| REPORT DOCUMENTATION PAGE | Form Approved OMB No. 0704-0188 |
|---|---|



| 1. AGENCY USE ONLY *(Leave blank)* | 2. REPORT DATE | 3. REPORT TYPE AND DATES COVERED |
|---|---|---|
| | 17.Jul.04 | DISSERTATION |

**4. TITLE AND SUBTITLE**
AN EVALUATION OF HIGH-RESOLUTION MODELING AND STATISTICAL FORECAST TECHNIQUES OVER COMPLEX TERRAIN

**5. FUNDING NUMBERS**

**6. AUTHOR(S)**
MAJ HART KENNETH A

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**
UNIVERSITY OF UTAH

**8. PERFORMING ORGANIZATION REPORT NUMBER**
CI04-432

**9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)**
THE DEPARTMENT OF THE AIR FORCE
AFIT/CIA, BLDG 125
2950 P STREET
WPAFB OH 45433

**10. SPONSORING/MONITORING AGENCY REPORT NUMBER**

**11. SUPPLEMENTARY NOTES**

**12a. DISTRIBUTION AVAILABILITY STATEMENT**
Unlimited distribution
In Accordance With AFI 35-205/AFIT Sup 1

**12b. DISTRIBUTION CODE**

**13. ABSTRACT** *(Maximum 200 words)*

**DISTRIBUTION STATEMENT A**
Approved for Public Release
Distribution Unlimited

20040730 027

**14. SUBJECT TERMS**

**15. NUMBER OF PAGES**
105

**16. PRICE CODE**

| 17. SECURITY CLASSIFICATION OF REPORT | 18. SECURITY CLASSIFICATION OF THIS PAGE | 19. SECURITY CLASSIFICATION OF ABSTRACT | 20. LIMITATION OF ABSTRACT |
|---|---|---|---|
| | | | |

## ABSTRACT

The accuracy of weather forecasts produced during the 2002 Olympic and Paralympic Games (23 Jan – 25 Mar 2002) by a multiply nested version of the PSU-NCAR Mesoscale Model (MM5) and associated model output statistics (MOS) system is evaluated using observations collected by the MesoWest cooperative network. Using traditional verification measures, the accuracy of MM5 wind and precipitation forecasts improved as model grid spacing decreased from 12 to 4 km. In contrast, temperature forecasts did not improve with decreasing model grid spacing due to the inability of the MM5 to properly simulate the evolution of persistent or nocturnal cold pools over lowland regions. Improved parameterization of the stable boundary layer processes may enable model skill to improve as grid spacing is decreased from 12 to 4 km. The need to improve model physics was also illustrated by the fact that MM5 MOS temperature, relative humidity, wind speed, and wind direction forecasts were considerably more accurate than those produced by either the 12- or 4-km MM5.

These results suggest that, contrary to studies in regions of broader topographic features, decreasing model grid spacing (< 10 km) does improve model skill over the finescale topography of the Intermountain West. These improvements may be valuable for a variety of environmental applications including fire weather forecasting, air quality prediction, and transport and dispersion modeling. Nevertheless, large biases frequently limit the direct application of numerical model output for weather prediction of surface

variables. Thus, it is recommended that traditional MOS or other statistical techniques based on high-density observations from the MesoWest cooperative networks be used to improve surface weather forecasts, including those generated by the National Weather Service Interactive Forecast Preparation System (IFPS).

## ABSTRACT

The accuracy of weather forecasts produced during the 2002 Olympic and Paralympic Games (23 Jan – 25 Mar 2002) by a multiply nested version of the PSU-NCAR Mesoscale Model (MM5) and associated model output statistics (MOS) system is evaluated using observations collected by the MesoWest cooperative network. Using traditional verification measures, the accuracy of MM5 wind and precipitation forecasts improved as model grid spacing decreased from 12 to 4 km. In contrast, temperature forecasts did not improve with decreasing model grid spacing due to the inability of the MM5 to properly simulate the evolution of persistent or nocturnal cold pools over lowland regions. Improved parameterization of the stable boundary layer processes may enable model skill to improve as grid spacing is decreased from 12 to 4 km. The need to improve model physics was also illustrated by the fact that MM5 MOS temperature, relative humidity, wind speed, and wind direction forecasts were considerably more accurate than those produced by either the 12- or 4-km MM5.

