 2.62 m/s and 2.44 m/s, respectively. As with Episode 1, wind speeds on the 4 km grid are roughly 40%-50% slower than the 1.33 km grid.

Figure 10-11 depicts the daily average index of agreement on the 4 km and 1.33 km grids. On the 4 km grid, the index ranges between 0.74 to 0.87 across the 9 modeling days. The 1.33 km grid results are poorer but still ranging from 0.50 to 0.73. Across the entire episode, the index of agreement for the 4 km and 1.33 km grids are 0.80 and 0.60, respectively. Thus, the index of agreement results for Episodes 1 and 2, at the 4 km and 1.33 km scales, are comparable.

Daily average root mean square errors (RMSE) vary between 2.46 m/s to 3.09 m/s on the 4 km grid (Figure 9-12) while they are somewhat poorer on the 1.33 km grid, ranging between 2.39 m/s to about 3.50 m/s (see Table 10-3). Across the entire episode, the RMSE errors the 4 km and 1.33 km grids are 2.77 m/s and 2.84 m/s, respectively.

Figures 10-13 and 10-14 present comparisons between the daily average predicted and observed wind direction for the two MM5 grids. On the 4 km grid (Figure 10-13), the wind directions match reasonably well for the first half of the episode, but beginning 30 June the mean predicted winds are veered more to the southwest compared with the average observations. This same feature (veering of the predicted wind relative to the observations) is also seen in the 1.33 km winds (Figure 10-14) after 28 June. Across the entire episode, the average wind direction discrepancies on the 1.33 km grid is 43 deg, while for the 4 km grid the mean difference is 29 deg. These wind direction results are somewhat better than those found in Episode 1.

The MM5's performance aloft was examined by reviewing the planetary boundary layer height (PBL) fields for each hour of the 24 June-2 July 2002 episode and by studying the skew-T plots for the sounding locations within the 4 km and 1.33 km domains. Figure 10-15 presents a typical pbl height field at 1400 MDT on 28 June 2002 over the 4 km domain. This plot shows the high afternoon boundary layer heights (2400m or more above ground level) over most of the 4 km modeling domain and lower depths (100m-800m) over the mountainous terrain of the Front Range and the San Juan Mountains to the south. Examining the full set of fields contained on the CD archive, we found no unusual or erratic behavior. As with Episode 1, we also examined the MM5's performance aloft by comparing the modeled and observed upper level skew-T plots containing horizontal winds, temperatures, and mixing ratios for at the standard 0000 and 1200 UTC sounding times each day at Denver for the 1.33 km grid and at Denver and Grand Junction on the 4 km grid domain. This brief review did not reveal any causes for concern with the MM5 predictions aloft as they might adversely affect the photochemical model applications.

10.4 Comparisons with Other Studies

Tables 10-1 through 10-3 summarize the episode composite temperature, wind speed, wind direction and mixing ratio statistics for the 24 June-2 July 2002 episode. Below, we compare the MM5 results on the 4 km grid with the much broader set of prognostic model applications studies listed in Table 8-1. As indicated in Chapter 9, the statistics for most of the studies reported in Table 8-1 were derived from model applications at scales between 4 km and 12 km so the current MM5 results for Episode 2 are from the lower bound of this range.

The MM5 average bias in hourly ground level temperatures over the 4 km grid is 0.65 deg C. The average across all studies is –0.2 deg C (Figure 8-15). Thus, the current simulation tends to have a systematic warm bias relative to the other studies. As with Episode 1, the source of this bias is the use of the Blackadar pbl scheme which replaced the PX land use module employed on the 36 km and 12 km grid meshes. The episode average gross error in hourly ground level temperatures for the 4 km MM5 simulation is 2.75 deg C. The mean temperature error over all studies is 2.0 deg C (Figure 8-16). Thus, the current simulation for Episode 2 on the 4 km grid tends to have somewhat higher gross errors in ground level hourly temperature predictions compared to other studies.

The MM5 average bias in hourly ground level mixing ratios on the 4 km grid is –0.17 gm/Kg. The average across all studies is 0.0 gm/Kg (Figure 8-17). This is a very small negative moisture bias relative to the other studies. The episode average error in hourly ground level mixing ratio for the current 4 km MM5 run is 1.59 gm/Kg, nearly identical to Episode 1. The mean error over all studies is 1.8 gm/Kg (Figure 8-18). As with the temperatures, the MM5 Episode 2 simulation tends to have comparable errors in ground level mixing ratios compared to other studies. The episode average error in hourly ground level wind speed on the 4 km domain is 33%. The average across all studies is 32% (Figure 8-19). Thus, the Episode 2 simulation has essentially the same wind speed errors on the 4 km grid, on average, compared to other studies. The episode average root mean square error (RMSE) in hourly ground level wind speed prediction is 2.77 m/s. The average RMSE error over all studies is 2.00 m/s (Figure 8-20). As with Episode 1, the Episode 2 grid simulation tends to have somewhat higher RMSE wind speed errors, on average, compared to other studies. The episode average IOA for hourly ground level wind speed prediction is 0.80, compared to 0.78 for Episode 1. The average index of agreement (I) over all studies is 0.71 (Figure 8-21). Thus, on the 4 km grid, the MM5 tends to yield a better a statistical score for I compared to other studies. The episode average discrepancy in hourly wind direction for the Denver EAC 4 km simulation is 29 deg C which is less than half the direction error in Episode 1 and is essentially comparable with the size of the average wind direction error (24 deg C) from the other studies (Figure 8-22). Thus, on the 4 km grid, the current MM5 simulation tends to have equivalent wind direction errors compared to other studies.

10.5 Assessment of the 24 June-2 July 2002 Episode

Table 10-4 compares the 4 km and 1.33 km MM5 results for the 16-22 July 2002 episode with the proposed meteorological modeling benchmarks. Shaded cells in the table correspond to those meteorological variables that fall just outside of the benchmark ranges. On the 4 km grid for the 24 June-2 July 2002 episode the bias and gross error in surface temperature prediction and the RMSE error in surface wind speed prediction fall outside the suggested model performance benchmarks and the average results for model applications at scales ranging from 4 km to 12 km. The remaining four statistical measures are within the suggested performance ranges. The results on the 1.33 km grid are only slightly poorer when compared with the benchmarks; again, this is attributed in part to the difference in scales. We conclude that the 4/1.33 km MM5 meteorological fields for the 24 June-2 July 2002 episode may be used, with reasonable caution, as input to the regional emissions and photochemical models for air quality impacts assessments for the Denver EAC study.

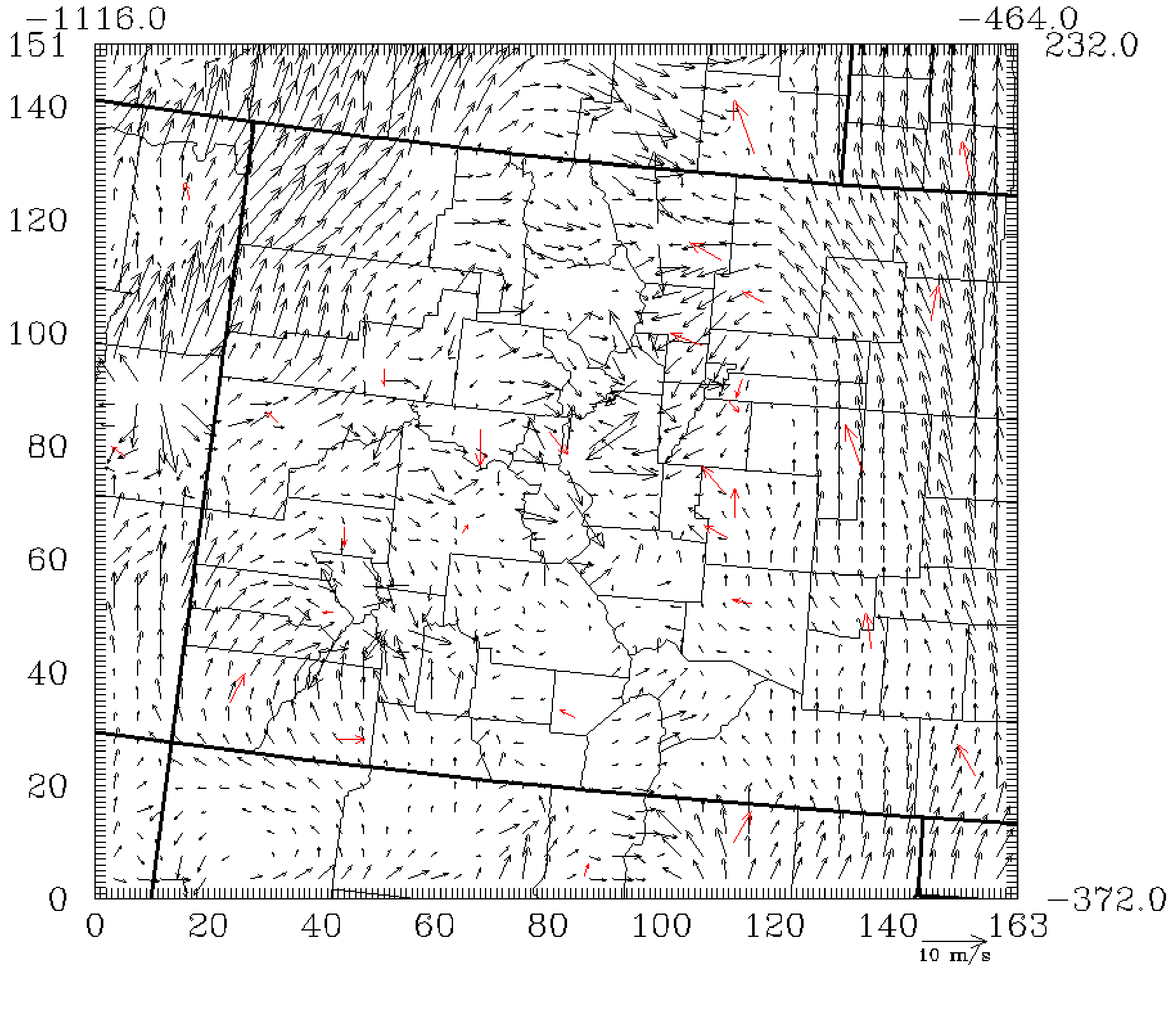


Figure 10-1. MM5 Surface Wind Fields at 1200 MDT on 28 June 2002 Over the 4 km Domain. (Predicted Vectors Plotted Every Fourth Grid Cell).

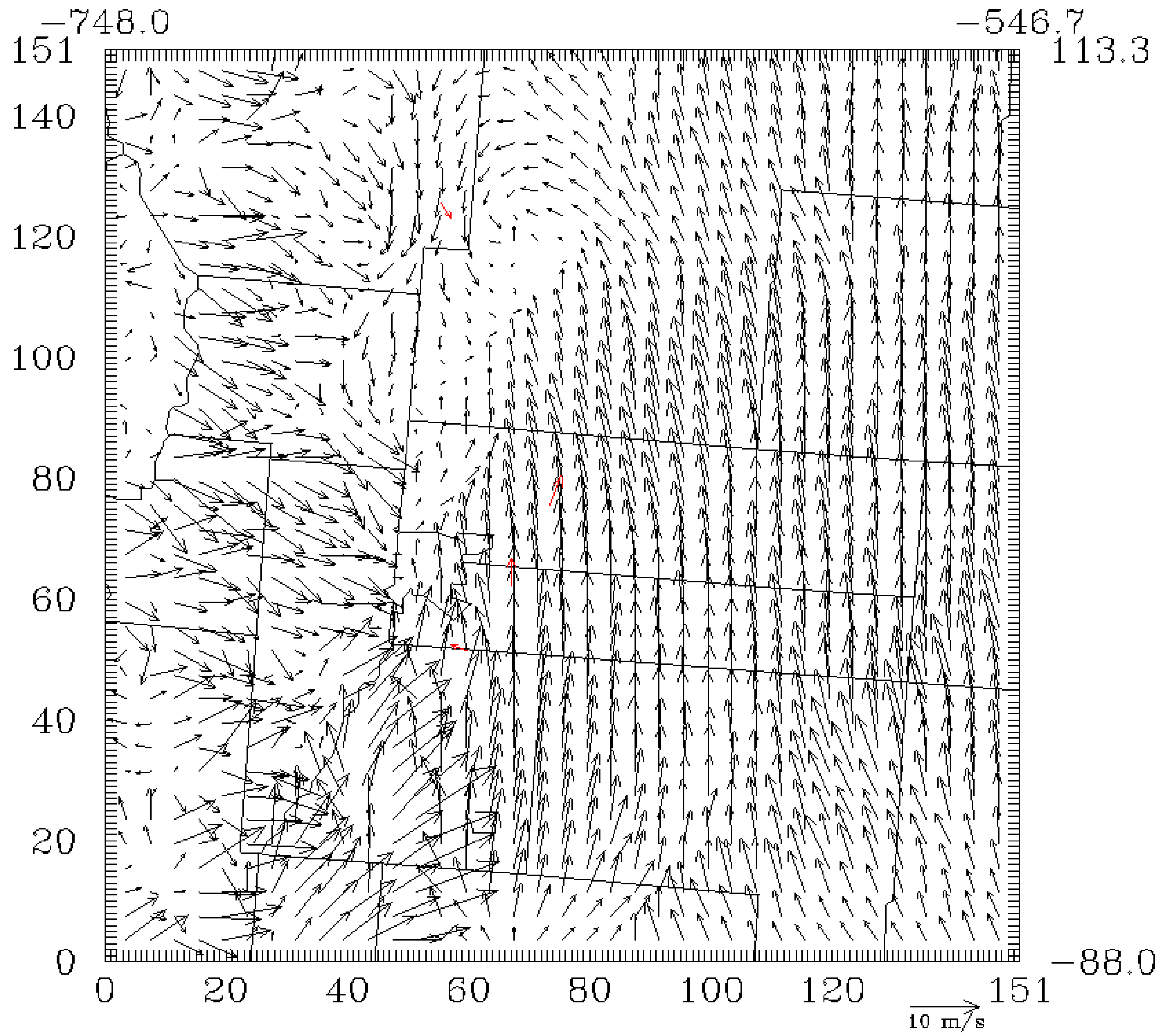


Figure 10-2. MM5 Surface Wind Fields at 0600 MDT on 28 June 2002 Over the 1.33 km Domain. (Predicted Vectors Plotted Every Fourth Grid Cell).

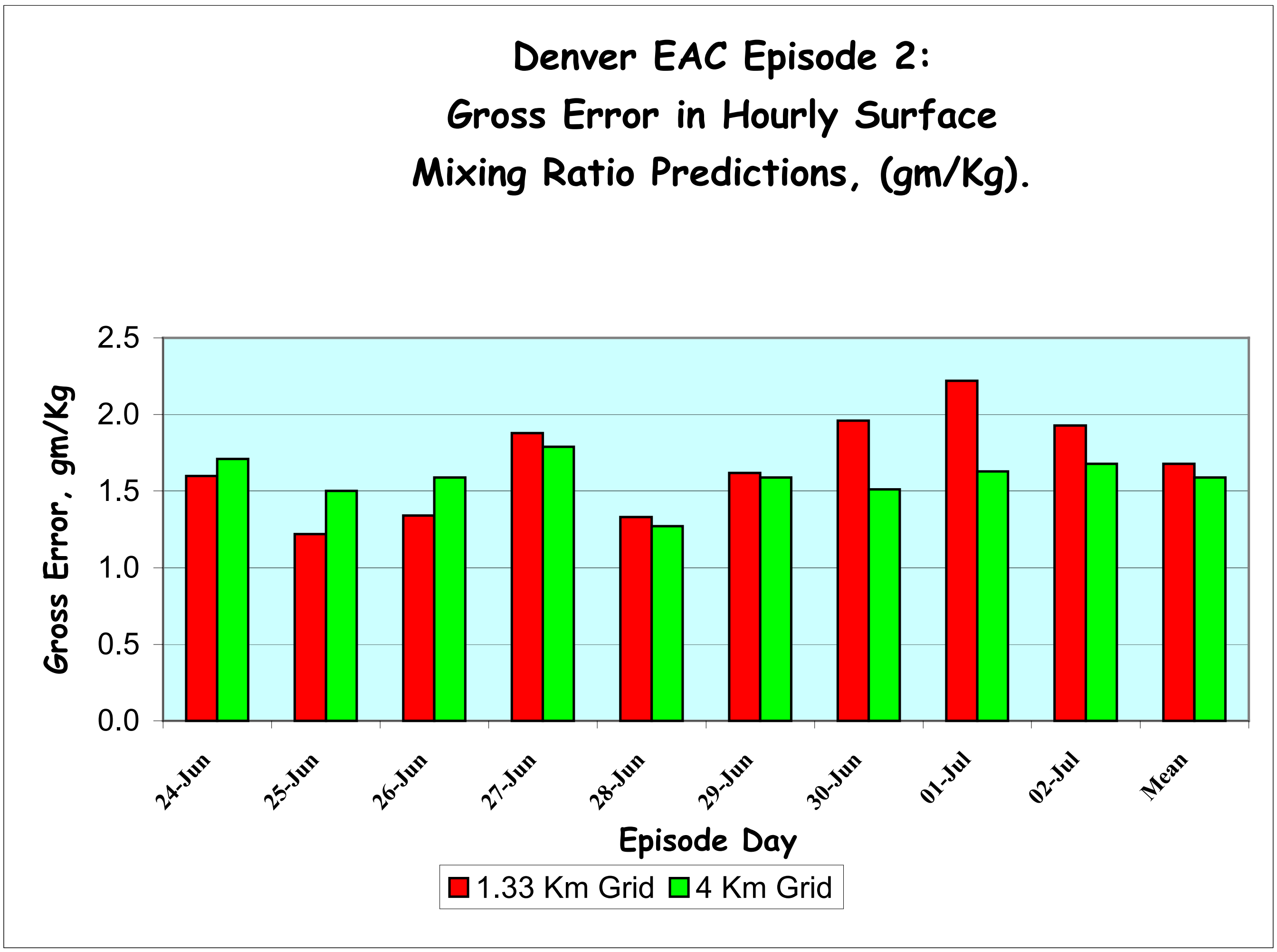


Figure 10-3. Gross Error In MM5 Hourly Surface Mixing Ratio (gm/Kg) for Episode 2.

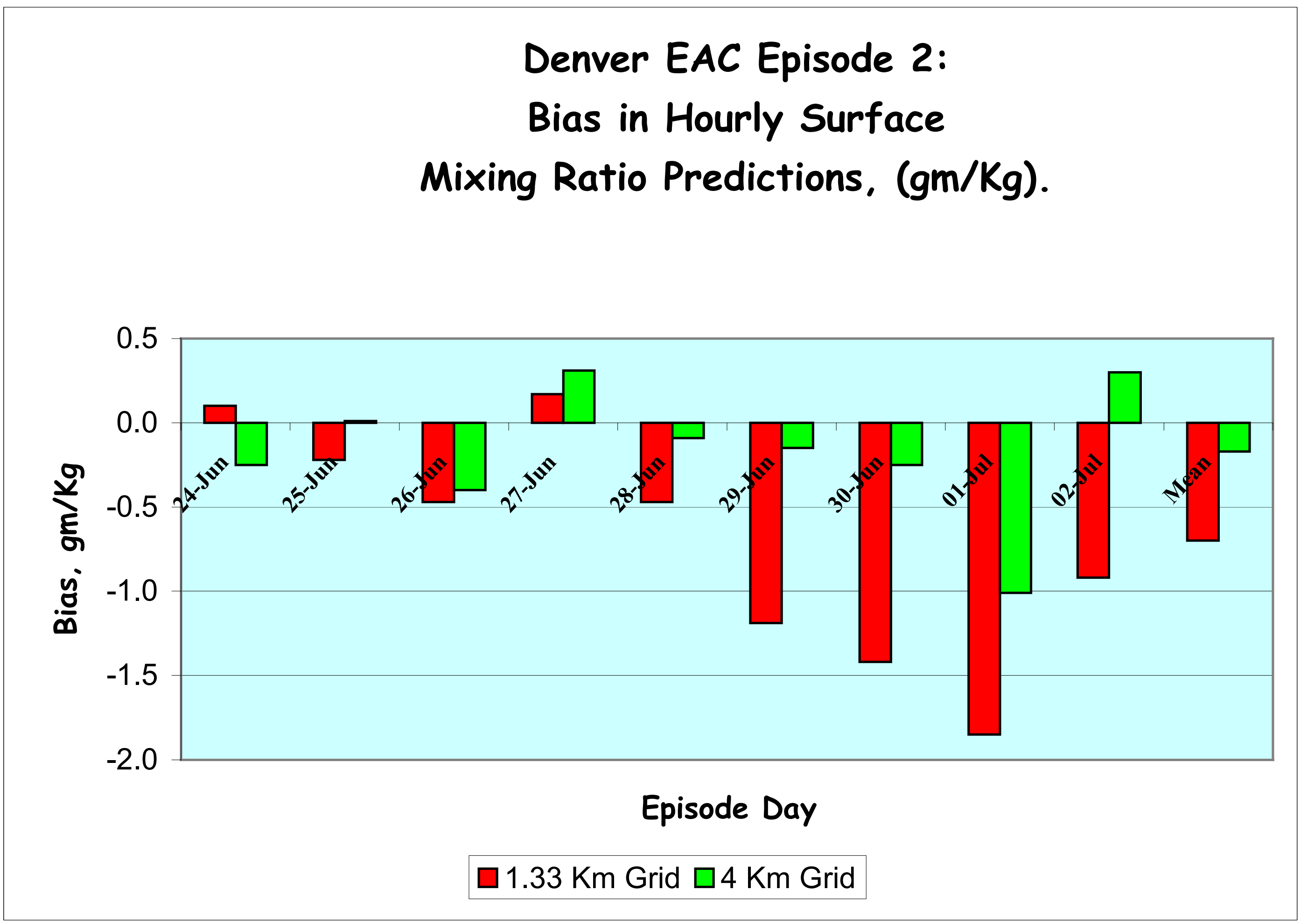


Figure 10-4. Bias In MM5 Hourly Surface Mixing Ratio (gm/Kg) for Episode 2.