These results suggest that, contrary to studies in regions of broader topographic features, decreasing model grid spacing (< 10 km) does improve model skill over the finescale topography of the Intermountain West. These improvements may be valuable for a variety of environmental applications including fire weather forecasting, air quality prediction, and transport and dispersion modeling. Nevertheless, large biases frequently limit the direct application of numerical model output for weather prediction of surface

variables. Thus, it is recommended that traditional MOS or other statistical techniques based on high-density observations from the MesoWest cooperative networks be used to improve surface weather forecasts, including those generated by the National Weather Service Interactive Forecast Preparation System (IFPS).

# AN EVALUATION OF HIGH-RESOLUTION MODELING AND STATISTICAL FORECAST TECHNIQUES OVER COMPLEX TERRAIN

by

Kenneth Alan Hart

A dissertation submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Meteorology

The University of Utah

August 2004

## ABSTRACT

The accuracy of weather forecasts produced during the 2002 Olympic and Paralympic Games (23 Jan – 25 Mar 2002) by a multiply nested version of the PSU-NCAR Mesoscale Model (MM5) and associated model output statistics (MOS) system is evaluated using observations collected by the MesoWest cooperative network. Using traditional verification measures, the accuracy of MM5 wind and precipitation forecasts improved as model grid spacing decreased from 12 to 4 km. In contrast, temperature forecasts did not improve with decreasing model grid spacing due to the inability of the MM5 to properly simulate the evolution of persistent or nocturnal cold pools over lowland regions. Improved parameterization of the stable boundary layer processes may enable model skill to improve as grid spacing is decreased from 12 to 4 km. The need to improve model physics was also illustrated by the fact that MM5 MOS temperature, relative humidity, wind speed, and wind direction forecasts were considerably more accurate than those produced by either the 12- or 4-km MM5.

These results suggest that, contrary to studies in regions of broader topographic features, decreasing model grid spacing (< 10 km) does improve model skill over the finescale topography of the Intermountain West. These improvements may be valuable for a variety of environmental applications including fire weather forecasting, air quality prediction, and transport and dispersion modeling. Nevertheless, large biases frequently limit the direct application of numerical model output for weather prediction of surface

variables. Thus, it is recommended that traditional MOS or other statistical techniques based on high-density observations from the MesoWest cooperative networks be used to improve surface weather forecasts, including those generated by the National Weather Service Interactive Forecast Preparation System (IFPS).

TABLE OF CONTENTS

# ACKNOWLEDGMENTS

I would like to thank my advisor, Jim Steenburgh, for his dedication and mentorship during the last three years. I am grateful for the guidance and support provided by my Ph.D. committee members: Larry Dunn, John Horel, Jay Mace, and Jan Paegle. Brett McDonald also served as an operations liaison and provided valuable feedback throughout this study. I also appreciate the MM5 and MOS development contributions of Daryl Onton and Andy Siffert along with their collaboration on the papers that make up a portion of this dissertation.

Computer time for the Olympic forecasts was provided by the University of Utah Center for High Performance Computing. MesoWest data were collected and processed by John Horel, Mike Splitt, Judy Pechmann, Brian Olsen, and Bryan White. I would like to thank the governmental agencies, commercial firms, and educational institutions that provide data to MesoWest. I greatly appreciate the support of Mark Eubank, KSL-TV, who provided the outdoor venue forecast data. Transportation corridor forecast data were provided by the National Weather Service forecast office at Salt Lake City.

I gratefully acknowledge the support of the United States Air Force AFIT program during this study. Other aspects of the project were sponsored by a series of grants provided by the National Weather Service C-STAR program to the NOAA Cooperative Institute for Regional Prediction at the University of Utah. Besides those listed above, Mark Jackson, Will Cheng, Dave Schultz, and two anonymous reviewers provided

insightful discussion and comments that enhanced and improved the content of the dissertation. The copyright release for Chapter 2 content was provided by the American Meteorological Society. The views expressed in this dissertation are those of the author and do not reflect the official policy or position of the United States Air Force, Department of Defense, or the U. S. Government.

I would like to extend my appreciation to colleagues at the University of Utah for their friendship and the insightful discussions that helped motivate me during this graduate work. They include, but are not limited to, Lee Byerle, Will Cheng, Justin Cox, Dave Myrick, Jennifer Roman, Jay Shafer, and Bryan White.

Finally, I would like to thank my loving wife and daughters who bring so much joy to my life. They never wavered from their commitment and support and, once again, bloomed where the Air Force planted us.