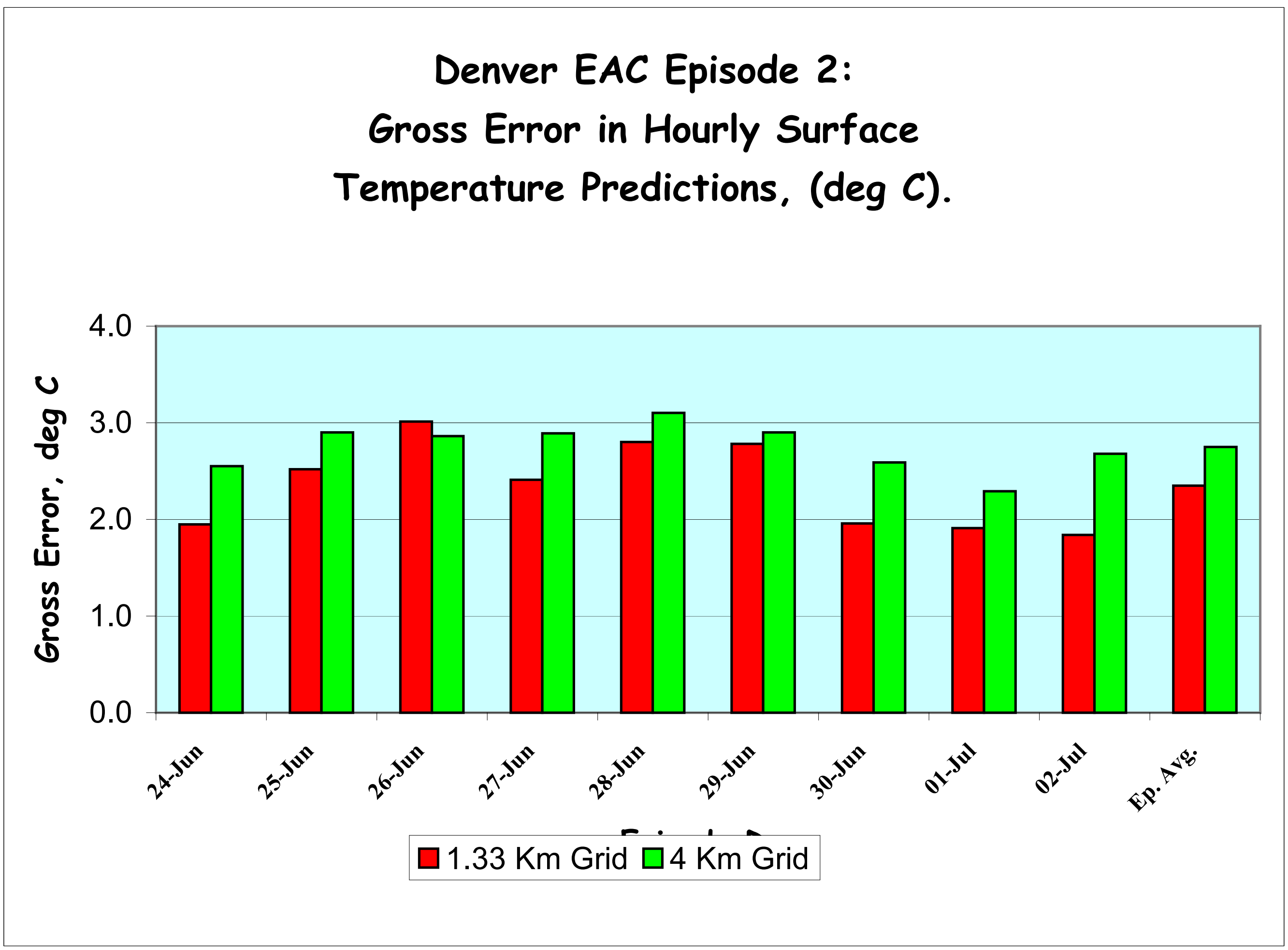


Figure 10-5. Gross Error In MM5 Hourly Surface Temperatures (deg C) for Episode 2.

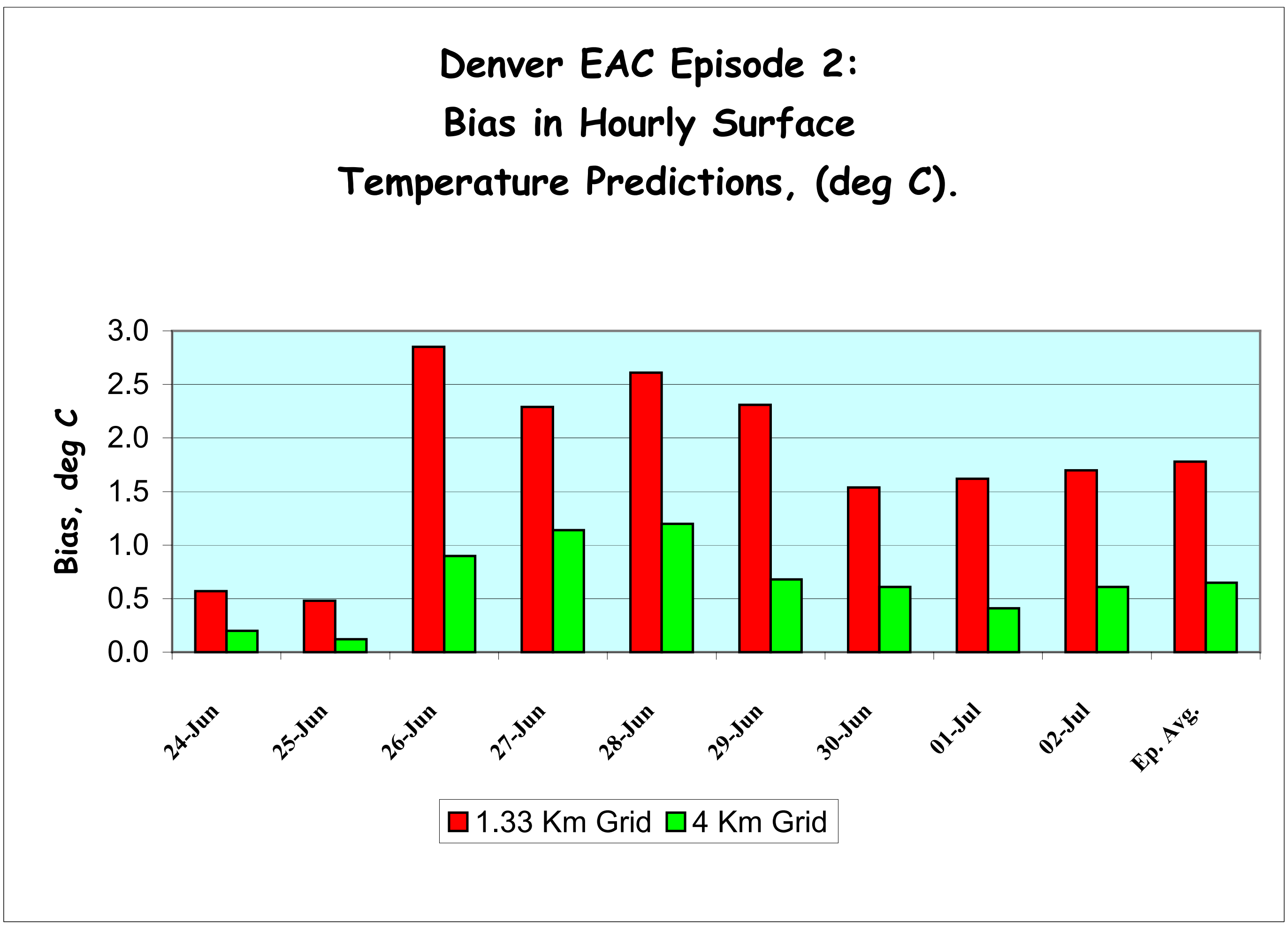
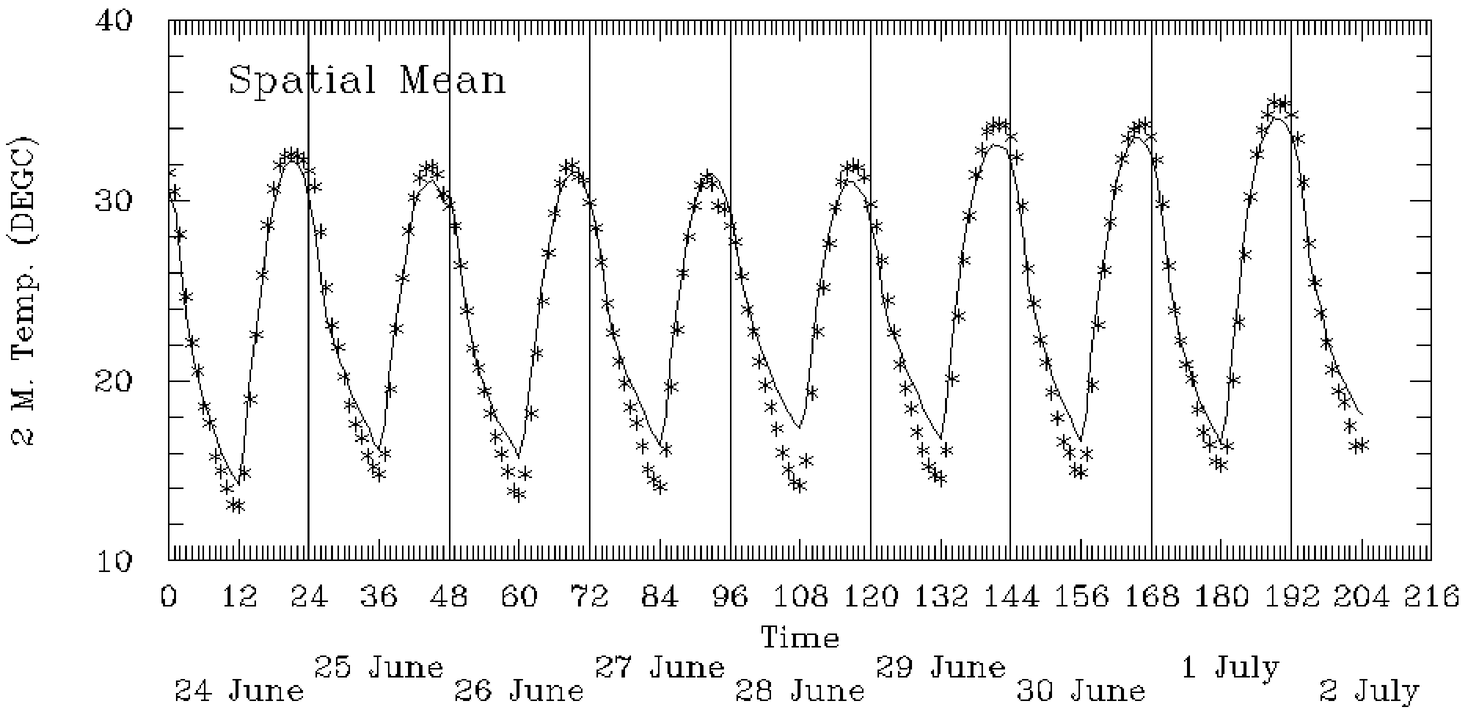
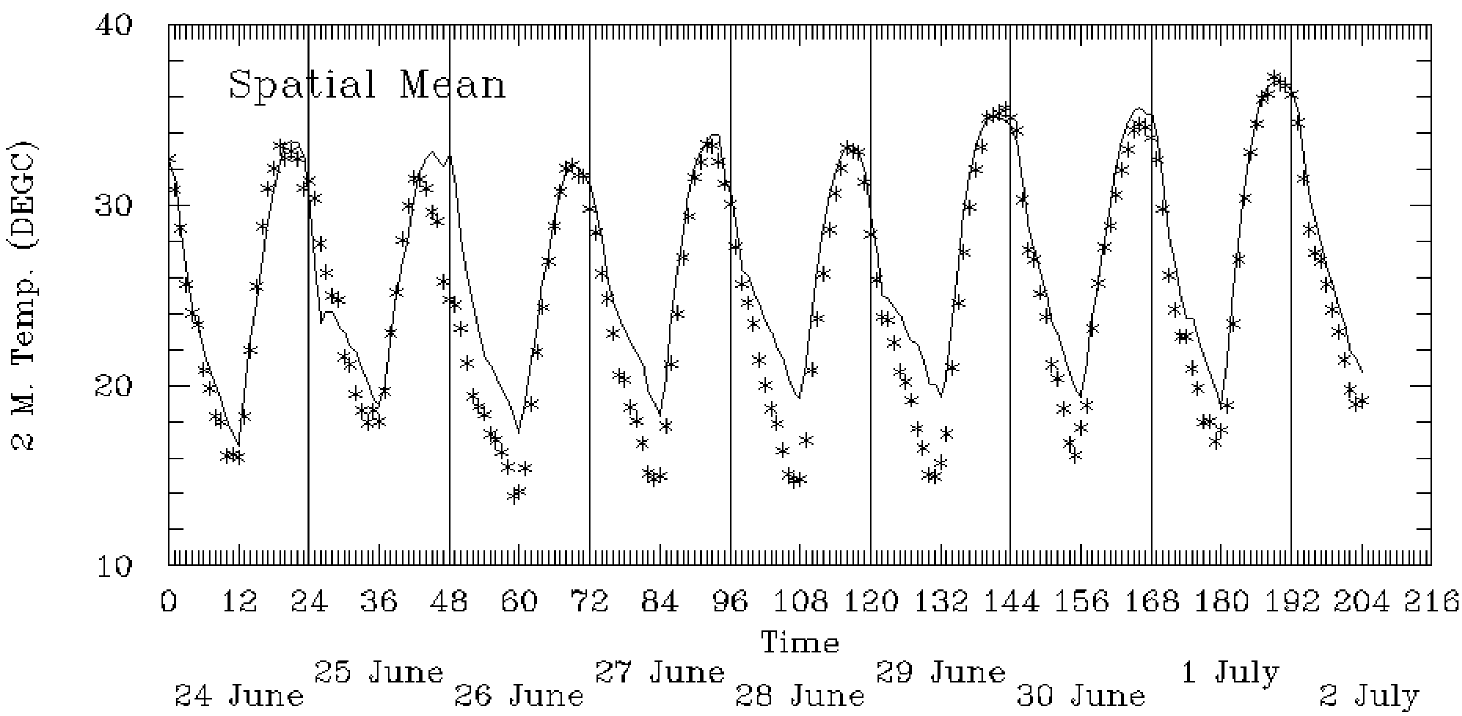


Figure 10-6. Bias In MM5 Hourly Surface Temperatures (deg C) for the Episode 2.



(a) 4 Km Grid



(b) 1.33 Km Grid

Figure 10-7. Diurnal Variation in Spatial Mean Surface Temperatures for Episode 2.

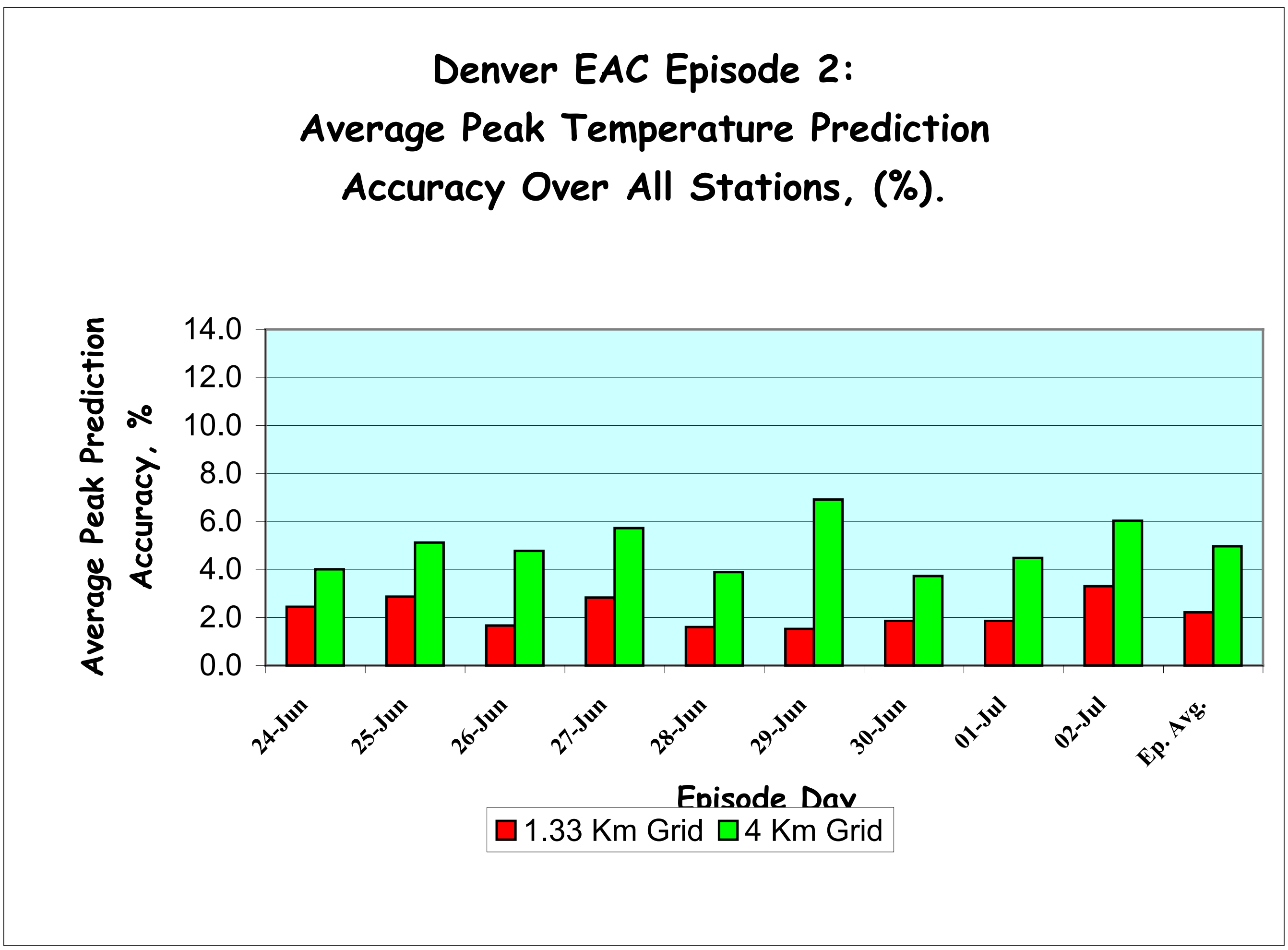


Figure 10-8. Average Peak Prediction Accuracy Over All Monitors for MM5 Hourly Surface Temperatures (deg C) for Episode 2.

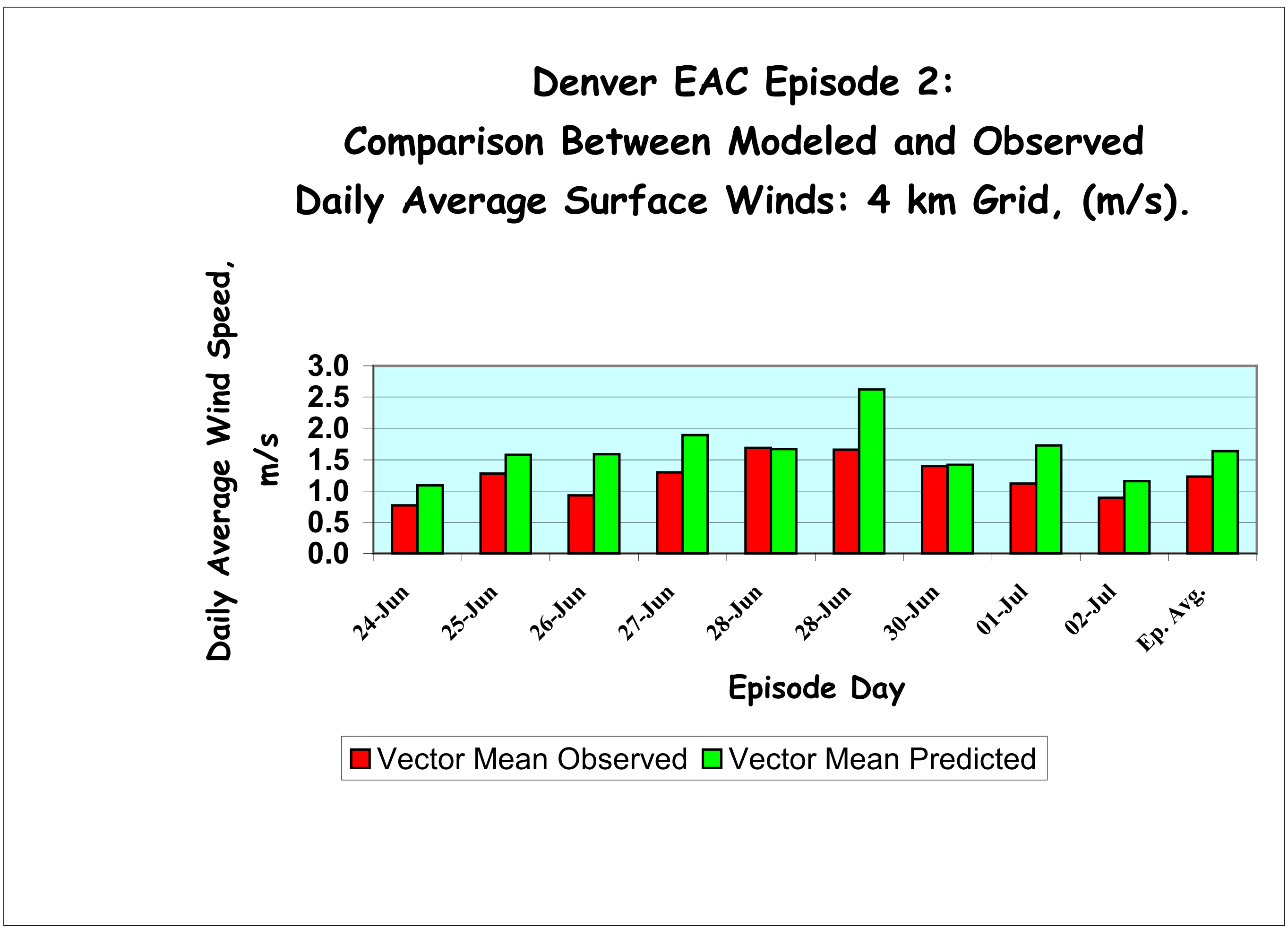


Figure 10-9. Daily Average Modeled and Observed Surface Winds (m/s) on the 4 km Grid for Episode 2.

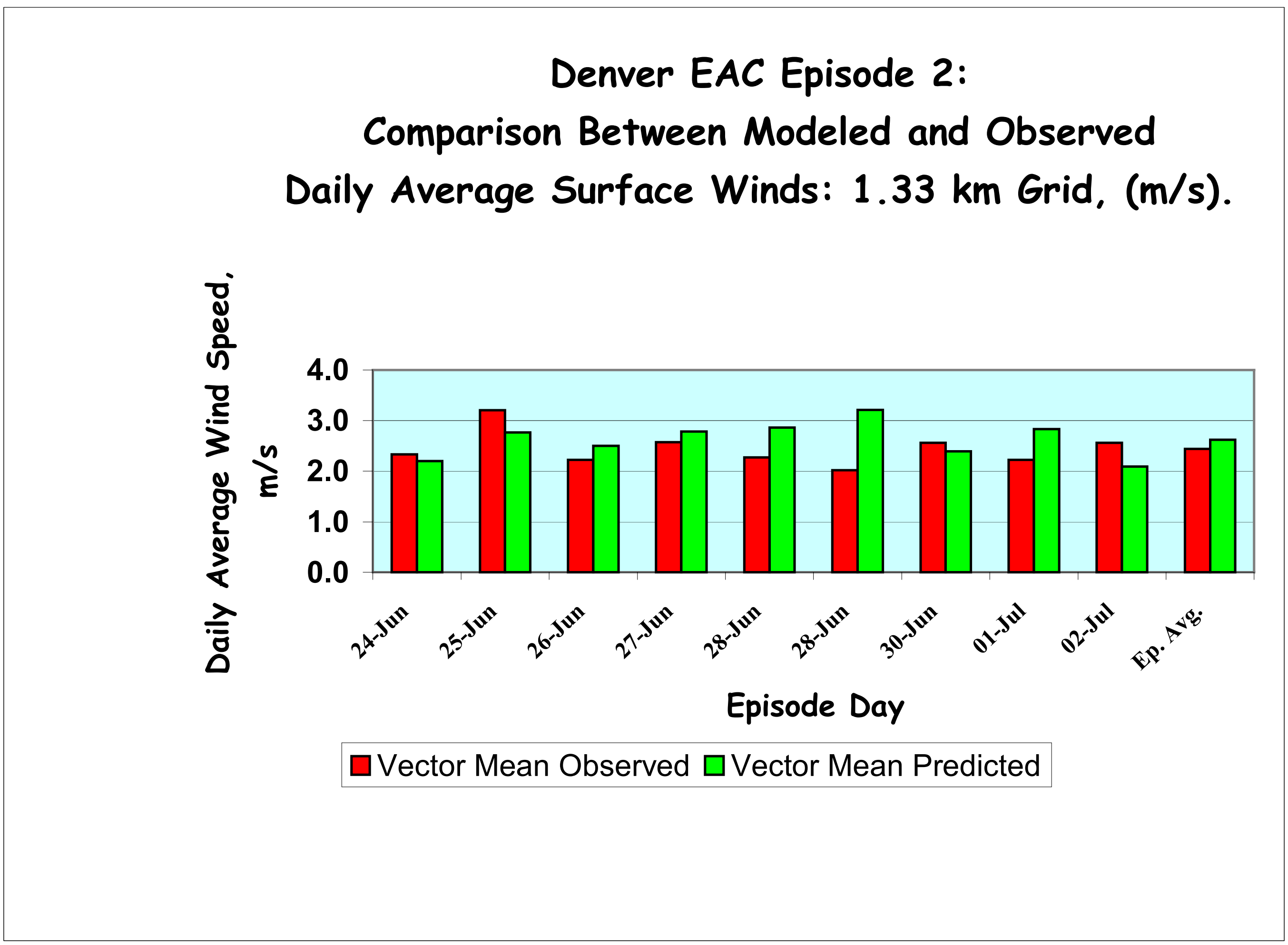


Figure 10-10. Daily Average Modeled and Observed Surface Winds (m/s) on the 1.33 km Grid for Episode 2

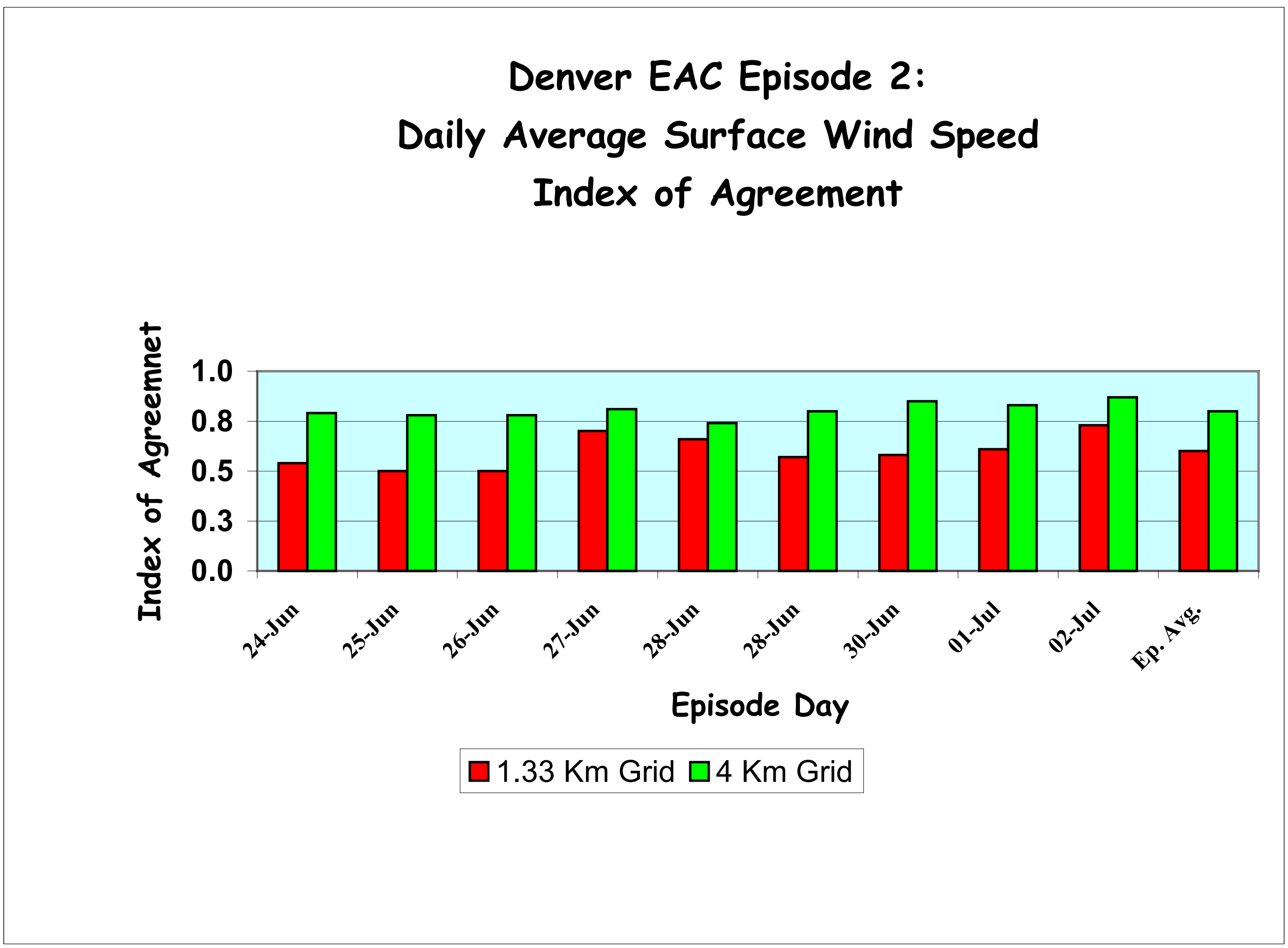


Figure 10-11. Daily Average Surface Wind Index of Agreement for Episode 2.

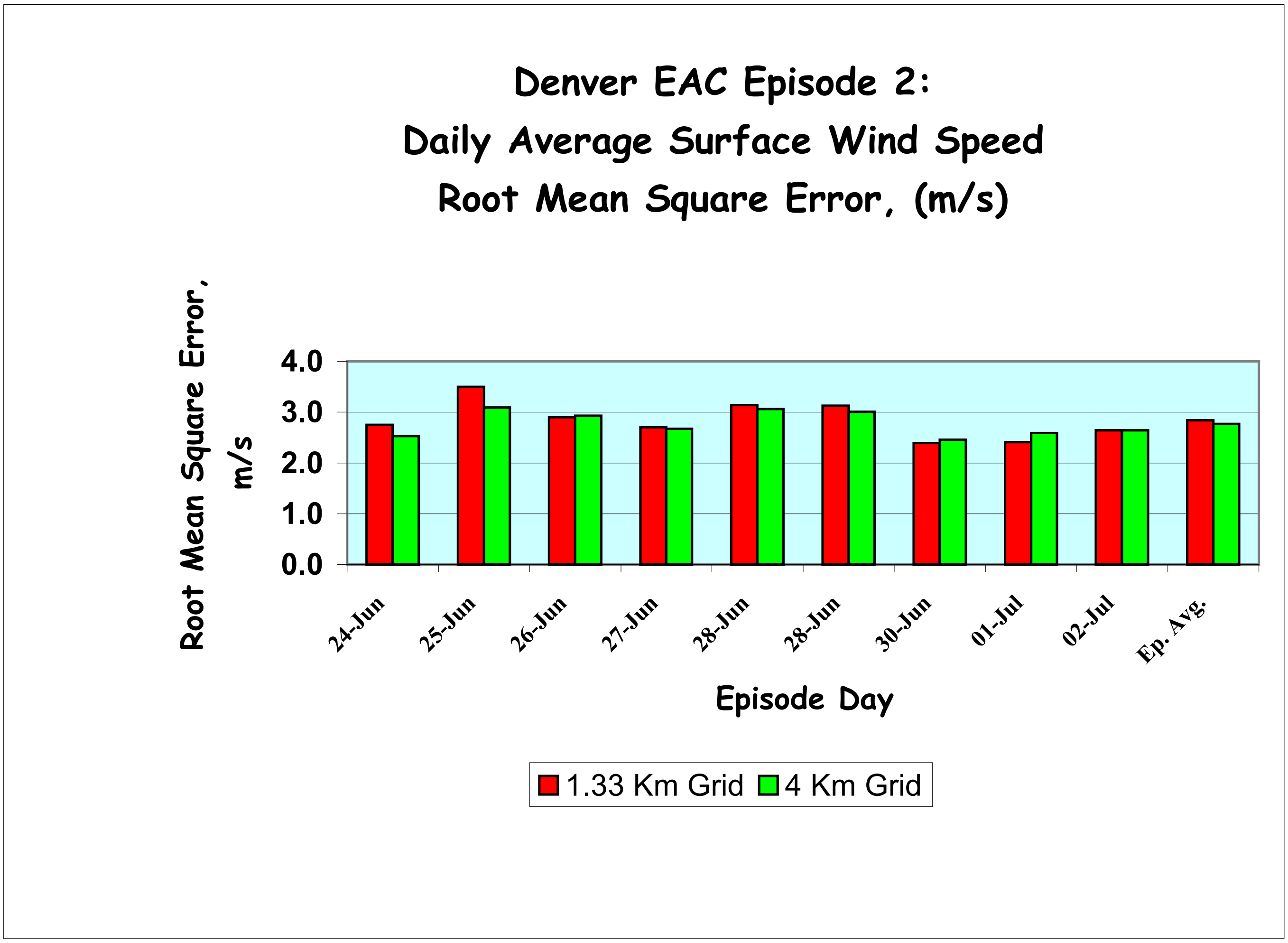


Figure 10-12. Daily Average Surface Wind Speed Root Mean Square Error (m/s) for Episode 2.

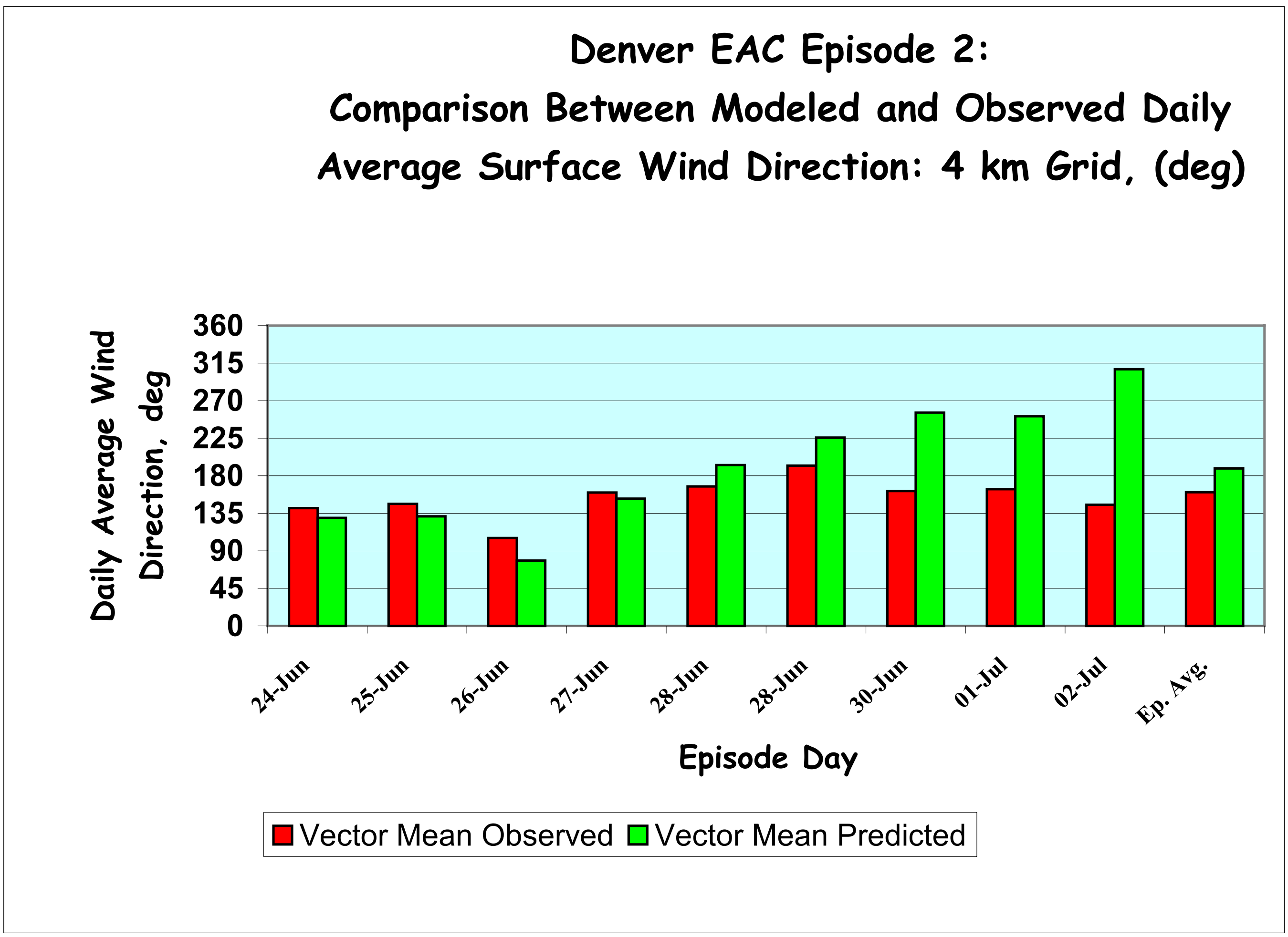


Figure 10-13. Daily Average Modeled and Observed Surface Wind Direction (deg) on the 4 Km Grid for Episode 2.

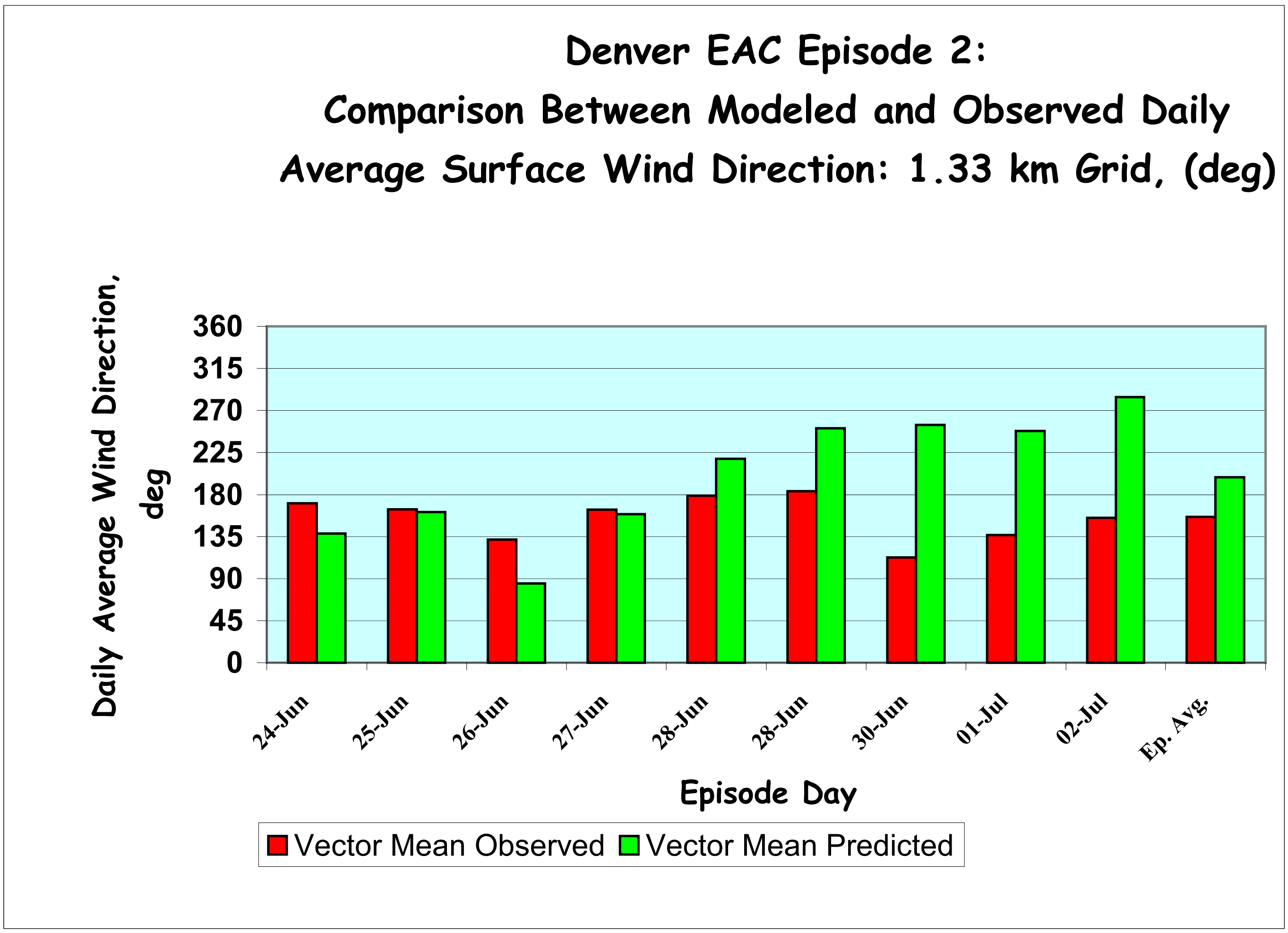


Figure 10-14. Daily Average Modeled and Observed Surface Wind Direction (deg) on the 1.33 Km Grid for Episode 2.

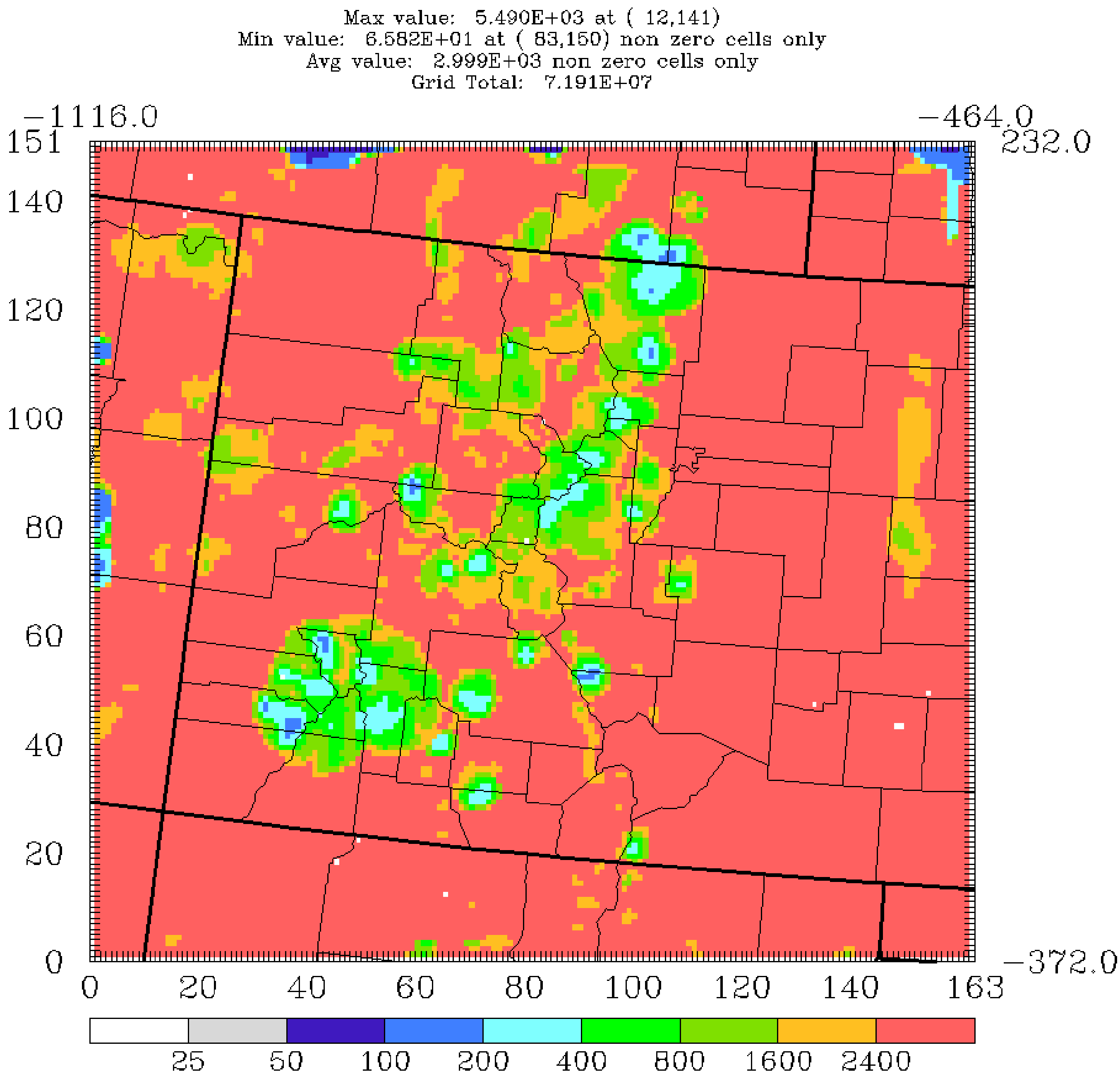


Figure 10-15. Planetary Boundary Layer Heights (m) at 1400 MDT on 28 June 2002 Over the 4 km Grid.

Table 10-1. MM5 Temperature MPE for the Denver EAC Ozone Episode 2: 4/1.3 km Grids.

1.33 Km Grid Domain													
Date	Day	ATS	AS	AT	AU	A-MEAN	N. Bias	Bias	N. Error	Error	Var	Max. O	Max. P
24-Jun	175	-4.96	-4.96	-4.35	-0.27	2.44	4.81	0.57	9.97	1.95	7.59	37.22	37.12
25-Jun	176	-15.62	-3.42	-11.10	1.85	2.86	3.90	0.48	11.18	2.52	11.96	36.11	36.78
26-Jun	177	-2.23	-1.80	-1.41	8.45	1.66	15.64	2.85	16.17	3.01	6.36	33.89	36.76
27-Jun	178	-2.52	-1.69	-1.87	1.40	2.82	12.66	2.29	13.05	2.41	3.92	36.11	36.61
28-Jun	179	-7.51	-5.70	-4.23	7.43	1.60	14.52	2.61	15.14	2.80	5.79	35.00	37.60
29-Jun	180	-3.32	-2.92	-2.67	0.96	1.52	13.77	2.31	15.37	2.78	7.25	37.78	38.14
30-Jun	181	-2.37	-1.46	-1.75	0.69	1.86	7.87	1.54	9.41	1.96	4.31	37.22	37.48
01-Jul	182	-4.17	-3.95	-2.05	1.91	1.85	8.46	1.62	9.48	1.91	4.38	38.89	39.63
02-Jul	183	0.19	1.49	0.90	2.94	3.30	8.26	1.70	8.77	1.84	2.89	37.78	38.89
Ep. Avg.	999	-4.72	-2.71	-3.17	2.82	2.21	9.99	1.78	12.06	2.35	6.05	38.89	39.63
4 Km Grid Domain													
Date	Day	ATS	AS	AT	AU	A-MEAN	N. Bias	Bias	N. Error	Error	Var	Max. O	Max. P
24-Jun	175	-1.99	-0.78	0.81	2.73	4.00	5.92	0.20	15.03	2.55	20.39	37.78	38.81
25-Jun	176	-3.42	-1.19	-0.44	1.95	5.12	4.75	0.12	15.00	2.90	24.52	37.78	38.52
26-Jun	177	-4.30	-4.30	-0.77	-0.29	4.77	7.17	0.90	15.31	2.86	20.93	38.89	38.78
27-Jun	178	-55.98	-55.98	0.88	3.28	5.71	9.11	1.14	15.62	2.89	21.34	37.22	38.44
28-Jun	179	-59.25	-52.24	-0.51	2.69	3.89	11.01	1.20	17.78	3.10	23.53	37.78	38.80
29-Jun	180	-54.92	-53.91	-2.61	0.12	6.90	7.83	0.68	15.85	2.90	24.82	40.00	40.05
30-Jun	181	-3.76	-3.43	-2.87	0.81	3.72	6.68	0.61	13.98	2.59	21.15	39.44	39.76
01-Jul	182	-6.00	-5.33	-2.43	1.27	4.48	4.32	0.41	11.39	2.29	15.49	39.44	39.94
02-Jul	183	-5.77	-3.99	-1.97	1.89	6.02	5.90	0.61	13.80	2.68	19.15	38.89	39.62
Ep. Avg.	999	-21.71	-20.13	-1.10	1.61	4.96	6.97	0.65	14.86	2.75	21.26	40.00	40.05

Table 10-2. MM5 Mixing Ratio MPE for the Denver EAC Ozone Episode 2: 4/1.3 km Grids.

Date	Day	ATS	AS	AT	AU	A-MEAN	N. Bias	Bias	N. Error	Error	Var	Max. O	Max. P
1.33 Km Grid Domain													
24-Jun	175	-55.37	-47.94	-51.10	12.26	30.98	4.53	0.10	25.19	1.60	4.19	12.38	13.90
25-Jun	176	-52.75	-47.70	-51.23	18.03	19.79	-0.95	-0.22	16.87	1.22	2.48	11.39	13.44
26-Jun	177	0.55	17.16	3.21	37.99	9.20	-6.13	-0.47	16.95	1.34	2.63	11.03	15.22
27-Jun	178	4.39	13.61	6.41	18.24	34.26	1.82	0.17	25.30	1.88	5.36	11.48	13.57
28-Jun	179	-37.82	-10.17	-31.99	12.24	22.25	-5.01	-0.47	17.54	1.33	2.77	11.55	12.96
29-Jun	180	-38.01	-38.01	-36.54	-10.46	15.79	-14.37	-1.19	21.39	1.62	2.26	12.03	10.77
30-Jun	181	24.96	47.61	32.82	55.08	32.95	-22.11	-1.42	29.56	1.96	3.66	9.94	15.41
01-Jul	182	-21.36	-4.59	-17.51	4.86	30.06	-25.40	-1.85	30.15	2.22	4.54	13.79	14.46
02-Jul	183	6.43	10.93	7.21	13.65	21.86	-11.86	-0.92	25.70	1.93	5.48	13.32	15.14
Mean	999	-18.78	-6.57	-15.41	17.99	24.13	-8.83	-0.70	23.18	1.68	3.71	13.79	15.41
4 Km Grid Domain													
Date	Day	ATS	AS	AT	AU	A-MEAN	N. Bias	Bias	N. Error	Error	Var	Max. O	Max. P
24-Jun	175	-59.01	-47.51	-49.89	69.49	34.65	-3.17	-0.25	30.58	1.71	5.11	12.38	20.98
25-Jun	176	-8.80	-4.27	-6.23	36.27	23.87	3.43	0.01	27.21	1.50	3.81	14.87	20.26
26-Jun	177	-13.80	-7.30	-5.28	36.14	17.54	-5.06	-0.40	26.44	1.59	4.06	13.11	17.85
27-Jun	178	-20.67	-17.75	-8.03	13.08	24.22	5.90	0.31	28.35	1.79	5.49	15.81	17.88
28-Jun	179	-42.69	3.97	0.06	36.22	18.30	1.02	-0.09	19.03	1.27	2.87	12.64	17.22
29-Jun	180	-62.47	-34.23	-60.25	29.21	24.18	2.20	-0.15	26.42	1.59	3.90	14.62	18.89
30-Jun	181	-21.32	12.35	20.97	65.16	31.35	-2.98	-0.25	26.70	1.51	4.06	12.22	20.18
01-Jul	182	-1.01	-1.01	2.21	43.88	24.91	-14.62	-1.01	25.42	1.63	3.61	13.87	19.96
02-Jul	183	-2.17	4.74	0.56	39.14	23.40	7.50	0.30	28.18	1.68	4.79	14.40	20.04
Mean	999	-25.77	-10.11	-11.76	40.96	24.71	-0.64	-0.17	26.48	1.59	4.19	15.81	20.98

Table 10-3. MM5 Surface Wind MPE for the Denver EAC Ozone Episode 3: 4/1.3 km Grids.

			1.33 Km Grid Domain				
Date	Day	VMOBS	VMEST	RMSE	IA	OBSDIR	ESTDIR
24-Jun	175	2.33	2.20	2.75	0.54	170	138
25-Jun	176	3.20	2.76	3.50	0.50	164	161
26-Jun	177	2.22	2.50	2.90	0.50	132	85
27-Jun	178	2.57	2.78	2.70	0.70	164	159
28-Jun	179	2.27	2.86	3.14	0.66	179	218
28-Jun	180	2.02	3.21	3.13	0.57	184	251
30-Jun	181	2.56	2.39	2.39	0.58	112	254
01-Jul	182	2.22	2.83	2.41	0.61	137	248
02-Jul	183	2.56	2.09	2.64	0.73	155	284
Ep. Avg.	999	2.44	2.62	2.84	0.60	156	199
			4 Km Grid Domain				
Date	Day	VMOBS	VMEST	RMSE	IA	OBSDIR	ESTDIR
24-Jun	175	0.77	1.09	2.53	0.79	141	129
25-Jun	176	1.28	1.58	3.09	0.78	146	131
26-Jun	177	0.93	1.59	2.93	0.78	105	78
27-Jun	178	1.30	1.89	2.67	0.81	160	153
28-Jun	179	1.69	1.67	3.06	0.74	167	193
28-Jun	180	1.66	2.62	3.01	0.80	192	226
30-Jun	181	1.40	1.42	2.46	0.85	162	256
01-Jul	182	1.12	1.73	2.59	0.83	164	251
02-Jul	183	0.89	1.16	2.64	0.87	145	308
Ep. Avg.	999	1.23	1.64	2.77	0.80	160	189

Table 10-4. Summary Results for the 24 June-2 July 2002 MM5 Simulation on the 4/1.33 km High Resolution Grids Compared with the Ad Hoc Performance Benchmarks and Fifty Recent Prognostic Model Performance Evaluations Throughout the U.S.

Episode Grid Resolution	Temperature, deg C		Mixing Ratio, kg/KG		Surface Winds, (m/s)		
	Bias	Error	Bias	Error	RMSE	I	WD diff
4 km	0.65	2.75	-0.17	1.59	2.77	0.80	29
1.33 km	1.78	2.35	-0.70	1.68	2.84	0.60	43
Benchmark	≤ ± 0.5	≤ 2.0	≤ ± 1.0	≤ 2.0	≤ 2.00	≥ 0.60	≤ 30
U.S. Average	-0.2	2.0	0.0	1.8	2.00	0.71	24

11.0 ADEQUACY OF THE MM5 FIELDS FOR CAMx AIR QUALITY MODELING

The central question addressed in this report concerns whether the MM5 meteorological fields are adequate for supporting the ozone modeling in the Denver EAC study. For the reasons discussed below, we believe this question can be answered positively, based on the significant body of information that has been developed particularly when examined in the context of other regulatory ozone studies in this country. This information includes not only the traditional comparisons of model statistical measures against past performance in similar applications but also the use of newer 'weight of evidence' examinations that seek to assess whether the overall modeling activity was performed in a sound and rigorous manner, consistent with the state-of-science in regional prognostic model applications in support of regulatory decision-making. We believe this information will prove useful to CDPHE and RAQC decision-makers in their efforts to assess the overall reliability and credibility of the atmospheric modeling results for public policy making purposes. As explained in this chapter, we believe the MM5 modeling results presented in earlier section of this report support the conclusion that the meteorological simulations are indeed suitable for use in the CAMx photochemical modeling although a number of important scientific questions have not been fully resolved within the time and resources available in this EAC study.

To begin, there is no simple way to answer the question of whether the MM5 fields are adequate as input to the CAMx photochemical model. First, there are no regulatory-approved performance benchmarks for prognostic meteorological models that, if passed, would allow one to declare the MM5 fields appropriate for use. For complex atmospheric modeling problems like the ones being addressed in this study, it is quite doubtful that a set of quantitative performance criteria will ever be completely sufficient. The question of meteorological data set adequacy depends, at a minimum, upon the specific host emissions and air quality models (EPS2x, CAMx in this instance) and the nature of the modeling episodes being used. Meteorological fields that might be adequate for use in the UAM-V model for an OTAG episode, for example, may be quite deficient in an ozone/PM episode for a Regional Planning Organization since the specific needs of the air quality model and the particular chemical and physical processes that must be simulated are different. Thus, quantitative statistical and graphical performance criteria, though helpful, are inherently insufficient in aiding modelers and decision-makers in deciding whether meteorological fields are adequate for air quality modeling. Other considerations must be brought to bear. Below, we present and then work through a process whereby the adequacy of the MM5 fields for use in the Denver EAC 8-hr ozone modeling can be evaluated. This process builds upon the more general evaluation process outlined by Roth, Tesche and Reynolds (1998) and recent suggestions by Tesche et al., (2001, 2002) and Emery et al., (2001) on potentially useful model performance benchmarks.

11.1 Framing the Questions to Be Addressed

Usually air quality simulations are quite sensitive to meteorological fields. Where this sensitivity is anticipated, it is important to make an effort to develop as accurate a representation of meteorological variables as possible. Special features of the flow fields, such as eddies, nocturnal jets, drainage flows, land-sea or land-bay breezes, and vertical circulations should be adequately characterized through the meteorological modeling. In circumstances where there are significant transitions in the meteorological variables over short distances, such as along shorelines or in areas of hilly terrain, the need for finer spatial resolution than is typically specified must be considered. If inadequate attention and care are accorded meteorological modeling, there is a significant risk of developing an inaccurate representation that will be propagated into the emissions and air quality models.

Several questions should be addressed for the specific application. Examples of these questions are as follows:

Appropriateness of Model Selection:

> **Modeling Requirements**: Was a carefully written characterization made of the most important physical and chemical processes relevant to successful air quality modeling of each episode (e.g., a "conceptual model" of each simulation period)?

> **Model Selection:** Did the model selection process ensure that a suitable modeling system was chosen, properly weighing the need for mature, well-tested, publicly-available model(s) against the constraints of the specific modeling problem, characteristics of the episodes to be simulated, and the limitations of schedule and resources?

> **Model Formulation Review:** Was a rigorous evaluation and inter-comparison made between the scientific formulation of the proposed meteorological modeling system (source codes plus pre- and post-processors) versus alternative contemporary prognostic models via an open, thorough scientific review process?

> **Code Verification**: Was the fidelity of the computer coding of the proposed model confirmed with respect to its scientific formulation, governing equations, and numerical solution procedures?

Identification of Air Quality Model Needs:

> **Air Quality Input Needs:** Were the meteorological input needs of the host air quality model and supporting emissions models (e.g., biogenic, motor vehicle, area source processors) clearly identified including specification of the requisite averaging times and nested grid scales for the specific modeling episodes?

> **Air Quality Model Sensitivities:** Was the air quality model's sensitivity to key meteorological inputs established through careful consideration (including air quality model sensitivity/uncertainty simulations) of the relevant modeling episodes over the specific domain of interest? Was the effect of uncertainty in those meteorological inputs to which the air quality model is demonstrated to be most sensitive adequately defined through appropriate numerical experiments or from previous relevant studies?

Note: Identification of air quality model needs is a crucial step in the meteorological model evaluation process, yet it is most often performed superficially if at all. Pragmatic constraints of time and resources necessitate that efforts be directed at achieving the best possible meteorological performance for those variables that matter most to the overall accuracy and reliability of the air quality model. There is little practical benefit to be gained in devoting considerable time to improving the accuracy of a particular meteorological variable if the air quality model – in the specific application at hand -- is insensitive to that variable. Particular attention should be given to those meteorological variables that have the largest uncertainty and to which the air quality model is most sensitive. This challenge can be particularly formidable when dealing with photochemical/aerosol models whose concentration and/or deposition estimates depend on several meteorological variables (mixing, transport, thermodynamic properties, precipitation) simultaneously.

Availability of Supporting Data Bases:

> **Adequate Data Available:** Were sufficient data available to test, at the ground and aloft and over all nested grid scales of importance, the model's dynamic, thermodynamic, and precipitation-related fields?

> **All Data Used**: Was the full richness of the available data base actually utilized in the input data file development, in FDDA, and in the evaluation of model performance?

Note: One of the main considerations underlying selection of modeling episodes for regulatory decision-making is the availability of special data collection programs to supplement the surface and aloft data routinely available from state and federal agencies. While attempts are made to select modeling episodes that coincide with intensive field measurement programs, in these situations it is common that the full set of supplemental measurements are not used thoroughly in the model input development and performance testing phases. At times, the availability of 'high-resolution' databases is touted in support of a particular episode selection choice yet when the modeling is actually performed and evaluated, only a fraction of the special studies data are actually used. This is most notably the case with air quality and meteorological data collected by aloft sampling platforms. Unless the high-resolution data are actually used to enhance the modeling and performance testing, their value is severely limited. Equally troublesome, selection of other candidate modeling days (supported by only routine information) may be overlooked which might otherwise be preferable modeling periods if a concerted effort to utilize special studies data is not made. Finally, as desirable as having supplemental meteorological measurements might be, unless the sampling was performed in the correct regions and includes the variables of primary importance to the air quality model, their potential to add meaningfully to the rigor of the modeling exercise will be limited. Thus, when judging the value of supplemental measurement programs, it is necessary to look beyond just their mere existence (relative to non-intensively monitored days); one must establish that these intensive data set indeed contribute to improved model performance and increased reliability. This necessitates a feedback loop to the air quality modeling exercise to ensure that the times, locations, and parameters associated with the supplemental measurements truly add to the overall quality and rigor of the study.

Results of Operational, Diagnostic, and Scientific Evaluations:

> **Full Model's Predictive Performance:** Was a full suite of statistical measures, graphical procedures, and phenomenological explorations performed with each of the model's state variables and diagnosed quantities for each pertinent grid nest to portray model performance against available observations and against model estimates from other relevant prognostic simulation exercises?

> **Performance of Individual Modules:** Was there an adequate evaluation of the predictive performance of individual process modules and preprocessor modules (e.g., advection scheme, sub-grid scale processes, closure schemes, planetary boundary layer parameterization, FDDA methodology)?

> **Diagnostic Testing:** Were sufficient and meaningful diagnostic, sensitivity, and uncertainty analyses performed to assure conformance of the meteorological modeling system with known or expected behavior in the real world?

> **Mapping Methods:** Were parallel evaluations made of: (a) the output from the prognostic model and (b) the output from the 'mapping' routines that interpolate the prognostic model output onto the host air quality model's grid structure? Were any important differences between the two reconciled?

> **Quality Assurance:** Was a credible quality assurance (QA) activity implemented covering both the prognostic modeling activity as well as the mapping programs that generate air quality-ready meteorological inputs? Was the full set of hourly, three-dimensional fields examined for reasonableness even though observational data for comparison were lacking or in short supply?

Note: Such an intensive performance evaluation process is rarely, if ever, carried out due to time, resource and data base limitations. Nevertheless, it is useful to identify the *ideal* evaluation framework so that the results of the *actual* evaluation can be judged in the proper perspective. This also allows decision-makers to establish realistic expectations regarding the level of accuracy and reliability associated with the meteorological and air quality modeling process.

Comparison with Other Relevant Studies:

> **Comparisons with Other Studies:** Were the model evaluation results (statistical, graphical, and phenomenological) compared with other similar applications of the same and alternative prognostic models to identify areas of commonality and instances of differences between modeling platforms?

Note: Reflecting limited data sets for performance testing and reliable criteria for judging a model's performance, meteorological model evaluations in recent years have emphasized comparisons with other RAMS and MM5 simulations over various modeling domains and episode types as a means of broadening the scope of the evaluation. While this insight into the model's performance – when gauged against other similar applications – is useful, caution must attend such comparisons which at times are at best anecdotal. Often the reporting of previous evaluations entails grossly composited performance statistics (episode averages or averages across episodes, for example), data bases and modeling efforts of widely varying and often unreported quality, different mathematical definitions of statistical quantities, and so on. Thus, these comparisons with other studies, while occasionally providing useful perspective, are by no means sufficient for declaring a meteorological model's performance to be reliable and acceptable in a particular application. Moreover, meteorological model evaluation benchmarks developed on the basis of such historical evaluation studies must also be applied thoughtfully with these limitations in mind.

Peer Review of Specific Modeling Exercise(s):

> **Scope of Peer Review**: Was an adequate, properly-funded, independent, in-depth peer review of the model set-up, application, and performance evaluation efforts conducted?

> **Findings of Peer Review**: Was the effort judged acceptable by the peer-review?

Note: Prognostic modeling requires considerable attention to detail, careful identification of options, and complete involvement in the work. Even with this commitment, critical aspects of a modeling exercise may be treated inadequately or overlooked, most often as the result of schedule or resource constraints. Consequently, an examination of the meteorological modeling effort conducted at arm's length by individuals with appropriate expertise and who have no personal involvement in the work can be essential to avoiding inadvertent oversights and problems. Such a peer review of the effort provides another check on the work as a whole. If concerns are raised about the reliability of the modeling, yet meteorological modeling results are to be used in applying air quality models despite these concerns (e.g., due to project schedule demands), the peer review can assist in suggesting to decision-makers the weight to be given the overall air quality results the planning and management context.

Often, when a professional paper is written describing the modeling study, it undergoes "peer review" by the journal. Such efforts do not constitute the review suggested here. Journal peer review usually entails a reading of the paper, thoughtful reflection, and written commentary, perhaps a 4- to 12-hour effort. Moreover, reporting in the professional literature is necessarily condensed, and much of the detail that should be scrutinized is omitted. This is especially true for complex atmospheric modeling projects. Peer review for pre-print volumes (e.g., American Meteorological Society or Air and Waste Management Association conferences) is even less rigorous, often consisting of a cursory reading of the paper by the Session Chairperson. Peer review, as used here, refers to detailed examination and evaluation of the work conducted by experts in the field. Such experts are generally, but not limited to, those with considerable direct experience in the development, evaluation, and application of the same or very similar meteorological models. This in depth review entails the independent scientists (a) thoroughly examining the conceptual model(s) and modeling protocols prepared for the study, (b) obtaining and examining the details of the model input and output files, and (c) in many cases even running the pre- and post-processor codes and the main simulation programs to corroborate reproducibility of results and to explore inevitable technical issues that arise in such comprehensive reviews. In essence, peer review refers to immersing oneself in the materials provided. Such an effort can take *several weeks* to carry out properly.

Overall Assessment:

> **Overall Reasonableness**: Has an adequate effort been made to evaluate the quality of representation of meteorological fields generated using the meteorological model, as revealed by the full suite of statistical, graphical, phenomenological, diagnostic, sensitivity, and uncertainty investigations? What were the strengths and limitations of the actual model performance evaluation?

> **Fulfillment of Air Quality Model Needs**: How well are the fields represented, particularly in areas and under conditions for which the air quality model is likely to be sensitive?

> **Appropriate Model**: Was a sound and suitable meteorological modeling system adopted?

> **Adequate Data Base**: Was the supporting database adequate to meet input and evaluation needs?

> **Adequate Application Procedures:** Was Four Dimensional Data Assimilation (FDDA) a part of the overall modeling approach and were sufficient data available to support the activity adequately?

> **Quality Assurance**: Were error-checking procedures instituted, followed, and the results reported?

> **Performance Evaluation:** Were suitable procedures specified and adopted for evaluating the quality (e.g., accuracy, precision, and uncertainty) of model estimates?

> **Judging the Overall Process**: Were the criteria (i.e., benchmarks) used to judge performance appropriate for the specific air quality model application, rigorously applied, and properly communicated?

11.2 Comparison of MM5 Performance Against Newly Proposed Meteorological Model Performance Benchmarks

As discussed previously in Chapter 8, there are no currently accepted performance criteria for prognostic meteorological models. Establishment of such criteria, unless accompanied with a careful evaluation process such as the one outlined in this section might lead to the misuse of such goals as is occasionally the case with the accuracy, bias, and error statistics recommended by EPA for judging photochemical dispersion models. In spite of this concern, there remains nonetheless the need for some benchmarks against which to compare new prognostic model simulations.

In Chapter 8 we identified the ad hoc benchmarks currently being used in the U.S. for regulatory ozone modeling studies. At the conclusions of Chapters 8, 9, and 10 we offered specific comparisons between the benchmarks and the MM5 modeling result for the entire Summer '02 episode at 36/12 km scale and Episodes 1 and 2 at 4/1.33 km scale. We also compared the modeling results against model performance in 50 other regulatory modeling studies in the U.S. Table 11-1 re-presents these results of the Denver EAC MM5 episode average statistical results (for those statistics that were produced in this study) compared against the ad hoc meteorological modeling benchmarks and historical performance levels. Cells in the table shaded gray correspond to those episodes and meteorological variables that fall outside of the benchmark ranges. (Recall that, strictly speaking, the benchmarks and the historical summaries pertain to modeling performed on grid ranging from roughly 4-12 km; thus the results on the 36 km and 1.33 km scales are presented largely for completeness).

From the Table 11-1, it is clear that the MM5 application to the Summer '02 episode in general and the two key intensive episodes in particular did not pass all of the ad hoc benchmarks and fell somewhat outside of the mean historical performance ranges of other regulatory studies. Several points are worth making. First, the technical challenges of modeling mesoscale summertime ozone episodes over the Colorado Front Range are significant and may, in fact, be more vexing than simulations in other parts of the country. To be sure, the Denver region is not influenced by local land-sea or land-lake breeze circulations which are obviously a great challenge. But the area is influenced by extreme topographic forcing and the occurrence of afternoon convective storms on some days which make meteorological modeling a challenge in most situations. Second, the period being modeled (Summer '02) was one of the driest periods on record (Waple and Lawrimore, 2003) and extreme drought conditions throughout the state induced widespread atypical soil moisture levels. Soil moisture, a critical MM5 input that directly influences the accuracy of the surface temperature predictions, is poorly characterized even in the best studies and in the present application we believe that a significant part of the temperature discrepancies are related to this parameter. Given adequate time and resources, through model sensitivity experimentation and diagnosis, we believe improved temperature simulations can be achived. However, this level of applied meteorological modeling research is clearly well beyond the scope and schedule constraints of an Early Action Compact photochemical modeling analysis.

Finally, while several of the model parameters (e.g., temperature bias, error, RMSE error in wind speed, and wind direction) fall somewhat outside of the mean of the performance statistics based on the historical RAMS and MM5 evaluations we have performed, the current Denver results for all measures nonetheless remain within the envelop of prognostic model performance that has been judged adequate for regulatory modeling studies elsewhere in the U. S. Therefore, just because a particular statistical measure falls outside a benchmark or the mean value from 50 other studies does not necessarily mean that the simulation is fatally flawed and cannot be used for regulatory purposes. Additional 'weight of evidence' information is needed arrive at this judgment and this is the subject we turn to next.

11.3 Weight of Evidence Assessment of the Denver EAC MM5 Application

Table 11-2 presents the results of our effort to judge the adequacy of the MM5 meteorological modeling process for the Denver EAC study and the specific results for: (a) the Summer '02 period, (b) Episode 1, and (c) Episode 2 against the set of two-dozen questions raised in the preceding section. Our overall conclusions about the adequacy of the MM5 modeling and the reliability of the meteorological fields supplied to the EPS2x emissions and CAMx photochemical models are as follows:

- The Denver EAC meteorological modeling activity selected an appropriate regional prognostic model for use in the assessment;

- The MM5 modeling was carried out in a logical, sound, well-documented manner that was consistent with good scientific principles and the procedures commonly used in the application of this sophisticated model;

- The suite of evaluation procedures employed to test the MM5 model were comprehensive and reflected several different model testing perspectives;

- The data base available to test the MM5 model was limited, precluding a number of meaningful, stressful tests of the model to ascertain whether it suffers from internal, compensating errors; as the result, model testing was confined principally to an operational evaluation;

- Generally, the MM5 performance for surface and aloft winds, temperatures, mixing ratios, and precipitation are consistent with contemporary modeling experience and with new proposed evaluation benchmarks;

- None of the performance testing results conducted have revealed serious flaws in MM5 performance of such a magnitude as to clearly indicate the presence of errors that would render the model inappropriate for use as input to regional air quality models.

We conclude that the MM5 meteorological fields may be used with appropriate cautions as input to the regional emissions and photochemical models for each of the Denver 8-hr EAC ozone episodes.

Table 11-1. Overall Summary of MM5 Performance, Benchmarks, and Previous Experience in Regulatory Modeling Studies.

Episode Grid Resolution	Temperature, deg C		Mixing Ratio, kg/KG		Surface Winds, (m/s)		
	Bias	Error	Bias	Error	RMSE	I	WD diff
Summer '02: 12 km	-1.9	3.0	0.5	1.9	2.34	0.85	27
Summer '02: 36 km	-0.8	2.1	-0.1	1.5	2.21	0.88	19
Episode 1: 4 km	0.45	2.30	-0.66	1.57	2.61	0.78	60
Episode 1: 1.33 km	0.81	1.60	-0.59	1.47	2.53	0.57	37
Episode 2: 4 km	0.65	2.75	-0.17	1.59	2.77	0.80	29
Episode 2: 1.33 km	1.78	2.35	-0.70	1.68	2.84	0.60	43
Benchmark	≤ ± 0.5	≤ 2.0	≤ ± 1.0	≤ 2.0	≤ 2.00	≥ 0.60	≤ 30
U.S. Average	-0.2	2.0	0.0	1.8	2.00	0.71	24

Table 11-2. Weight of Evidence Assessment of the MM5 Fields As Input to CAMx for the Denver EAC.

No.	Question	Assessment
	Appropriateness of Model Selection	
1	Was a careful written characterization made of the most important physical and chemical processes relevant to successful air quality modeling of each episode (e.g., a "conceptual model" of each simulation period)?	No.
2	Did the model selection process ensure that a suitable modeling system was chosen, properly weighing the need for mature, well-tested, publicly-available model(s) against the constraints of the specific modeling problem, characteristics of the episodes to be simulated, and the limitations of schedule and resources?	Yes. The other alternative model, RAMS, is proprietary.
3	Was a rigorous evaluation and inter-comparison made between the scientific formulation of the proposed meteorological modeling system (source codes plus pre- and post-processors) versus alternative contemporary prognostic models via an open, thorough scientific review process?	No. We are not aware of any detailed comparisons being performed between MM5 and alternative models (e.g., RAMS) including their respective pre- and post-processor systems. Model selection was based on general attributes of the MM5 model, its public availability, and the extensive experience of the model's use in supplying inputs to the CAMx air quality model
4	Has the fidelity of the proposed model's computer code's scientific formulation, governing equations, and numerical solution procedures been confirmed?	The MM5 modeling system is well established with a rich development and refinement history spanning more than two decades. The model has seen extensive use worldwide by many agencies, consultants, university scientists and research groups. Thus, the current version of the model and its predecessor versions have been extensively "peer-reviewed" and considerable algorithm development and module testing has been carried out with all of the important process components.

Table 11-2. Continued.

No.	Question	Assessment
	Identification of Air Quality Model Needs	
5	Were the meteorological input needs of the host air quality model and supporting emissions models (e.g., biogenic, motor vehicle, area source processors) clearly identified including specification of the requisite average times and nested grid scales for the specific modeling episodes?	Since the MM5 model has been used successfully before with both EPS2x and CAMx, there was no apparent need to conduct a separate study at the beginning of the Denver EAC study to re-affirm this.
6	Was the air quality model's sensitivity to key meteorological inputs established through careful consideration (including air quality model sensitivity/uncertainty simulations) of the relevant modeling episodes over the specific domain of interest? Was the effect of uncertainty in those meteorological inputs to which the air quality model is demonstrated to be most sensitive adequately defined through appropriate numerical experiments or from previous relevant studies?	No. Study resource constraints precluded this analysis.
	Availability of Supporting Data Bases	
7	Were sufficient data available to test, at the ground and aloft and over all nested grid scales of importance, the model's dynamic, thermodynamic, and precipitation-related fields?	No. Data were adequate to set up, operate, and evaluate (operationally) the MM5 model with standard surface and aloft NWS data sets and other information from established surface reporting networks . No supplemental meteorological data sets, particularly for aloft processes, were available.
8	Was the full richness of the available data base actually utilized in the input data file development, in FDDA, and in the evaluation of model performance?	No. Some supplemental data sets were available.

Table 11-2. Continued.

No.	Question	Assessment
	Results of Operational, Diagnostic, and Scientific Evaluations	
9	Was a full suite of statistical measures, graphical procedures, and phenomenological explorations performed with each of the models state variables and diagnosed quantities for each pertinent grid nest to portray model performance against available observations and predictions from other relevant prognostic modeling exercises?	Yes, for the most part. An extensive set of operational evaluation statistics and graphical displays were produced focusing on point comparisons, residual analyses, and comparisons between spatial fields of measurements and predictions. The operational evaluations were carried out at all MM5 spatial scales but the level of analysis and reporting varied from one grid scale to the next. The 4 and 1.33 km scales received the greatest attention.
10	Was there an adequate evaluation of the predictive performance of individual process modules and preprocessor modules (e.g., advection scheme, sub-grid scale processes, closure schemes, planetary boundary layer parameterization, FDDA methodology)?	No. Lack of data to perform these experiments and Denver EAC schedule and resource allocations prevented these explorations from being carried out.
11	Were sufficient and meaningful diagnostic, sensitivity, and uncertainty analyses performed to assure conformance of the meteorological modeling system with known or expected behavior in the real world?	Very limited. Detailed diagnostic sensitivity experiments were performed with a few episodes (especially the first one) but subsequently, little diagnostic or sensitivity experimentation was performed as the result of schedule and budget considerations.
12	Were parallel evaluations made of (a) the output from the prognostic model and (b) the output from the 'mapping' routines that interpolate the prognostic model output onto the host air quality model's grid structure? Were sources of differences between the two reconciled?	No. The Denver EAC schedule and resource constraints precluded an in-depth comparison of "raw" MM5 output fields vs. the CAMx-ready meteorological fields resolved to the air quality model grid mesh.
13	Was a credible quality assurance activity implemented covering both the prognostic modeling activity as well as the mapping programs that generate air quality-ready meteorological inputs? Was the full set of hourly, three-dimensional fields examined for reasonableness even though observational data for comparison are lacking or in short supply?	Partially. Quality assurance activities consisted principally of routine plotting of surface fields, calculation of summary statistics (to reveal outliers or anomalies), and related graphical display methods to provide a cursory check of the model inputs and outputs. However, once the final set of data preparation procedures were established, the files were constructed mostly in a hands-off manner. Quality assurance activities of the MM5 output fields was performed as an integral part of the statistical and graphical performance examinations.

Table 11-2. Continued.

No.	Question	Assessment
	Comparison with Other Relevant Studies	
14	Were the model evaluation results (statistical, graphical, and phenomenological) compared with other similar applications of the same and alternative prognostic models to identify areas of commonality and instances of differences between modeling platforms?	Partially. Episode average statistics over the 12 km grid were compared with over 50 RAMS and MM5 model applications elsewhere in the U.S. (including the Denver/Northern Front Range Region), primarily involving summertime ozone episodes with typical grid scales in the 4– 16 km range. No detailed comparisons were performed between the Denver EAC embedded or Super Summer '02 episode and others reported in the literature to elucidate areas of similar performance and areas of disparate performance.
	Peer Review of Specific Modeling Exercise(s)	
15	Was an adequate, independent, in-depth peer review of the model set-up, application, and performance evaluation efforts conducted?	No. Study resource constraints precluded an in-depth peer review.
16	Was the effort judged acceptable by the peer-review?	Not applicable since no formal peer review was performed.

Table 11-2. Continued.

No.	Question	Assessment
	Overall Assessment	
17	Has an adequate effort been made to evaluate the quality of representation of meteorological fields generated using the meteorological model, as revealed by the full suite of statistical, graphical, phenomenological, diagnostic, sensitivity, and uncertainty investigations? What were the strengths and limitations of the actual model performance evaluation?	Generally yes. A rich variety of analytical procedures, statistical metrics and graphical tools were used. All of the statistical and graphical presentation methods have been used extensively and effectively in past evaluations reported broadly in the literature.
18	How well are the fields represented, particularly in areas and under conditions for which the air quality model is likely to be sensitive?	Uncertain. The Denver EAC schedule and resources precluded detailed diagnosis and model performance improvement exercises with all nine episodes. There was very little opportunity for a sustained cycle of MM5 diagnosis and performance improvement, followed by an investigation of the CAMx model response, producing yet another round of meteorological model diagnosis and performance improvement.
19	Was a sound and suitable meteorological modeling system adopted?	Yes. The MM5 model used in the Denver EAC application is clearly representative of the state-of-the-science in mesoscale prognostic models suitable for air quality applications.
20	Was the supporting database adequate to meet input and evaluation needs?	No. While the available data base was sufficient to set up, exercise, and evaluate operationally the model, it was clearly deficient in supporting rigorous testing, aimed at identifying potential sources of internal, compensating errors.

Table 11-2. Concluded.

No.	Question	Assessment
	Overall Assessment	
21	Was Four Dimensional Data Assimilation (FDDA) a part of the overall modeling approach and were sufficient data available to support the activity adequately?	Partially. The routinely available data were sufficient to utilize data assimilation in the MM5 simulations. However, lack of high-resolution data (e.g., radar wind profilers) and budget/time constraints precluded the infusion of this information into the routine FDDA methodologies that were ultimately used.
22	Were error-checking procedures instituted, followed, and the results reported?	Partially.
23	Were suitable procedures specified and adopted for evaluating the quality (e.g., accuracy, precision, and uncertainty) of model estimates?	Partially. Useful statistical measures and graphical procedures were employed to quantify performance for key dynamic and thermodynamic variables (e.g., bias, gross errors, root-mean-square-error, Index of Agreement, skill scores). However, little quantitative information was produced relative to model uncertainty. No formal uncertainty analysis was conducted of the MM5 simulations. Hence, the estimates of "uncertainty" in the MM5 outputs are based on the ranges in the various statistics (e.g., the range in the model's surface temperature bias over the three embedded ozone episodes) as compared to quantitative estimate of model uncertainty arising due to: (a) formulation, (b) procedures for developing inputs or processing outputs, and (c) measurement error and spatial representativeness issues.
24	Were the criteria used to judge performance appropriate for the specific air quality model application, rigorously applied, and properly communicated?	Partially. A credible effort was made to identify the key components of the meteorological model evaluation process and to address each one subject to the constraints of project schedule, resources, and the information available from the meteorological and air quality modeling activities. While the present Denver EAC MM5 evaluation is arguably more comprehensive and systematic than previous prognostic model evaluation study supporting 1-hr or 8-hr ozone air quality applications, many areas of the evaluation were not adequately explored, principally as the result of these limitations.

12.0 SUMMARY AND CONCLUSIONS

12.1 Summary

This report describes the results of a meteorological model evaluation study carried out as part of the Denver Early Action Compact (EAC) Study, described in detail in the modeling protocol by Tesche et al., (2003a). As part of the Denver EAC study, the Mesoscale Meteorological Model (MM5) was applied to a fifty-day long summer ozone period in central Colorado spanning the 6 June-25 July 2002 timeframe. Within the Summer '02 episode, three embedded high 8-hr ozone air pollution episodes occurred in the Denver-Northern Front Range Region (DNFRR):

> Episode 1: (18-21 July 2002);

> Episode 2: (25 June–1 July 2002); and

> Episode 3: (8-12 June 2002).

Nested MM5 meteorological simulations were performed by modelers at Alpine Geophysics in technical consultation with staff at ENVIRON International Corporation (the modeling prime contractor). In this report, we have presented the results of an operational and limited scientific evaluation of the MM5 model for the Summer '02 episode and the first two intensive embedded periods.[1]

We have assessed the model's performance in simulating three-dimensional fields of wind, temperature, and moisture (i.e. mixing ratio) using a combination of statistical measures and benchmarks, graphical tools, and more qualitative 'weight of evidence' considerations for one primary purpose: *to judge the adequacy of the meteorological results as input to regulatory 8-hr ozone modeling for the Denver EAC.* This has been accomplished, in part, by comparing the MM5's performance in simulating the two primary episodes with results from fifty other recent regulatory modeling studies carried out in U.S. using the MM5 (or similar models) in direct support of 1-hr or 8-hr ozone NAAQS decision-making.

12.2 Conclusions

There are no currently accepted performance criteria for prognostic meteorological models used in regulatory decision-making. In this study we have utilized recently proposed *ad hoc* benchmarks and evaluation results from fifty recent regulatory ozone modeling studies to assess the current MM5 modeling results developed for the Summer '02 episode at 36/12 km scale and Episodes 1 and 2 at 4/1.33 km scale. The MM5 application to the Summer '02 episode and the two primary 8-hr modeling episodes exceeded many but not all of the *ad hoc* statistical benchmarks. In other instances, the model's performance for temperature and/or wind composite statistical measures fell somewhat outside of the typical performance levels achieved in other regulatory studies. Through diagnostic analyses, we have attributed these performance issues to: (a) the technical challenges of modeling mesoscale ozone episodes over the Colorado Front Range, (b) the extreme topography, (c) the occurrence of one of the driest periods on record and the concomitant widespread atypical soil moisture levels, and (d) the effect of drought

[1] The results for Episode 3 will be presented when the technical analyses and evaluations are completed later this summer.

conditions on modeled surface temperature predictions. Given adequate time and resources, thorough model sensitivity experimentation and diagnosis, we believe that the current MM5 simulations could be improved somewhat from their current performance levels. However, given that this level of applied meteorological modeling research is clearly well beyond the scope and schedule constraints of an Early Action Compact photochemical modeling analysis and the fact that, in our judgment, the present MM5 model simulations are quite sufficient for regulatory 8-hr ozone modeling, we believe that the subsequent emissions and photochemical modeling tasks in the Denver study can proceed on schedule using these inputs.

Finally, although certain of the MM5 statistical measures such as temperature bias and error, wind speed RMSE error, wind direction error fall somewhat outside the *average* performance levels achieved in other regulatory evaluations, the current Denver MM5 results for all measures are still *well within the envelop of prognostic model performance* that has been judged acceptable for 1-hr and 8-hr regulatory ozone modeling studies elsewhere in the U. S. Supplemented with the 'weight of evidence' information summarized in Chapter 11, we conclude that the MM5 meteorological fields may indeed be used as input to the regional emissions and photochemical models for the two high priority Denver 8-hr EAC ozone episodes.

REFERENCES

AG, 1995. “The Emissions Modeling System (EMS-95) User's Guide", Alpine Geophysics, Boulder, CO.

Anthes, R. A., 1977. “A Cumulus Parameterization Scheme Utilizing a One-Dimensional Cloud Model”, Monthly Weather Review, Vol. 105, pp. 270-286.

Anthes, R. A., and T. T. Warner, 1978. “The Development of Mesoscale Models Suitable for Air Pollution and Other Mesometeorological Studies”, Monthly Weather Review, vol. 106, pp, 1045-1078.

Arakawa, A. and W. Schubert, 1974. “Interaction of a Cumulus Cloud Ensemble with the Large Scale Environment, Part I”. J. Atmos. Sci., Vol. 31, pp. 674-701.

Arnold, J. R., R. L. Dennis, and G. S. Tonnesen, 2003. “Diagnostic Evaluation of Numerical Air Quality Models with Specialized Ambient Observations: Testing the Community Multiscale Air Quality Modeling System (CMAQ) at Selected SOS 95 Ground Sites”, ***Atmospheric Environment***, Vol. 37, pp. 1185-1198.

Barchet, W. R., and R. L. Dennis, 1990. “NAPAP Model Evaluation, Volume 1: Protocol”, prepared for the U.S. Environmental Protection Agency, prepared by Battelle Pacific Northwest Laboratories, Richland, WA.

Benjamin, S. G., and N. L. Seaman, 1985. "A Simple Scheme for Objective Analyses in Curved Flow", Mon. Wea. Rev., Vol. 113, pp. 1184-1198.

BLM, 2003. “Farmington Proposed Resource Management Plan and Final Environmental Impact Statement”, Volume I, prepared by the U.S. Department of Interior, Bureau of Land Management, Farmington Field Office, Farmington, NM.

Bott, A., 1989. “A Positive Definite Advection Scheme Obtained by Nonlinear Renormalization of the Advective Fluxes”, Monthly Weather Review, Vol. 117, pp. 1006-1015.

Breiman, L., et al., 1984. Classification and Regression Trees, Wadsworth Press, Belmont, CA.

Byun, D. W., and J. K. S. Ching, 1999. “Science Algorithms of the EPA Models-3 Community Multiscale Air Quality (CMAQ) Modeling System”, EPA/600/R-99/030.

CDPHE, 2003. “Episode Selection for the Denver Early Action Compact”, prepared by the Colorado Department of Health and Environment, Denver, CO.

Coats, C. J., 1995. “Sparse Matrix Operator Kernel Emissions (SMOKE) Modeling System”, MCNC Environmental Programs, Research Triangle Park, NC.

Coella, P., and P. L., Woodward, 1984. “The Piecewise Parabolic Method (PPM) for Gas-Dynamical Simulations”, J. of Computational Physics, Vol. 54, pp. 174-201.

Cooke, G. A., 2002. “Protocol for Early Action Compacts Designed to Achieve and Maintain the 8-Hour Ozone Standard”, EPA Region 6 Administrator, Dallas, TX., letter of 19 June 2002.

Cox, R. et al., 1998. "A Mesoscale Model Intercomparison", Bulletin of the American Meteorological Society, Vo.. 79, No. 2., pp. 265-283.

Dennis, R. L., et al., 1990. "Evaluation of Regional Acid Deposition Models", State-of-Science/Technology Report No. 5, National Acid Precipitation Assessment Program, Washington, D.C.

Douglas, S. G., and A. B. Hudischewskyi, 1999a. "Episode Selection Analysis for 8-Hour Ozone for Selected Areas Along the Eastern Gulf Coast", Systems Applications, Int., San Rafael, CA. (SYSAPP-99/07d)

Douglas, S. G., et al., 1997. "Investigation of the Effects of Horizontal Grid Resolution on UAM-V Simulation Results for Three Urban Areas", prepared for the Southern Company Services and Cinergy Corporation, prepared by Systems Applications, Inc., San Rafael, CA.

Douglas, S. G., et al., 1999b. "Process-Based Analysis of the Role of the Gulf Breeze in Simulating Ozone Concentrations Along the Eastern Gulf Coast". 11th Joint Conference on the Applications of Air Pollution Meteorology with the AWMA, American Meteorological Society, Long Beach, CA, 9-14 January.

Dudhia, J., 1989. "Numerical Study of Convection Observed During the Winter Monsoon Experiment Using a Mesoscale Two-Dimensional Model", J. Atmos. Sci., Vol. 46. pp. 3077-3107.

Dudhia, J., 1993. "A Non-hydrostatic Version of the Penn State/NCAR Mesoscale Model: Validation Tests and Simulation of an Atlantic Cyclone and Cold Front", Mon. Wea. Rev., Vol. 121. pp. 1493-1513.

Dunker A., G. Yarwood, J. Ortmann, and G.M. Wilson. 2002a. "The Decoupled Direct Method for Sensitivity Analysis in a Three-dimensional Air Quality Model – Implementation, Accuracy and Efficiency." Environmental Science and Technology, 36, 2965-2976.
Dunker A., G. Yarwood, J. Ortmann, and G.M. Wilson. 2002b. "Comparison of Source Apportionment and Source Sensitivity of Ozone in a Three-Dimensional Air Quality Model." Environmental Science and Technology, 36, 2953-2964.

Durrenberger, C. J., et al., 1999a. "Regional Photochemical Modeling in Texas", Texas Natural Resources Conservation Commission, Austin, TX.

Durrenberger, C. J., et al., 1999b "Comparison of Performance of Several Photochemical Models", Texas Natural Resources Conservation Commission, Austin, TX.

Emery, C. et al., 1999. "Ozone Modeling for the Kansas City Nonattainment Area: Final Protocol", prepared for the Kansas Department of Health and Environment, prepared by ENVIRON International Corporation and Alpine Geophysics.

Emery, C. et al., 2001. "Enhanced Meteorological Modeling and Performance Evaluation for Two Texas Ozone Episodes", prepared for the Texas Natural Resource Conservation Commission, prepared by ENVIRON International Corporation, Novato, CA.

Emigh, R. A., 1995. "Development of a Draft PM-10 Emissions Inventory in the SARMAP Region Using the EMS-95 Emission Estimates Modeling System". Presented at the International Conference of the Emissions Inventory, Research Triangle Park, NC.

Emigh, R. A., 1997. "The EMS-95 Emissions Modeling Advanced Workshop," prepared for the Victoria Environmental Protection Authority, Melbourne, Australia.

Emigh, R. A., and G. J. Wilkinson, 1995. "Baltimore/Washington, D.C. Emissions Quality Assurance, Volume I: Data Analysis", prepared for the Maryland Department of Environment, prepared by Alpine Geophysics, Boulder, CO.

Emigh, R.A., et al., 1997. "Comparison of CEM-Enhanced Emissions with OTAG Base1c Emissions and the Impact on Ozone Concentrations", Final Report prepared for the Electric Power Research Institute, prepared by Alpine Geophysics, LLC, Boulder, CO.

ENSR, 1993. "Model Code Verification of Air Quality and Meteorological Simulation Models for the Lake Michigan Ozone Study", prepared for the Lake Michigan Air Directors Consortium, prepared by ENSR Consulting and Engineering, Hartford, CN.

ENVIRON, 1998, "Meteorological Types Associated with Gulf Coast Ozone Episodes", prepared for the Minerals Management Service, prepared by ENVIRON, Intl., Novato, CA.

ENVIRON, 2000. "User's Guide to the Comprehensive Air Quality Model with Extensions (CAMx), Version 3.00", ENVIRON International Corporation, Novato, CA.

ENVIRON, 2002. "User's Guide to the Comprehensive Air Quality Model with Extensions (CAMx), Version 3.01", ENVIRON International Corporation, Novato, CA.

ENVIRON, 2003. "User's Guide to the Comprehensive Air Quality Model with Extensions (CAMx), Version 4.00", ENVIRON International Corporation, Novato, CA.

ENVIRON, 2001. "Development of a Joint CAMx Photochemical Modeling Database for the Four Southern Texas Near Non-Attainment Areas". ENVIRON International Corporation, Novato, California. Prepared for Alamo Area Council of Governments (AACOG), San Antonio, Texas. March 12.

ENVIRON, 2002a. "Ozone Modeling Protocol for FY 2000/2001 Projects in the Tyler/Longview/Marshall Area of East Texas". ENVIRON International Corporation, Novato, California. Prepared for East Texas Council of Governments (ETCOG), Kilgore, Texas. August 8.

ENVIRON, 2002b. "Modeling Protocol – Development of A Photochemical Modeling Database to Address 8-Hour ozone Attainment in the Tulsa and Oklahoma City Areas". ENVIRON International Corporation, Novato, California. Prepared for Oklahoma Department of Environmental Quality (ODEQ), Oklahoma City, Oklahoma. August 8.

EPA, 1991. "Guidance for Regulatory Application of the Urban Airshed Model (UAM)". Office of Air Quality Planning and Standards, U.S. Environmental Protection Agency, Research Triangle Park, N.C.

EPA, 1995. "User's Guide for the CALPUFF Dispersion Model", U.S. EPA/OAQPS, EPA454-B-95-006, Research Triangle Park, NC.

EPA, 1999. "Draft Guidance on the Use of Models and Other Analyses in Attainment Demonstrations for the 8-hr Ozone NAAQS". Draft (May 1999), U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards, Research Triangle Park, N.C

EPA, 2000. "EPA Regional and State Acid Rain CEM Contact List," http://www.epa.gov/acidrain/cems/contact.html

EPA, 2001. "Draft Guidance for Demonstrating Attainment of Air Quality Goals for PM2.6 and Regional Haze", U. S. EPA, Office of Air Quality Planning and Standards, Research Triangle Park, NC.

EPA, 2002. "EPA 8-hr Ozone Plan and Guidance Update", prepared by the U.S. EPA Region IV, December 2002.

Gayno, G. A., et al., 1994. "Forecasting Visibility Using a 1.5-Order Closure Boundary Layer Scheme in a 12 Km Non-Hydrostatic Model", 10th AMS Conference on Numerical Weather Prediction, Portland, OR. 18-22 July. American Meteorological Society, Boston, MA.

Gill, D. O., 1992. "A User's Guide to the Penn State/NCAR Mesoscale Modeling System", NCAR Tech. Note 381 +IA, National Center for Atmospheric Research, Boulder, CO, 233 pp.

Green, M. A., et al., 1998. "Mesoscale Transport of Mercury in Southern Florida", 10th Joint Conference on Applications of Air Pollution Meteorology, 11-16 January, Phoenix, AZ.

Grell. G. A., et al., 1991. "Semi-prognostic Tests of Cumulus Parameterization Schemes in the Middle Latitudes", Monthly Weather Review, Vol. 119, pp. 5-31.

Grell, G. A., J. Dudhia, and D. R. Stauffer, 1994. "A Description of the Fifth Generation Penn State/NCAR Mesoscale Model (MM5). NCAR Tech. Note, NCAR TN-398-STR, 138 pp.

Guenther, A., B. Baugh, G. Brasseur, J. Greenberg, P. Harley, L. Klinger, D. Serca, and L. Vierling, 1999a. "Isoprene emission estimates and uncertainties for the Central African EXPRESSO study domain," *J. Geophysical Research*, in press.

Guenther, A., C. Geron, T. Pierce, B. Lamb, P. Harley, and R. Fall, 1999b. "Natural emissions of non-methane volatile organic compounds, carbon monoxide, and oxides of nitrogen from North America," *Atmospheric Environment*, in press.

Hanna, S. R., et al., 1998. "Evaluations of Numerical Weather Prediction (NWP) Models from the Point of View of Inputs Required by Atmospheric Dispersion Models", 5th International Conference on Harmonization within Atmospheric Dispersion Modeling for Regulatory Purposes, 18-21 May, Rhodes, Greece.

Houyoux M. R., and J. M. Vukovich, 1996. "Updates to the Sparse Matrix Operator Kernel Emissions (SMOKE) Modeling System and Integration with Models-3", The Emissions Inventory: Regional Strategies for the Future, Air and Waste Management Association, 26-28 October, Raleigh, NC.

Kain, J. S., and J. M. Fritsch, 1990. "A One-Dimensional Entraining/Detraining Plume Model and Its Application in Convective Parameterization", Journal of Atmospheric Science, Vol. 47, pp. 2784-2802.

Kain, J. S., and J. M. Fritsch, 1993. "Convective Parameterization for Mesoscale Models: The Kain-Fritsch Scheme", The Representation of Cumulus Convection in Numerical Models, Meteor. Monogr., Vol. 46, American Meteorological Society, pp. 165-170.

Kinnee, E., C. D. Geron, and T. E. Pierce, 1997. "United States Land Use Inventory for Estimating Biogenic Ozone Precursor Emissions," *Ecological Applications*, 7, 1:46-58.

Kumar, N. and F. W. Lurmann, 1997. "Peer Review of ENVIRON's Ozone Source Apportionment Technology and the CAMx Air Quality Model", prepared for the Ohio Environmental Protection Agency, prepared by Sonoma Technology, Inc., Santa Rosa, CA.

Kumar, N. et al., 1996. "Development and Application of a Three Dimensional Aerosol Model", Presented at the AWMA Specialty Conference on Computing in Environmental Resource Management, Research Triangle Park, NC.

Kuo, H. L., 1974. "Further Studies of the Parameterization of the Effect of Cumulus Convection on Large-Scale Flow", J. Atmos. Sci., Vol. 31., pp. 1232-1240.

Lehmann, Elfrun, 1998. "The Predictive Performance of the Photochemical Grid Models UAM-V and CAMx for the Northeast Corridor", Air and Waste Management 91st Annual Meeting, San Diego, CA. 14-18 June.

Liu, Gang, C. Hogrefe, and S. T. Rao, 2003, "Evaluating the Performance of Regional-Scale Meteorological Models: Effects of Clouds Simulation on Temperature Prediction", ***Atmospheric Environment***, Vol. 36, pp. 1691-1705.

Loomis, C. F., 1997. "Comparison of CAL-MoVEM and DTIM2 Motor Vehicle Emissions Estimate Models", prepared for the California Air Resources Board, prepared by Alpine Geophysics, LLC, Arvada, CO.

Loomis, C. F., and J. G. Wilkinson, 1996. "Review of Current Methodologies for Estimating Ammonia Emissions", prepared for the California Air Resources Board, prepared by Alpine Geophysics, LLC, Arvada, CO.

Loomis, C. F., and D. E. McNally, 1998. "Cincinnati-Hamilton Ozone Attainment Demonstration Study: Volume 6: Development of Base Year Emissions Estimates for Three Modeling Episodes", prepared for the Ohio Environmental Protection Agency, prepared by Alpine Geophysics, LLC, Arvada, CO.

Loomis, C. F., D. E. McNally, and T. W. Tesche, 1999. "Verification of the EMS-95 Benchmark Simulations of the July 1995 SAMI Oxidant Episode over the Southeastern U.S.", prepared for the Southern Appalachian Mountains Initiative and the Tennessee Valley Authority, prepared by Alpine Geophysics, LLC, Arvada, CO.

Loomis, C. F., et al., 1996. "Pittsburgh Regional Ozone Attainment Study: Volume XI C Emissions Modeling Results", prepared for the Southwestern Pennsylvania Clean Air Stakeholders Group, prepared by Alpine Geophysics, LLC, Ft. Wright, KY.

Loomis, C. F., et al., 1997a. "Protocol for the Development of a 2007 SIP call Emissions Inventory"

(Version 2.0), prepared by Alpine Geophysics, LLC, Arvada, CO.

Loomis, C. F., et al., 1997b. "Development of the 2007 Emissions Inventory Reflecting the EPA Section 110 SIP call Requirements for the Eastern U.S.", prepared by Alpine Geophysics, LLC, Arvada, CO.

Louis, J. F., 1979. "A Parametric Model of Vertical Eddy Diffusivity Fluxes in the Atmosphere", Boundary Layer Meteorology, Vol. 17, pp. 187-202.

Lurmann, F. W., and N. Kumar, 1996. "Development of Chemical Transformation Algorithms for Annual PM-10 Dispersion Models". prepared for the South Coast Air Quality Management District, prepared by Sonoma Technology, Inc., Santa Rosa, CA.

Lurmann, F. W., and N. Kumar, 1997. "Evaluation of the UAM-V Model Performance in OTAG Simulations: Phase I: Summary of Performance Against Surface Observations", prepared for Science Applications International Corporation, prepared by Sonoma Technology, Inc., Santa Rosa, CA.

Loveland, T. R., J. W. Merchant, D. O. Ohlen, and J. F. Brown, 1991. "Development of a land-cover characteristics database for the conterminous US," *Photogrammetric Engineering and Remote Sensing*, 57:1453-1463.

Madronich, S., and G. Weller, 1990. "Numerical Integration Errors in Calculated Tropospheric Photodissociation Rate Coefficients", Journal of Atmos. Chem. Vol. 10, pp. 289-300.

Mansell G.E. and G.M. Wilson. 2002. Technical Memorandum to Tyler Fox U.S. Environmental Protection Agency, RTP, NC. March 1.

Mass, C. F., and Y. H. Kuo, 1998. "Regional Real-Time Numerical Weather Prediction: Current Status and Future Potential", Bulletin of the American Meteorological Society, Vol. 79, No. 2, pp. 253-263.

Maul, P. R., 1980. "Atmospheric Transport of Sulfur Compound Pollutants", Central Electricity Generating Board, MID/SSD/80/0026/R, Nottingham, England.

McNally, D. E., 1997. "Development of Methodology for Mapping MM5 Fields onto Arbitrary Eulerian Photochemical Air Quality Simulation Models (PAQSM)", Alpine Geophysics, LLC, Arvada, CO.

McNally, D. E., and T. W. Tesche, 1994. "MAPS2.3 User's Guide", Alpine Geophysics, LLC, Golden, CO.

McNally, D. E., and T. W. Tesche, 1996a. "Evaluation of the MM5 Model for the 1-11 July 1988 OTAG Episode over the Northeastern United States", prepared for Pennsylvania Power and Light Co., prepared by Alpine Geophysics, LLC, Ft. Wright, KY.

McNally, D. E., and T. W. Tesche, 1996b. "Evaluation of the MM5 Model for the July 1988 and July 1995 Episodes and Comparison with the OTAG Meteorological Model, RAMS", 89th Annual Meeting of the Air and Waste Management Association, 23-28 June 1996, Nashville, TN.

McNally, D. E., and T. W. Tesche, 1996c. "Pittsburgh Regional Ozone Attainment Study: Evaluation of the MM5 Model for Three Episodes", prepared for the Pennsylvania Department of Environmental Protection, prepared by Alpine Geophysics, LLC, Ft. Wright, KY.

McNally, D. E., and T. W. Tesche, 1996d. “Evaluation of the MM5, SAQM, UAM-IV, and UAM-V Models over the Northeast U.S. for Four Ozone Episodes Using Routine and Intensive NARSTO-NE and LMOS Aerometric Data Sets”, First NARSTO-NE Data Analysis Symposium and Workshop, Washington, D.C., 10-12 December.

McNally, D. E., and T. W. Tesche, 1997c. “Modeled Effects of Indiana Point Source NOx Emissions Reductions on Local and Regional 1-hr and 8-hr Ground Level Ozone Concentrations in 1995 and 2007 Using Two OTAG Oxidant Episodes”, prepared for the Indiana Electric Utility Air Workgroup, prepared by Alpine Geophysics, LLC, Arvada, CO.

McNally, D. E., and T. W. Tesche, 1997d. “Comparative Evaluation of the CAMx and UAM-V Models Over the Northeastern U.S. Using the July 1995 OTAG Episode and the NARSTO-NE Intensive Field Study Data”, prepared for the Virginia Department of Environmental Quality, prepared by Alpine Geophysics, LLC, Arvada, CO.

McNally, D. E., and T. W. Tesche, 1998a. “Effects of Indiana Point Source NOx Emissions Reductions on Ground Level Ozone Concentrations Using the 07EPA1a Basecase Inventory”, prepared for the Indiana Electric Association, prepared by Alpine Geophysics, LLC, Ft. Wright, KY.

McNally, D. E., and T. W. Tesche, 1998b. “Comparative Evaluation of the CAMx and UAM-V Models Over the Northeastern U.S. Using the July 1995 OTAG Episode and the NARSTO-NE Intensive Field Study Data”, prepared for the Mid-Atlantic Regional Air Management Association, prepared by Alpine Geophysics, LLC, Arvada, CO.

McNally, D. E., and T. W. Tesche, 1998c. “Evaluation of the MM5 Meteorological Model over the Greater Denver Front Range Region for Two Wintertime Episodes”, prepared for the Denver Regional Air Quality Council, prepared by Alpine Geophysics, LLC, Arvada, CO.

McNally, D. W., and T. W. Tesche, 1999a. “Impact of Stack Parameter Errors on Ground Level Ozone Metrics in the OTAG Domains”, prepared for General Motors Corporation, prepared by Alpine Geophysics, LLC, Arvada, CO.

McNally, D. E., and T. W. Tesche, 1999c. “MM5 Performance Evaluation for the 15-24 June 1995 and Assessment of Model Suitability for 8-Hr Ozone Attainment Demonstration over the Kansas City-St. Louis, Missouri Domain”, prepared for the Kansas Department of Health and Environment and the Missouri Department of Natural Resources, prepared by Alpine Geophysics, LLC, Arvada, CO.

McNally, D. E., T. W. Tesche, and A. G. Russell, 1996. "Comparative Evaluation of the URM and UAM-V Regional Models Over the LMOS Domain Using Two High-Resolution Episodic Data Bases", Ninth Joint Conference on the Applications of Air Pollution Meteorology, American Meteorological Society and the Air and Waste Management Association, 28 January-2 February, 1996, Atlanta, GA.

McNally, D. E., et al., 1996. "Evaluation of the URM, UAM-V, UAM-IV, and ROM2.2 Photochemical Models Over Lower Lake Michigan for Two 1991 LMOS Oxidant Episodes", Ninth Joint Conference on the Applications of Air Pollution Meteorology, American Meteorological Society and the Air and Waste Management Association, 28 January-2 February, 1996, Atlanta, GA.

McNally, D. E., et al., 1997a. "Photochemical Modeling Analysis of the Pittsburgh-Beaver Valley Ozone Nonattainment Area: Volume IV: -- Interim Final Report", prepared for the Southwestern Pennsylvania Clean Air Stakeholders Group, prepared by Alpine Geophysics, LLC, Ft. Wright, KY.

McNally, D. E., et al., 1997b. "Modeled Effects of Indiana Point Source NOx Emissions Reductions on Local and Regional 1-hr and 8-hr Ground-Level Ozone Concentrations in 1995 and 2007 Using Two OTAG Oxidant Episodes", prepared for the Indiana Electric Utility Air Workgroup, prepared by Alpine Geophysics, LLC, Ft. Wright, KY.

McNally, D. E. et al., 1997c. "Comparative Evaluation of the CAMx and UAM-V Models Over the Northeastern U.S. Using the July 1995 OTAG Episode and the NARSTO-NE Intensive Field Study Data", prepared for the Virginia Department of Environmental Quality, prepared by Alpine Geophysics, LLC, Arvada, CO.

McNally, D. E. and T. W. Tesche, 1997d. "Greater Denver Front-Range Meteorological Modeling Protocol: MM5 Modeling Protocol", prepared for the Denver Regional Air Quality Council, prepared by Alpine Geophysics, LLC, Arvada, CO.

McNally, D. E. et al ., 1998a. "Photochemical Modeling Analysis of the Effects of Electric Utility NOx Emissions Reductions in Eastern Missouri on 1-Hr and 8-Hr Ozone Concentrations", prepared for the Missouri Electric Utility Environmental Committee, prepared by Alpine Geophysics, LLC, Arvada, CO.

McNally, D. E., et al., 1998b. "Nested Regional Photochemical Modeling in Support of the Pittsburgh-Beaver Valley Ozone SIP", 10th Joint Conference on the Applications of Air Pollution Meteorology with the Air and Waste Management Association, 11-16 January, Phoenix, AZ.

McNally, D. E., et al., 1998c. "Photochemical Modeling of the Effects of VOC and NOx Emissions Controls in the Baltimore, Washington Ozone Nonattainment Area", prepared for the Maryland Department of Environment, prepared by Alpine Geophysics, LLC, Ft. Wright, KY.

McNally, D. E. and T. W. Tesche, 1998d. "Evaluation of the MM5 Meteorological Model Over the Greater Denver Front Range for Two Wintertime Episodes", prepared for the Denver Regional Air Quality Council, prepared by Alpine Geophysics, LLC, Arvada, CO.

McNally, D. E., and T W. Tesche, 2002. "Annual Meteorological Modeling Protocol (ver 1.0): Annual Application of MM5 to the Continental United States", prepared for the U. S. Environmental Protection Agency, Office of Air Quality Planning and Standards, prepared by Alpine Geophysics, LLC, Arvada, CO.

McNally, D. E., and T W. Tesche, 2003. "Annual Application of MM5 to the Continental United States", prepared for the U. S. Environmental Protection Agency, Office of Air Quality Planning and Standards, prepared by Alpine Geophysics, LLC, Arvada, CO.

Mellor, B. L., and T. Yamada, 1974. "Hierarchy of Turbulence Closure Models for Planetary Boundary Layers", J. Atmos. Sci., Vol. 31, pp. 1791-1806.

Mellor, B. L., and T. Yamada, 1982. "Development of a Turbulence Closure Model for Geophysical Fluid Problems", Review of Geophysics and Space Physics, Vol. 20, pp. 851-875.

Modica, L., et al., 1985. “Flexible Regional Emissions Data System (FREDS) Documentation for the 1985 NAPAP Emissions Inventory”, prepared for the U.S. Environmental Protection Agency, prepared by Alliance Technologies, Chapel Hill, N.C.

Morris, R. E., et al., 1998. “Assessment of the Contribution of Industrial and Other Source Sectors to Ozone Exceedances in the Eastern United States”, Final Report to the Division of Air Pollution Control, Ohio Environmental Protection Agency, prepared by ENVIRON International, Novato, CA.

Morris R.E., C.A. Emery and E. Tai. 2003. “Sensitivity Analysis and Intercomparison of the Models-3/CMAQ and CAMx Models for the July 1995 NARSTO-Northeast Episode” presented at AWMA 2003 Annual Meeting and Exhibition, San Diego, California. June.

Morris, R. E., T. W., Tesche, and F. L. Lurmann, 1999. “Evaluation of the CAMx and MAQSIP Models Over the NARSTO-NE Region with Inputs from the MM5 Model”, prepared for the Coordinating Research Council, prepared by ENVIRON International Corporation, Alpine Geophysics, and Sonoma Technology.

Morris, R. E., et al., 2002. “Model Evaluation and Sensitivity of the Models-3/CMAQ and CAMx Modeling Systems for the July 1995 NARSTO-Northeast Episode, Models-3/CMAS Workshop, Research Triangle Park, NC, 21 October.
NMED, 2002. “Four Corners Region Ozone”, prepared by the Air Quality Bureau, New Mexico Environment Department, Santa Fe, NM.

OTAG, 1996. “Ozone Transport Assessment Group Modeling Protocol”, Version 3.0, Prepared by the Regional and Urban-Scale Modeling Workgroup, Lake Michigan Air Directors Consortium and the EPA Office of Air Quality Planning and Standards, Research Triangle Park, NC.

Pai, P. K., et al., 1998: Modeling Air Pollution in the Los Angeles Basin Using the MM5-SAQM Modeling System: Part II: Air Quality Simulations”, 10th Joint Conference on Applications of Air Pollution Meteorology, 11-16 January, Phoenix, AZ.

Reynolds, S. D., and P. M. Roth, 1997. “Peer Review of the CAMx Ozone Source Apportionment Technology”, Reprot from the EPA Source Attribution Workshop, 16-18 July 1997. U.S. Environmental Protection Agency, Research Triangle Park, N.C.

Pielke, R. A., 1984. Mesoscale Meteorological Modeling, Academic Press, New York, NY.

Pielke, R. and R. Pearce, 1994. Mesoscale Modeling of the Atmosphere, Meteor. Monogr., No. 47, American Meteorological Society, Boston, MA.

Pielke, R. A., and M. Uliasz, 1998. “Use of Meteorological Models as Input to Regional and Mesoscale Air Quality Models C Limitations and Strengths”, Atmospheric Environment, Vol 32, No. 8, pp. 1455-1466.

Pierce, T. E., 1996. “Documentation for BEIS2,” anonymous ftp at ftp://monsoon.rtpnc.epa.gov/pub/beis2/SOS/AAREADME.

Pierce, T. E. and C. D. Geron, 1996. “The personal computer version of the Biogenic Emissions Inventory System (PCBEIS2.2),” anonymous ftp at ftp://monsoon.rtpnc.epa.gov/pub/beis2/pcbeis22.

Pleim, J. D., and J. S. Chang, 1992. “A non-local closure model for vertical mixing in the convective boundary layer”. ***Atmos. Environ***. Vol.26A, pp. 965-981.

Pleim, J. E., et al., 2001. “A coupled land-surface and dry deposition model and comparison to field measurements of surface heat, moisture, and ozone fluxes”. ***Water, Air, and Soil Pollution: Focus***, Vol. 1, pp. 243-252.

Roth, P. M., T. W. Tesche, and S. D. Reynolds, 1998. “A Critical Review of Regulatory Air Quality Modeling for Tropospheric Ozone”, Prepared for the North American Strategy for Tropospheric Ozone and the American Petroleum Institute.

Russell, A. G., and R. L. Dennis, 2000. “NARSTO Critical Review of Photochemical Models and Modeling”, Atmospheric Environment, Vol. 34, No. 12-14, pp. 2283-2324.

SAI, 1990. “User’s Guide for the Urban Airshed Model: Volume IV, Emissions Preprocessor System”, prepared for the U.S. Environmental Protection Agency, prepared by Systems Applications, Inc., San Rafael, CA.

Seaman, N. L., 1995. “Status of Meteorological Pre-Processors for Air Quality Modeling”, International Conf. On Particulate Matter, Air and Waste Mgt. Assn., Pittsburgh, PA.

Seaman, N. L., 1995. “Status of Meteorological Pre-Processors for Air Quality Modeling”, International Conf. On Particulate Matter, Air and Waste Mgt. Assn., Pittsburgh, PA.

Seaman, N. L., 1996. “Study of Meteorological Variables Needed in Air-Quality Modeling’, Annual Progress Report prepared for the Coordinating Research Council (CRC), Interim Report, Project A-11, prepared by the Department of Meteorology, Penn State University, State College, PA.

Seaman, N. L., 2000. “Meteorological Modeling for Air Quality Assessments”, Atmospheric Environment., Vol. 34, No. 12-14, 2231-2260.

Seaman, N. L., and D. R. Stauffer, 1996. "SARMAP Meteorological Model Final Report", prepared for the San Joaquin Valleywide Air Pollution Study Agency, prepared by the Department of Meteorology, Pennsylvania State University, University Park, PA.

Seaman, N. L., and S. A. Michelson, 1998. “Mesoscale Meteorological Structure of a High-Ozone Episode During the 1995 NARSTO-Northeast Study”, Journal of Applied Meteorology, (submitted).

Seaman, N. L., D. R. Stauffer, and T. W. Tesche, 1992. "The SARMAP Meteorological Model: A Four-Dimensional Data Assimilation Technique Used to Simulate Mesobeta-Scale Meteorology During a High-Ozone Episode in California", International Specialty Conference on Tropospheric Ozone Nonattainment and Design Value Issues, U.S. EPA/AWMA, 27-30 October, Boston, MA.

Seaman, N. L., D. R. Stauffer, and L. M. Lario, 1995. "A MultiScale Four-Dimensional Data Assimilation System Applied to the San Joaquin Valley During SARMAP. Part I: Modeling Design and Basic Performance Characteristics", J. Appl. Meteo., Vol. 34, pp. 1739-1761.

Seaman, N. L. D. R. Stauffer, and D. E. McNally, 1996a. "Application of the MM5-FDDA Meteorological Model to the Southern California SCAQS-1997 Domain: Preliminary Test Using the SCAQS August 1987 Case", Ninth Joint Conference on Applications of Air Pollution Meteorology, American Meteorological Society, 28 January-2 February, Atlanta, GA.

Seaman, N. L., et al., 1996. "Application of the MM5-FDDA Meteorological Model to the Southern California SCAQS-1997 Domain: Preliminary Test Using the SCAQS August 1987 Case", Ninth Joint Conference on Applications of Air Pollution Meteorology, American Meteorological Society, 28 January-2 February, Atlanta, GA.

Seaman, N. L., et al., 1997. "The Use of the San Joaquin Valley Meteorological Model in Preparation of a Field Program in the South Coast Air Basin and Surrounding Regions of Southern California: Volume II -- Numerical Modeling Studies for the Development and Application of a Guidance Technique to Support of the Design of the 1997 Southern California Ozone Study Field Program", prepared for Alpine Geophysics, LLC and the California Air Resources Board, prepared by the Department of Meteorology, Pennsylvania State University, University Park, PA.

Seigneur, C. et al., 2000. "Guidance for the Performance Evaluation of Three-Dimensional Air Quality Modeling systems for Particulate Matter and Visibility", ***Journal of the Air and Waste Management Association,*** Vol. 50. pp. 588-599.

Seigneur, C. et al., 2002. "Development of New Science In Models-3/CMAQ", Models-3/CMAS Workshop, Research Triangle Park, NC, 21 October.

Shafran, P. C., and N. L. Seaman, 1998. "Comparison of Numerical Predictions of Boundary-Layer Structure Over the Midwest During the Lake Michigan Ozone Study (LMOS)", 10th Joint Conference on the Applications of Air Pollution Meteorology with the Air and Waste Management Association, Phoenix, AZ. 11-16 January.

Smagorinsky, J., 1963. "General Circulation Experiments with the Primitive Equations: I. The Basic Experiment", Mon. Wea. Rev, Vol. 91, pp. 99-164.

Smolarkiewicz, P. K. 1983. "A Simple Positive Definite Advection Scheme with Small Implicit Diffusion", Mon. Wea. Rev., Vol. 111, pp. 479-486.

Sonoma Technology, Inc., 1997a. "Peer Review of ENVIRON's Ozone Source Apportionment Technology and the CAMx Air Quality Model", Final Report STI996203-1732-FR. Prepared for the Division of Air Pollution Control, Ohio Environmental Protection Agency, Columbus, OH.

Sonoma Technology, Inc., 1997b. "Comparison of CAMx and UAM-V Model Performance for Two Ozone Episodes in the Eastern United States", Final Report STI996203-1733-FR. Prepared for the Division of Air Pollution Control, Ohio Environmental Protection Agency, Columbus, OH.

Stauffer, D.R. and N.L. Seaman, 1990. "Use of Four-Dimensional Data Assimilation in a Limited-Area Mesoscale Model. Part I: Experiments with Synoptic Data". Mon. Wea. Rev., 118, 1250-1277.

Stauffer, D.R., N.L. Seaman and F.S. Binkowski, 1991. "Use of Four-Dimensional Data Assimilation in a Limited-Area Mesoscale Model. Part II: Effects of Data Assimilation Within the Planetary Boundary

Layer." Mon. Wea. Rev., 119, 734-754.

Steyn, D. G., and I. G. McKendry, 1988, "Quantitative and Qualitative Evaluation of a Three-Dimensional Mesoscale Numerical Model Simulation of a Sea Breeze in Complex Terrain", Monthly Weather Review, Vol. 116, pp. 1914-1926.

Tang, Y. 2002, "A Case Study of Nesting Simulation for the Southern Oxidants Study 1999 at Nashville", ***Atmospheric Environment***, Vol. 36, pp. 1691-1705.

Tanrikulu, S. et al., 1999. "Numerical Simulation of Meteorology for the July 26-30, 1990 Ozone Episode in the San Joaquin Valley", 11th Joint Conference on the Applications of Air Pollution Meteorology with the AWMA, American Meteorological Society, Long Beach, CA, 9-14 January.

Tesche, T. W., 1976. "Theoretical, Numerical, and Physical Techniques for Characterizing the Fate of Emissions from Power Plant Plumes", prepared for the Electric Power Research Institute, prepared by Systems Applications, Inc., San Rafael, CA.

Tesche, T. W., 1985. "Photochemical Dispersion Modeling: A Review of Model Concepts and Recent Applications Studies," Environment International, Vol. 9, pp. 465-489.

Tesche, T. W., 1991b. "Development and Application of the SARMAP Emissions Modeling System", prepared for the California Air Resources Board and Pacific Gas and Electric Company, prepared by Alpine Geophysics, Placerville, CA.

Tesche, T. W., 1991c. "Evaluation of Regional Atmospheric Models", State-of-Science Synthesis Paper, Comprehensive Modeling System Workshop, Sponsored by EPRI, U.S. DOE, NOAA, U.S. EPA, Environment Canada, API, and the Ontario Ministry of Environment, 7-8 November, Atlanta, GA.

Tesche, T. W., 1991d. "Evaluation Procedures for Using Numerical Meteorological Models as Input to Photochemical Models". 7th Joint Conference on Applications of Air Pollution Meteorology; American Meteorological Society, 14-18 January, New Orleans, LA.

Tesche, T. W., and J. Ogren, 1976. "Determination of the Feasibility of Ozone Formation in Power Plant Plumes", prepared for the Electric Power Research Institute, prepared by Systems Applications, Inc., and Meteorology Research, Inc., San Rafael, CA.

Tesche, T. W., and D. E. McNally, 1993a. "Operational Evaluation of the CAL-RAMS Meteorological Model for LMOS Episode 1: 26-28 June 1991", prepared for the Lake Michigan Air Directors Consortium, prepared by Alpine Geophysics, LLC, Crested Butte, CO.

Tesche, T. W., and D. E. McNally, 1993b. "Operational Evaluation of the CAL-RAMS Meteorological Model for LMOS Episode 2: 17-19 July 1991", prepared for the Lake Michigan Air Directors Consortium, prepared by Alpine Geophysics, LLC, Crested Butte, CO.

Tesche, T. W., and D. E. McNally, 1993c. "Operational Evaluation of the CAL-RAMS Meteorological Model for LMOS Episode 3: 25-26 August 1991", prepared for the Lake Michigan Air Directors Consortium, prepared by Alpine Geophysics, LLC, Crested Butte, CO.

Tesche, T. W., and D. E. McNally, 1993d. "Operational Evaluation of the CAL-RAMS Meteorological Model for LMOS Episode 4: 20-21 June 1991", prepared for the Lake Michigan Air Directors Consortium, prepared by Alpine Geophysics, LLC, Crested Butte, CO.

Tesche, T. W., and D. E. McNally, 1993e. "Operational Evaluation of the SARMAP Meteorological Model (MM5) for Episode 1: 3-6 August 1990", prepared for the Valley Air Pollution Study Agency, prepared by Alpine Geophysics, Crested Butte, CO.

Tesche, T. W., and D. E. McNally, 1993f "Operational Evaluation of the SARMAP Meteorological Model (MM5) for Episodes 2: 27-29 July 1990", prepared for the Valley Air Pollution Study Agency, prepared by Alpine Geophysics, Crested Butte, CO.

Tesche, T. W. and D. E. McNally, 1996a. “Superregional Ozone Modeling and Analysis Study C Phase I: Work Element 3: Assessment of the OTAG Data Sets -- Task 2 Technical Memorandum: Review of the OTAG Meteorological Inputs and Outputs”, prepared for the Midwest Ozone Group, prepared by Alpine Geophysics, LLC, Ft. Wright, KY.

Tesche, T. W. and D. E. McNally, 1996b. “Superregional Ozone Modeling and Analysis Study C Phase II: Work Element 5 Technical Report: Comparative Evaluation of the MM5 and RAMS Models for the July 1991 OTAG Episode”, prepared for the Midwest Ozone Group, prepared by Alpine Geophysics, LLC, Ft. Wright, KY.

Tesche, T. W. and D. E. McNally, 1996c. “Superregional Ozone Modeling and Analysis Study C Phase I: Work Element 3: Assessment of the OTAG Data Sets -- Task 3 Technical Memorandum: Review of the OTAG UAM-V Inputs and Results”, prepared for the Midwest Ozone Group, prepared by Alpine Geophysics, LLC, Ft. Wright, KY.

Tesche, T. W. and D. E. McNally, 1996d. “Superregional Ozone Modeling and Analysis Study C Phase II: Status Report on SAQM/UAM-V Comparative Model Evaluation”, prepared for the Midwest Ozone Group, prepared by Alpine Geophysics, LLC, Ft. Wright, KY.

Tesche, T. W. and D. E. McNally, 1996f. “Evaluation of the MM5 Model for Three 1995 Regional Ozone Episodes over the Northeast United States”, prepared for the Southwestern Pennsylvania Clean Air Stakeholders Group, prepared by Alpine Geophysics, LLC, Ft. Wright, KY.

Tesche, T. W., and D. E. McNally, 1996i “Assessment of UAM-IV Boundary Conditions and Mass Fluxes for Four LMOS Episodes”, prepared for the Indiana Electric Association, prepared by Alpine Geophysics, LLC, Ft. Wright, KY.

Tesche, T. W., and D. E. McNally, 1997a. “Superregional Ozone Modeling and Analysis Study C Final Report: Assessment of the Reliability of the OTAG Modeling System”, prepared for the Midwest Ozone Group, prepared by Alpine Geophysics, LLC, Ft. Wright, KY.

Tesche, T. W., and D. E. McNally, 1997b. “Methodology for Climatological Analysis of Multi-Year Aerometric Data to Support the Development of the Breton Conceptual Model and Field Program Design”, prepared for Walk Heydel Environmental, prepared by Alpine Geophysics, LLC, Ft. Wright, KY.

Tesche, T. W., and D. E. McNally, 1997c. "The Use of the San Joaquin Valley Meteorological Model in Preparation of a Field Program in the South Coast Air Basin and Surrounding Regions of Southern California: Volume I: Final MM5 Evaluation for the 3-6 August 1990 SARMAP Episode", prepared for the California Air Resources Board, prepared by Alpine Geophysics, LLC, Ft. Wright, KY.

Tesche, T. W. and D. E. McNally, 1998a. "Examination of CAMx, SAQM, and UAM-V Performance and Response to Emissions Changes Over the Eastern U.S. for Various OTAG, LMOS, and NARSTO Episodes", 91st Annual Meeting of the Air and Waste Management Association, San Diego, CA, 14-19 June 1998.

Tesche, T. W., and D. E. McNally, 1998b. "Modeled Effects of Indiana Point Source NOx Emissions Reductions on Local and Regional 1-Hr and 8-Hr Ground Level Ozone Concentrations in 1995 and 2007 Using Two OTAG Oxidant Episodes", prepared for the Indiana Electric Utility Air Workgroup, prepared by Alpine Geophysics, LLC, Ft. Wright, KY.

Tesche, T. W., and D. E. McNally, 1998c. "Evaluation of the MM5 Model for Two 1993 Regional Ozone Episodes Over the Gulf Coast", prepared for the Offshore Operators Committee, prepared by Alpine Geophysics, LLC, Ft. Wright, KY.

Tesche, T. W., and D. E. McNally, 1998d. "Recommendations for Air Quality Dispersion Models and Related Aerometric Data Sets in Support of the Breton Aerometric Monitoring Program (BAMP)", prepared for the Offshore Operators Committee, prepared by Alpine Geophysics, LLC, Ft. Wright, KY.

Tesche, T. W., and D. E. McNally, 1998e. "Cincinnati-Hamilton Ozone Attainment Demonstration Study: Volume 5: Evaluation of the MM5 Model for the 18-22 June 1994 Episode", prepared for the Ohio EPA, prepared by Alpine Geophysics, LLC, Ft. Wright, KY.

Tesche, T. W. and R. E. Morris, 1998f. "Cincinnati-Hamilton Ozone Attainment Demonstration Study: Volume IV – Review and Synthesis of Regional Modeling Studies Related to the EPA SIP Call", prepared for the Ohio Environmental Protection Agency, prepared by Alpine Geophysics, LLC, Ft. Wright, KY.

Tesche, T. W., and D. E. McNally, 1998g. "Modeled Effects of NOx Controls on Electric Utility Sources in Western Missouri on 1-Hr Ozone Concentrations in Nonattainment Areas in the Eastern U.S.", prepared for Spencer Fane Britt & Browne, LLP, prepared by Alpine Geophysics, LLC, Ft. Wright, KY.

Tesche, T. W., and D. E. McNally, 1999a. "Comparison of Photochemical Model Performance in Several Studies in the Lower Lake Michigan Region", prepared for the Coordinating Research Council, prepared by Alpine Geophysics, LLC, Ft. Wright, KY.

Tesche, T. W., and D. E. McNally, 1999b. "Comparison of Photochemical Model Performance in Several Studies in the Northeast U.S.", prepared for the Coordinating Research Council, prepared by Alpine Geophysics, LLC, Ft. Wright, KY.

Tesche, T. W. and D. E. McNally, 1999c. "Comparative Evaluation of the MM5 and RAMS3c Prognostic Meteorological Models Over the Midwestern U.S. for Two 1999 LMOS Intensive Measurement Episodes", prepared for the Coordinating Research Council, Project A-25 draft Final Report, prepared by Alpine Geophysics, LLC, Ft. Wright, KY.

Tesche, T. W., and D. E. McNally, 1999d. “Impact of Incorrect Grid M Point Source Emissions Data on Ozone Model Performance and Year 2007 CAA and SIP Call Projections”, prepared for the Indiana Electric Association, prepared by Alpine Geophysics, LLC, Ft. Wright, KY.

Tesche, T. W. and D. E. McNally, 2003a. “Application and Evaluation of the MM5/CAMx Regional Modeling System for Two 8-hr Ozone Episodes Over the Denver-Northern Front Range and San Juan/Four Corners Region”, prepared by Alpine Geophysics, LLC, Ft. Wright, KY.

Tesche, T. W. and D. E. McNally, 2003b. “Application and Evaluation of the MM5 Prognostic Meteorological Model at 1.33km Scale Over the San Juan Basin/Four Corners Region for Two 8-hr Ozone Episodes”, prepared by Alpine Geophysics, LLC, Ft. Wright, KY.

Tesche, T. W., D. E. McNally, and C. F. Loomis, 1998f. “Cincinnati-Hamilton Ozone Attainment Demonstration Study: Volume 3: Interim Attainment Demonstration”, prepared for the Ohio EPA, prepared by Alpine Geophysics, LLC, Ft. Wright, KY.

Tesche, T. W., D. E. McNally, and C. F. Loomis, 1998g. “Cincinnati-Hamilton Ozone Attainment Demonstration Study: Volume 8: Final Report”, prepared for the Ohio EPA, prepared by Alpine Geophysics, LLC, Ft. Wright, KY.

Tesche, T. W., D. E. McNally, and C. F. Loomis, 1998h. “Photochemical Modeling Analysis of the Effects of Electric Utility NOx Emissions Trading on Eastern U.S. Ozone Concentrations”, prepared for First Energy Corporation, prepared by Alpine Geophysics, LLC, Ft. Wright, KY.

Tesche, T. W., D. E. McNally, and R. E. Morris, 1998i, “Review of the EPA Ozone Transport SIP Call and Recent Post-OTAG Modeling and Analysis Studies”, prepared for the Ohio EPA and the States of Kentucky and West Virginia, prepared by Alpine Geophysics, LLC and ENVIRON International Corporation.

Tesche, T. W., D.E. McNally, and C. F. Loomis, 1999a. “Ozone Modeling Protocol for the Pennsylvania Stakeholders Study”, prepared for the Pennsylvania Department of Environmental Protection, prepared by Alpine Geophysics, LLC, Ft. Wright, KY.

Tesche, T. W., D.E. McNally, and C. F. Loomis, 1999b. “Estimation of Emissions Reduction Targets for the Pennsylvania Stakeholders Study”, Final Report for the South-Central and Lehigh Valley Stakeholders Groups, and Pennsylvania Department of Environmental Protection, prepared by Alpine Geophysics, LLC, Ft. Wright, KY.

Tesche, T. W., et al., 1991. "Improved Treatment of Procedures for Evaluating Photochemical Models", prepared for the California Air Resources Board, prepared by Alpine Geophysics, Crested Butte, CO. Contract No. A832-103.

Tesche, T. W., et al., 1992. "Scientific Assessment of the Urban Airshed Model (UAM-IV)", report by Alpine Geophysics, Envair, and Sonoma Technology, Inc, to the American Petroleum Institute. Report No. AG-90/TS16, Crested Butte, CO.

Tesche, T. W. et al., 1997d. “Photochemical Modeling Analysis of the Pittsburgh-Beaver Valley Ozone Nonattainment Area: Volume VII -- Final Report”, prepared for the Southwestern Pennsylvania

Stakeholders and the Pennsylvania Department of Environmental Protection, prepared by Alpine Geophysics, LLC, Ft. Wright, KY.

Tesche, T. W., et al., 1997e. "The Use of the San Joaquin Valley Meteorological Model in Preparation of a Field Program in the South Coast Air Basin and Surrounding Regions of Southern California", prepared for the California Air Resources Board, prepared by Alpine Geophysics, LLC, and Penn State University.

Tesche, T. W., et al., 1998a. "Assessment of the Reliability of the OTAG Modeling System", 10th Joint Conference on Applications of Air Pollution Meteorology, 11-16 January, Phoenix, AZ.

Tesche, T. W., et al., 1998b. "Final Evaluation of the MM5 Model for the August 3-6, 1990 SARMAP Episode over Central California", 10th Joint Conference on Applications of Air Pollution Meteorology, 11-16 January, Phoenix, AZ.

Tesche, T. W., et al., 1998c. "Tri-State Regional Ozone Modeling Study: Results of the Sub-Regional Modeling of the EPA Section 110 SIP call", prepared for the Greater Cincinnati Chamber of Commerce, prepared by Alpine Geophysics, LLC, Ft. Wright, KY.

Tesche, T. W., et al., 1998d. "Application of EPA's Flexible Attainment Demonstration Guidance to the Pittsburgh-Beaver Valley Ozone Nonattainment Area", 10th Conference on Applications of Air Pollution Meteorology with the Air and Waste Management Association, American Meteorological Society, 11-16 January, Phoenix, AZ.

Tesche, T. W., et al., 1998e, "Photochemical Modeling Analysis of the Effects of VOC and NOx Emissions Reductions on 1-hr and 8-hr Ozone Concentrations in Kentucky", prepared for Louisville Gas and Electric Co., prepared by Alpine Geophysics, LLC, Ft. Wright, KY.

Tesche, T. W., et al., 1998f, "Photochemical Modeling Analysis of the Effects of VOC and NOx Emissions Reductions in the Kansas City Nonattainment Area on 1-hr and 8-hr Ozone Concentrations", prepared for Kansas City Gas and Electric Co., prepared by Alpine Geophysics, LLC, Ft. Wright, KY.

Tesche, T. W., et al., 1998g, "Photochemical Modeling Analysis of the Subregional Effects of the EPA Section 110 SIP Call Within and Downwind of the State of Virginia", prepared for AlliedSignal, Inc., prepared by Alpine Geophysics, LLC, Ft. Wright, KY.

Tesche, T. W., et al., 1998h, "Analysis of the Effects of VOC and NOx Emissions Reductions in the Eastern United States on Peak 1-hr and 8-hr Ozone Concentrations", prepared for the Midwest Ozone Group, prepared by Alpine Geophysics, LLC, Ft. Wright, KY.

Tesche, T. W., et al., 1999a. "Impact of Incorrect Grid M Point Source Emissions Data on Ozone Model Performance and Year 2007 CAA and SIP Call Projections", prepared for the Indiana Electric Association, prepared by Alpine Geophysics, LLC, Ft. Wright, KY.

Tesche, T. W., et al., 1999b. "Evaluation of the CAMx and MAQSIP Models Over the Lower Lake Michigan Region with Inputs from the RAMS3c and MM5 Models: Volume I – Final Report", prepared for the Coordinating Research Council, prepared by Alpine Geophysics, LLC, ENVIRON International, and Sonoma Technology, Inc.

Tesche, T. W., et al., 2000. "Ozone Modeling Protocol for the Peninsular Florida Ozone Study (Version 1.0)", prepared for the Florida Department of Environmental Protection, prepared by Alpine Geophysics, LLC, Ft. Wright, KY.

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.

Tesche, T. W., D. E. McNally, C. F. Loomis, R. E. Morris, and G. E. Mansell, 2003a. "Air Quality Modeling Analysis for the Denver Early Action Ozone Compact: Modeling Protocol, Episode Selection, and Domain Definition", prepared for the Denver Regional Air Quality Council, prepared by Alpine Geophysics, LLC and ENVIRON International Corporation, Ft. Wright, KY.

Tesche, T. W., D. E. McNally, and R. E. Morris, 2003b. "Air Quality Modeling Analysis for the San Juan County Early Action Ozone Compact: Episode Selection and Conceptual Model", prepared for the New Mexico Environment Department, Air Quality Bureau, prepared by Alpine Geophysics, LLC and ENVIRON International Corporation, Ft. Wright, KY.

Tesche, T. W., et al., 2003c. "Peninsular Florida Ozone Study Final Report", prepared for the Florida Department of Environmental Protection, prepared by Alpine Geophysics, LLC, Ft. Wright, KY.

Tesche, T. W. et al., 2003d. "Assessment of Potential 8-hr Ozone and PM10/PM2.5 Impacts from the We Energies 'Power the Future' Proposal", prepared for We Energies, Prepared by Alpine Geophyiscs, LLC and ENVIRON International Corp, Ft. Wright, KY.

TNRCC, 1998. "Comparative Evaluation of CAMx and UAM for the Houston/Beaumont COAST Domain", Letter from James W. Thomas, P.E., Texas Natural Resources Conservation Commission to Thomas Diggs, EPA Region VI, dated 13 April, 1998.

Tonnesen, G. et al., 2001. "Modeling Protocol and Quality Assurance Plan for the WRAP Visibility Modeling Project", prepared for the Western Regional Air Partnership Modeling Forum, prepared by University of California, Riverside and ENVIRON International Corporation.

Tremback, C. and D. E. McNally, 2003. "Alternative Meteorological Datasets for North America", prepared for the U. S. Environmental Protection Agency, Office of Air Quality Planning and Standards, prepared by ATMET, LLC, Boulder, CO and Alpine Geophysics, LLC, Arvada, CO.

Waple, A. M., and J. H. Lawrimore, 2003. "State of the Climate in 2002", Bulletin of the American Meteorological Society, Volume 84, No. 6, pp. S1-S68.

Wesley, M. L., 1989. "Parameterization of Surface Resistances to Gaseous Dry Deposition in Regional-Scale Numerical Models", Atmos. Environ., Vol. 23, pp. 1293-1304.

Wilkinson, J. G., 1997a. "Uncertainty Assessment of Biogenic Emissions Estimates and its Impact on Ozone Attainment Control Strategy Selection," Ph.D. Qualifier, Dept. of Engineering and Public Policy, Carnegie Mellon University, Pittsburgh, PA, 1997.

Wilkinson, J.G., 1997b. "Uncertainty Assessment of Biogenic Emissions Estimates and Its Impact on Ozone Attainment Control Strategy Selection," Air Pollution in the Ural Mountains: Environmental, Health, and Policy Aspects (2. Environment Volume 40), edited by I. Linkov and R. Wilson, Kluwer Academic Publishers, Boston.

Wilkinson, J. G., A. G. Russell, 1999. "The Biogenic Model For Emissions (BIOME): A Model To Estimate Biogenic VOC And NO," submitted to *Environmental Science & Technology*

Wilkinson, J.G., T W. Pierce, and R. A. Emigh, 1995. "An Intercomparison of Biogenic Emissions Estimates from BEIS2 and BIOME: Reconciling the Differences", 1995 AWMA Emissions Inventory Conference, Research Triangle Park, N.C.

Wilkinson, J. G., A. G. Russell, D. McNider, 1998a "Emissions Modeling Protocol: Meteorological, Emissions and Air Quality Modeling for an Integrated Assessment Framework in Support of the Southern Appalachians Mountain Initiative," Prepared For The Southern Appalachian Mountains Initiative (SAMI) Technical Oversight Committee, Ashford, NC.

Wilkinson, J. G., M. T. Odman, A. G. Russell, 1998b "Air Quality Modeling Protocol: Meteorological, Emissions and Air Quality Modeling for an Integrated Assessment Framework in Support of the Southern Appalachians Mountain Initiative," Prepared For The Southern Appalachian Mountains Initiative (SAMI) Technical Oversight Committee, Ashford, NC.

Wilkinson, J. G., et al., 1994. Technical Formulation Document: SARMAP/LMOS Emissions Modeling System (EMS-95). AG-90/TS26 & AG-90/TS27. Prepared for the Lake Michigan Air Directors Consortium, Des Plaines, IL & The Valley Air Pollution Study Agency, Technical Support Division, Sacramento, CA. Prepared by Alpine Geophysics, Pittsburgh, PA.

Wilkinson, J. G., et al., 1996. "Application of the Emissions Modeling System EMS-95 to the Southern California SCAQS-97 Domain", 9th Joint Conference on Applications of Air Pollution Meteorology, Atlanta, GA, 28 January-2 February.

Wilmont, C. J., 1981. "On the Validation of Models", Phys. Geog., Vol. 2, No. 2, pp. 168-194.

Yarwood, G. and C. S. Burton, 1993. "An Update to the Radical-Radical Termination Reactions in the CBM-IV", prepared by Systems Applications, Int., San Rafael, CA.

Yarwood, G. et al., 1996a. "Development of a Methodology for Source Apportionment of Ozone Concentration Estimates from a Photochemical Grid Model", 89th Annual Meeting of the Air and Waste Management Association, Nashville, TN. 23-28 June.

Yarwood, G. et al., 1996b. "User's Guide to the Ozone Tool: Ozone Source Apportionment Technology for UAM-IV", prepared by ENVIRON International Corporation, Novato, CA.

Yocke, M. A., 1996. "Future-Year Boundary Conditions for Urban Airshed Modeling for the State of Texas", Final Report. Prepared for the Texas Natural Resource Conservation Commission, prepared by ENVIRON International, Novato, CA.

Zhang, D. L., and R. A. Anthes, 1982. "A High-Resolution Model of the Planetary Boundary Layer

Sensitivity Tests and Comparisons with SESAME-79 Data", Journal of Applied Meteorology, Vol. 21. Pp. 1594-1609

Zhang, K. and S T. Rao, 1999. "Meteorological Modeling with MM5/FDDA for Use in Ozone Studies: Part I: Mesoscale FDDA Analyses on Meteorological Characteristics of High Ozone Episodes in 1995". Prepared by the New York State Department of Environmental Conservation, Albany, NY.

APPENDIX A

MM5 MODEL EVALUATION PROCEDURES

A.1 OVERVIEW

Before an air quality modeling system is applied to emission control strategy investigations, it must be tested in accordance with EPA's model evaluation guidelines. This provides some assurance to decision-makers that the model is producing the right answer for the right reasons. EPA's recommended model evaluation process (EPA, 1991; 1999; 2000) includes the calculation and analysis of several routine statistical measures and the plotting of specific graphical displays to characterize the basic performance attributes of the model. Among the statistics examined are: different measures for characterizing the model's accuracy in estimating the maximum one-hour average concentration; mean normalized bias to indicate the degree to which calculated one-hour concentrations are over- or underestimated; the variance, describing the dispersion of the residual distribution about the mean; and the mean normalized gross error, which quantifies the average absolute signed deviation of the concentration residuals. Evaluation of the MM5 model for the Denver EAC study was performed using Alpine Geophysics' MAPS software, described briefly below.

The Model Performance Evaluation, Analysis, and Plotting Software (MAPS) system package was developed for urban- and regional-scale meteorological, emissions, and photochemical model evaluations. The MAPS system embodies a variety of the statistical and graphical model testing methods for photochemical, fine particulate aerosol, and meteorological models recommended by various agencies including the California Air Resources Board (ARB) and the EPA (see, for example, ARB, 1992; EPA, 1991, 1999). MAPS also contains a variety of statistical and graphical tools for analyzing emissions model estimates. The performance measures calculated with MAPS are consistent with the definitions contained in Appendix C of the "Guideline for Regulatory Application of the Urban Airshed Model" and with EPA's recent 8-hour ozone and PM/regional haze modeling guidelines (EPA, 1999; 2000).

MAPS consists of a set of special-purpose FORTRAN codes, the National Center for Supercomputer Applications (NCSA) Hierarchial Data Format (HDF) data management libraries (ported to SUN and IBM RS/6000 platforms) and National Center for Atmospheric Research (NCAR) Graphics, Version 3.01. The formulation of the general package of statistical measures and graphical procedures available within MAPS are presented in this appendix. Not all of these techniques are used in every applications; some are tailored to the specific need. In some of the definitions below, the variable Φ represents a model-estimated or derived quantity, e.g., ozone, NOx, nitrate, sulfate, or H_2O_2 concentration, wind speed, wind direction, PBL height, ambient temperature. The subscripts e and o correspond to model-estimated and observed quantities, respectively. The subscript i refers to the ith hour of the day.

A.2 MEAN AND GLOBAL STATISTICS

Several statistical measures are calculated to provide an overall summary of photochemical and meteorological model estimates and observations and to support calculation of other statistical measures.

Mean Estimation (M_e). The mean model estimate is given by:

$$M_e = \frac{1}{N} \sum_{i=1}^{N} \Phi_{ei}$$

where N is the product of the number of simulation hours and the number of ground-level monitoring locations providing hourly-averaged observational data. Φ_{ei} represents the model-estimate at hour i. As noted above, this variable might be a gas-phase or particulate aerosol concentration, a meteorological state variable, or some other quantity.

Mean Observation (M_o). The mean observation is given by:

$$M_o = \frac{1}{N} \sum_{i=1}^{N} \Phi_{oi}$$

Here, Φ_{oi} represents the observations at hour i.

Average Wind Direction. Because wind direction has a crossover point between 0 degrees and 360 degrees, standard linear statistical methods cannot be used to calculate the mean or standard deviation. Evaluations by the EPA (Turner, 1986) suggest that the method proposed by Yamartino (1984) performs well in estimating the wind direction standard deviation. Specifically, this quantity is calculated by:

$$\sigma_\alpha = \arcsin(\beta)\ [1 + 0.1547\ \beta^3]$$

where:

$$\beta = \left[1.0 - [(\overline{\sin \alpha})^2 + (\overline{\cos \alpha})^2] \right]^{1/2}$$

Here, alpha is the measured hourly or instantaneous wind direction value.

Standard Deviation of Estimation (Sde). The standard deviation of the model estimates is given by:

$$SD_e = \left[\frac{1}{N} \sum_{i=1}^{N} |\Phi_{ei} - M_e|^2 \right]^{1/2}$$

Standard Deviation of Observations (SDo). The standard deviation of the observations is given by:

$$SD_o = \left[\frac{1}{N} \sum_{i=1}^{N} |\Phi_{oi} - M_o|^2 \right]^{1/2}$$

Least Square Slope and Intercept Regression Statistics. A linear least-squares regression is performed to calculate the intercept (a) and slope (b) parameters in the following equation:

$$\hat{\Phi}_{ei} = a + b \ \Phi_{oi}$$

This regression is performed for each set of hourly (or instantaneous) data to facilitate calculation of several error and skill statistics.

Maximum Ratio (R_{max}). The maximum ratio is defined as the quotient of the maximum one-hour averaged model estimated concentration and the maximum hourly-averaged measurement, i.e.,

$$R_{\max} = \frac{c_e(x, t)}{c_o(\hat{x}, \hat{t})}$$

where c_e is the estimated one-hour averaged pollutant concentration, c_o is the observed hourly averaged concentration, $\hat{x}$ refers to the peak monitoring station location, $\hat{t}$ is the time of the peak observation. The caret, ˆ, denotes the time or location of the maximum observed concentration. There is no requirement that the maximum estimated and observed concentrations be paired in either time or space but for this measure we require that the maximum modeled concentration be taken from a monitoring station.

A.3 DIFFERENCE STATISTICS

Residual (d_i). For quantities that are continuous in space and time (i.e., wind speed, temperature, pressure, pbl height, species concentrations) difference (or residual) statistics are very useful. Difference statistics are based on the definition of a residual quantity. A concentration residual, for example, is defined as:

$$d_i = c_e(x_i, t) - c_o(x_i, t)$$

where d_i is the i-th residual based on the difference between model-estimated (c_e) and observed (c_o) concentration at location x and time i.

Standard Deviation of Residual Distribution (SDr). The standard deviation of the residual distribution is given by:

$$SD_r = \left(\frac{1}{N - 1} \sum_{i=1}^{N} (d_i - MBE)^2 \right)^{0.5}$$

where the concentration residual is defined as:

$$d_i = c_e(x_i, t) - c_o(x_i, t)$$

and MBE is the first moment, i.e., the mean bias error, defined shortly. This statistic describes the "dispersion" or spread of the residual distribution about the estimate of the mean. The standard deviation is calculated using all estimation-observation pairs above the cutoff level. The second moment of the residual distribution is the variance, the square of the standard deviation. Since the standard deviation has units of concentration, it is used here as the metric for dispersion. The standard deviation and variance measure the average "spread" of the residuals, independent of any systematic bias in the estimates. No direct information is provided concerning subregional errors or about large discrepancies occurring within portions of the diurnal cycle although in principle these, too, could be estimated.

Accuracy of Peak Model Estimates (A). Five related methods are used to evaluate the accuracy of the model's estimate of the maximum value of a spatially-distributed variable. This may be, for example, temperature, wind speed, pressure, or concentration. In the definitions below we use the peak one-hour average concentrations for discussion purposes; however, these measures may be applied to several of the meteorological variables as well.

Several accuracy measures are used because there are different, informative, and plausible ways of comparing the peak measurement on a given day with model estimates. These five accuracy measures provide complimentary tests of the model's performance. When applied to ozone simulations, they are particularly useful from a regulatory perspective since they deal with peak ozone (or precursor) concentration levels.

Paired Peak Estimation Accuracy. The paired peak estimation accuracy, A_{ts}, is given by:

$$A_{ts} = \frac{c_e(\hat{x},\hat{t}) - c_o(\hat{x},\hat{t})}{c_o(\hat{x},\hat{t})} 100\%$$

A_{ts} quantifies the discrepancy between the magnitude of the peak one-hour average concentration measurement at a monitoring station, $c_o(\hat{x},\hat{t})$, and the estimated concentration at the same location, $\hat{x}$, and at the same time, $\hat{t}$. Model estimates and observations are thus "paired in time and space." The paired peak estimation accuracy is a stringent model evaluation measure. It quantifies the model's ability to reproduce, at the same time and location, the highest observed concentration during each day of the episode. The model-estimated concentration used in all comparisons with observations is derived from bi-linear interpolation of the four ground level grid cells nearest the monitoring station.

A_{ts} is very sensitive to spatial and temporal misalignments between the estimated and observed concentration fields. These space and time offsets may arise from spatial displacements in the transport fields resulting from biases in wind speed and direction, problems with the "timing" of photochemical oxidation and removal processes, or subgrid-scale phenomena (e.g., ozone titration by local NO_x emission sources) that are not intended to be resolvable by grid-based photochemical models.

Temporally-Paired Peak Estimation Accuracy. The temporally-paired peak estimation accuracy, A_t,

$$A_t = \frac{c_e(x,\hat{t}) - c_o(\hat{x},\hat{t})}{c_o(\hat{x},\hat{t})} x\ 100\%$$

is given by:

A_t quantifies the discrepancy between the highest concentration measurement at a monitoring station and the highest model estimate at the same station or any other grid cell within a distance of, say, 4 - 5 grid cells. This measure examines the model's ability to reproduce the highest observed concentration in the same subregion at the correct hour.

Spatially-Paired Peak Estimation Accuracy. The spatially-paired peak estimation accuracy, A_s, is given by:

$$A_s = \frac{c_e(\hat{x},t) - c_o(\hat{x},\hat{t})}{c_o(\hat{x},\hat{t})} \; x\ 100\,\%$$

A_s quantifies the discrepancy between the magnitude of the peak one-hour average concentration measurement at a monitoring station and the highest estimated concentration at the same monitor, within 3 hours (before or after) the peak hour.

Unpaired Peak Estimation Accuracy. The unpaired peak estimation accuracy, A_u, is given by:

$$A_u = \frac{c_e(x,t) - c_o(\hat{x},\hat{t})}{c_o(\hat{x},\hat{t})} \; x\ 100\,\%$$

A_u quantifies the difference between the magnitude of the peak one-hour average measured concentration and the highest estimated value in the modeling domain, whether this occur at a monitoring station or not. The unpaired peak estimation accuracy tests the model's ability to reproduce the highest observed concentration anywhere in the region. This is the least stringent of the above four peak estimation measures introduced thus far. It is a weak comparison relative to the previous ones but is useful in coarse screening for model failures. This measure quickly identifies situations where the model produces maximum ozone concentrations in the air basin that significantly exceed the highest observed values within the network.

Average Station Peak Estimation Accuracy. The average station peak estimation accuracy, $\underline{A}$, is given by:

$$\overline{A} = \frac{1}{N} \sum_{i=1}^{N} |A_{si}|$$

where:

$$A_{si} = \frac{c_e(\hat{x}_i,t) - c_o(\hat{x},\hat{t})}{c_o(\hat{x},\hat{t})} \; x\ 100\,\%$$

Here, x_i is the ith monitoring station location. $\underline{A}$ is calculated by first determining the spatially-paired peak estimation accuracy, Asi, at each monitoring station. Thus, the average station peak estimation accuracy is simply the mean of the absolute value of the A_{si} scores, where the temporal offset between estimated and observed maxima at any monitoring station does not exceed three hours.

Mean Bias Error (MBE). The mean bias error is given by:

$$MBE = \frac{1}{N} \sum_{i=1}^{N} (c_e(x_i, t) - c_o(x_i, t))$$

where N equals the number of hourly estimate-observation pairs drawn from all valid monitoring station data on the simulation day of interest.

Mean Normalized Bias Error (MNBE). The mean normalized bias error, often just called the bias, is given by:

$$MNBE = \frac{1}{N}\sum_{i=1}^{N}\frac{(c_e(x_i,t) - c_o(x_i,t))}{c_o(x_i,t)} \text{ x } 100\ \%$$

Mathematically, the bias is derived from the average signed deviation of the concentration residuals and is calculated using all pairs of estimates and observations above the cutoff level. Cutoff levels of 40-60 ppb for ozone and 20 ppb for NO_2 are often used in modeling studies to reduce the influence that low measured or modeled concentrations (often occurring at night or on the upwind boundaries) have on the normalized bias statistics. In regions of exceptionally high ozone, e.g., the South Coast Air Basin or Houston-Galveston, cutoff levels as high as 100 ppb are commonly used. For this study, an ozone cutoff level of 60 ppb will be used, consistent with EPA (1991) guidance.

Mean Absolute Gross Error (MAGE). The mean gross error is calculated in two ways, similar to the bias. The mean absolute gross error is given by:

$$MAGE = \frac{1}{N}\sum_{i=1}^{N}|c_e(x_i, t) - c_o(x_i, t)|$$

Mean Absolute Normalized Gross Error (MANGE). The mean absolute normalized gross error is:

$$MANGE = \frac{1}{N}\sum_{i=1}^{N}\frac{|c_e(x_i, t) - c_o(x_i, t)|}{c_o(x_i, t)} \text{ x } 100\ \%$$

The gross error quantifies the mean absolute deviation of the concentration residuals. It indicates the average unsigned discrepancy between hourly estimates and observations and is calculated for all pairs above the cutoff level of 60 ppb. Gross error is a robust measure of overall model performance and provides a useful basis for comparison among model simulations across different air basins or ozone episodes. Unless calculated for specific locations or time intervals, gross error estimates provide no direct information about sub-regional errors or about large discrepancies occurring within portions of the diurnal cycle.

Root Mean Square Error (RMSE). The root mean square error is given by:

$$RMSE = \left[\frac{1}{N}\sum_{i=1}^{N}|\Phi_{ei} - \Phi_{oi}|^{2}\right]^{1/2}$$

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

Systematic Root Mean Square Error ($RMSE_s$). A measure of the model's linear (or systematic) bias may be estimated from the systematic root mean square error given by:

$$RMSE_s = \left[\frac{1}{N} \sum_{i=1}^{N} |\hat{\Phi}_{ei} - \Phi_{oi}|^{2} \right]^{1/2}$$

Unsystematic Root Mean Square Error ($RMSE_u$). A measure of the model's unsystematic bias is given by the unsystematic root mean square error, that is:

$$RMSE_u = \left[\frac{1}{N} \sum_{i=1}^{N} |\Phi_{ei} - \hat{\Phi}_{ei}|^{2} \right]^{1/2}$$

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 root mean square error, RMSE, explaining most of the variation in the observations. The systematic error, $RMSE_s$ should approach zero and the unsystematic error $RMSE_u$ 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 bias might be removed, thereby reducing the bias transferred to the photochemical calculation. On the other hand, if the RMSE consists largely of the unsystematic component ($RMSE_u$), this indicates further error reduction may require model refinement and/or data acquisition. It also provides error bars that may used with the inputs in subsequent sensitivity analyses.

A.4 SKILL MEASURES

Index of Agreement (I). Following Willmont (1981), the index of agreement is given by:

$$I = 1 - \left[\frac{N\ (RMSE)^{2}}{\sum_{i=1}^{N} \left(|\Phi_{ei} - M_o| + |\Phi_{oi} - M_o| \right)^{2}} \right]$$

This metric condenses all the differences between model estimates and observations into one statistical quantity. It is the ratio of the cumulative difference between the model estimates and the corresponding observations to the sum of two differences: between the estimates and observed mean and the observations and the observed mean. Viewed from another perspective, the index of agreement is a measure of how well the model estimates departure from the observed mean matches, case by case, the observations' departure from the observed mean. Thus, the correspondence between estimated 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.

RMS Skill Error ($Skill_e$). The root mean square error skill ratio is defined as:

$$Skill_E = \frac{RMSE_u}{SD_o}$$

Variance Skill Ratio ($Skill_{var}$). The variance ratio skill is given by:

$$Skill_{Var} = \frac{SD_e}{SD_o}$$

A.5 GRAPHICAL TOOLS

Many features of photochemical and meteorological model simulations are best analyzed through graphical means. In addition to revealing important qualitative relationships, graphical displays also supply quantitative information. The main graphical displays that may be used to analyze air quality model performance results are as follows:

> The relationships among the five accuracy measures;

> The temporal correlation between estimates and observations;

> The spatial distribution of estimated concentration fields;

> The correlation among hourly pairs of estimates, observations and residuals;

> The variation in bias and error estimates as functions of time and space; and

> The degree of mismatch between volume-averaged model estimates and point measurements.

> The distributional relationships between rank-ordered observations and rank-ordered model estimates.

Brief discussions of these plotting methods used in MAPS are as follows.

Accuracy Plots. Two accuracy plots are used. One depicts relationships between the peak five accuracy measures while the other plot summarizes the peak estimation accuracy at all monitoring stations, revealing the presence of subregional estimation bias if it occurs. The first plot is a histogram that displays the calculated values of A_{ts}, A_t, A_s, A_u, and $\underline{A}$. The second plot is also a histogram showing the peak observed and estimated concentrations (unpaired in time) at each monitoring station above the cutoff concentration of 60 ppb. Also contained on the plot is a shaded region corresponding to the normalized gross error.

Time Series Plots. Probably the most useful graphical procedure for depicting air quality model results is the time series plot. Developed for each monitoring station for which observed concentrations are available, this plot presents the hourly estimates and observations throughout the simulation period. The time series plot consists of the hourly averaged observations (boxes) and the hourly averaged estimates, the latter being fitted by a smooth continuous line. The model estimates are derived from bi-linear interpolation of the nearest four grid cells to the monitor. At each hour, the absolute value of the concentration residual will be calculated and plotted as a dashed line on the same plot.

With the time series plot one may determine the model's ability to reproduce the peak estimation, the presence or absence of significant bias and errors within the diurnal cycle, and whether the "timing" of the estimated concentration maximum agrees with the observation. By including the residual plot on the same graph, estimation biases are more apparent.

Spatial Time Series Plots. Conventional time series plots do not reveal situations where the model estimates concentrations comparable in magnitude to the observations a short distance away from the monitoring station. A second time series display, called a "spatial time series plot", are used for this purpose. These plots provide information about the degree to which model discrepancies result from the procedure for selecting the estimated values. There is no a priori reason to select the four-cell bi-linear average estimate over the estimate in the specific grid cell containing the monitor (i.e., the "cell value"), or perhaps the grid cell estimate within any of the four adjacent cells that is closest in magnitude to the observed value (i.e. the "best" estimate). Spatial time series plots are constructed for each monitoring station by plotting the hourly observations together with an envelope defined by the highest and lowest grid cell estimate within one cell of the monitoring station. MAPS can easily examine multiple grid cell distances as well.

The spatial time series plots provide diagnostic information about the "steepness" of the concentration gradients in the simulated fields. A small envelope indicates relatively flat concentration gradients. Conversely, steep gradients may produce a fairly large envelope. Ideally, the measurement points will fall within the envelope. Spatial time series plots are one method of revealing the correspondence or "commensurability" between volume-averaged model estimates and point measurements.

Ground Level Isopleths. Ground-level ozone isopleths are developed for each hour of the episode to display the spatial distribution of estimated concentration fields. The isopleth plots are developed by computer-contouring the hourly, gridded ozone estimates. The information content of these plots are enhanced by including the following:

- A base map identifying significant geophysical and political boundaries;
- Locations of air monitoring stations;
- The observed concentrations at each monitoring station by a bold numeral;
- The location of the peak estimate (signified by an asterisk); and
- The magnitude of the peak grid cell estimate.

Ground-level isopleths are also constructed based on the daily maximum concentration estimate in each grid cell. These "maximum" ozone isopleths supply direct information about the magnitude and location of pollutant concentrations and help to identify situations where sub-regional biases may be attributed to spatial misalignment of the estimated and observed concentration fields.

Scatterplots of Estimates and Observations. Scatterplots are a useful means of visually assessing the extent of bias and error in hourly ozone estimate-observation pairs. Hourly scatterplots are developed by plotting all hourly-averaged estimate-observation pairs for which the observed concentration exceeds the cutoff value. Similarly, daily maximum scatterplots are developed from the pairs of maximum hourly estimated and observed values at each monitoring station. The estimated maximum is the highest value simulated within three hours of the observed maximum. In these plots, the solid diagonal line with 1:1 slope will be used to identify the perfect correlation line and the dashed lines enclose the region wherein estimates and observations agree to within a factor of two. The lines of agreement can be made more stringent if desired.

The scatterplot is used to give a quick visual indication of the extent of over-or underestimation in the hourly estimates and whether there appear to be strong nonlinearities in model estimates and observations over the concentration range studied. Bias is indicated by the preponderance of data points falling above or below the perfect correlation line. The dispersion (spread) of points provides a visual indication of the general error pattern in the simulation. Scatterplots help identify outlier estimate-observation pairs, i.e., a seemingly discrepant estimate-observation pair that may result from erroneous data, a fundamental flaw in the model, or some other cause that requires investigation. These plots provide little diagnostic information about sub-regional performance problems, temporal or spatial misalignments, or other inadequacies in the simulation. In addition, scatterplots mask the temporal correlation between various estimate-observation pairs.

Scatterplot of Residuals and Observations. Residual scatterplots are developed to describe the distribution of hourly average model discrepancies (positive and negative) as a function of concentration level. This graphical display is constructed from the data elements that make up the bias and error calculations. Hourly concentration residuals for all monitoring stations are plotted as a function of observed concentration for all pairs above the cutoff value. A daily maximum residual plot is also constructed based on data pairs involving the maximum observed concentration at a monitor station and the maximum estimated value at the same station within three hours of the peak.

Residual scatterplots are used to characterize estimation discrepancy throughout the observed concentration range. The plot does not reveal the existence or causes of sub-regional or timing performance problems. Absence of bias is suggested by no systematic tendency for the data points to fall above or below the ordinate; however, as noted previously, important subregional biases may still exist in the presence of a zero overall bias estimate.

Bias Stratified by Concentration. Bias-concentration plots are derived from the residual distribution to depict the degree of systematic bias in hourly-averaged model estimates (paired in time and space) as a function of observed concentration level. This plot (and the companion error-concentration plot) aids in model diagnosis. The observed concentration range is divided into several equal-sized concentration bins and the normalized bias within each bin is calculated and plotted as a function of concentration level. A smooth line is then fitted through the bin-averaged values. The bias-concentration plot is used to reveal the existence of under- or over-estimation throughout the concentration range.

Gross Error Stratified by Concentration. Gross error-concentration plots is derived from the residual distribution to depict the error in model estimation (paired in time and space) as a function of observed concentration level. The observed concentration range is divided into several equal-sized concentration bins. Then, the average value of the normalized gross error within each bin is calculated and the bin averages are plotted as a function of the observed concentration level. MAPS will display the mean normalized gross error on the plot for easy reference.

The gross error-concentration plot is used to reveal the variation in model error at various intervals throughout the concentration range. The plot must be interpreted carefully, however, remembering that the concentration residual is normalized by the observed value.

Bias Stratified by Time. Bias-time plots are developed to help identify specific time periods within the photochemical simulation when systematic patterns of under- or overestimation occur. The bias-time plot is constructed in a manner similar to the bias-concentration plot, except that the simulation period is discretized into a number of time intervals, usually 1-2 hours in duration. Systematic bias in model estimates during specific periods within the diurnal cycle may have several causes: biases in vertical mixing or wind transport; "timing" problems with the chemistry; non-representative temporal distributions assumed in the emissions inventory, and so on. While the bias-time plots may not clearly pinpoint the causes of bias, they may be helpful in defining the time intervals when the bias is most apparent. This helps focus subsequent diagnostic investigations.

Gross Error Stratified by Time. Gross error-time plots are developed to help identify specific time periods when gross errors in the model estimates may be a problem. This plot is constructed in a similar manner as the error-concentration plot, except that the simulation period is discretized into a number of time intervals, usually 1-2 hours in duration. When interpreting the gross error-time and bias-time plots, one must remember that the concentration levels of all pollutants vary throughout the diurnal cycle.

Quantile-Quantile Plots. Quantile-quantile plots are cumulative frequency distributions that provide a graphical characterization of the distribution of observed and modeled values over their entire ranges. Quantitative information that can be obtained from these distributions include estimates of the mean, median, and standard deviation. The plots also provide a visual characterization of how the estimates and observations are spread out with respect to the central value. They also readily display unpaired bias.