# CHAPTER 1

# INTRODUCTION

Local topographic effects and spatial land-surface variability present unique challenges for weather forecasting in areas of complex terrain. The primary goal of this research is to improve the accuracy of weather forecasts over the fine-scale topography of the Intermountain West using statistical and numerical modeling techniques. Specifically, we seek to determine where, when and why decreasing the horizontal grid spacing of numerical models improves forecast accuracy and to evaluate the added benefits of using statistical techniques to correct for systematic numerical model deficiencies.

The study is based on a multiply nested version of the PSU–NCAR fifth-generation Mesoscale Model (MM5), which was used for weather prediction during the 2002 Olympic and Paralympic Games. The verification region includes most of northern Utah and all of the outdoor Olympic venues. This region features the Great Salt Lake and the steep and complex terrain of the Wasatch, Oquirrh, and Stansbury Mountains, and is known for a variety of microclimates. The verification is based on high-density surface observations collected by the MesoWest cooperative networks (Horel et al. 2002a).

The remainder of this dissertation is organized as follows: An evaluation of MM5-based MOS forecasts for 18 sites including outdoor Olympic venues and transportation corridor locations, is described in Chapter 2. This chapter derives from the published work of Hart et al. (2004). An expanded verification study of the 12- and 4-km MM5 domains,

using approximately 200 temperature and wind verification sites and nearly 100 precipitation observation locations, is found in Chapter 3. A brief summary in Chapter 4 concludes the dissertation.

# CHAPTER 2

# AN EVALUATION OF MESOSCALE-MODEL-BASED MODEL OUTPUT STATISTICS (MOS) DURING THE 2002 OLYMPIC AND PARALYMPIC WINTER GAMES

## Introduction

Weather-sensitive organizations and individuals frequently desire temporally detailed, point-specific forecasts. Such forecasts provide a unique challenge in areas of complex terrain where meteorological conditions exhibit dramatic spatial variability because of local topographic effects. Over the western United States, for example, dramatic temperature, wind speed, and wind direction differences are found over distances as small as a few tens of meters (e.g., Whiteman 1990, 2000; Doran et al. 2002; Lazarus et al. 2002), and large contrasts in precipitation are commonly observed across distances of a few kilometers (e.g., Rangno 1986; Schultz et al. 2002). As a result, forecasts for transportation corridors or areas (e.g., a ski resort) must describe a wide range of meteorological conditions.

The 2002 Olympic and Paralympic Winter Games, which were held in Salt Lake City, Utah, and the adjacent Wasatch Mountains, provided several challenges of the type described above. During the Games, Olympic Forecast Team meteorologists were responsible for providing nowcast, short-range, and medium-range forecasts for five outdoor venues and six transportation corridors (Horel et al. 2002b). End users included

venue managers, sports managers, transportation planners, athletes, and athletic support teams. At most venues, local topographic effects made it necessary to predict a wide range of meteorological conditions. For example, at Snowbasin ski area, considerable variability in surface weather was found along the men's downhill course, which began on a meridionally oriented ridge, traversed a zonally oriented ridge, and crossed slopes with eastern, northeastern, and southeastern exposure while falling 900 vertical meters. Wind, air temperature, and snow temperature exhibited considerable variability along the course depending on the prevailing synoptic, mesoscale, and local weather conditions.

The 12-km grid spacing of the National Centers for Environmental Prediction (NCEP) Eta Model, which was available to Olympic forecasters on a 40-km grid, and the 4-km grid spacing of the real-time fifth-generation Pennsylvania State University–National Center for Atmospheric Research Mesoscale Model (MM5), provided by the University of Utah (Horel et al. 2002b), could not accurately capture local effects of the type described above. In fact, the resolution necessary to explicitly resolve the influence of ridges, canyons, and other topographic features on local meteorology will not likely be obtainable for many years. Even if sufficient computer power were available to explicitly resolve finescale topographic details, considerable improvement is needed to existing models to accurately simulate local boundary layer structure and evolution (e.g., Doran et al. 2002). As a result, current models exhibit systematic biases that limit their application toward detailed point-specific forecasting (e.g., Mass et al. 2002; Mass 2003).

One approach that has been used for more than three decades to produce objective, point-specific forecast guidance from relatively coarse-resolution numerical model output is model output statistics (MOS; Glahn and Lowry 1972). Currently, MOS based on the

Eta Model and the Aviation (AVN) run of the Global Spectral Model provides forecasts in the United States for about 1300 and 1400 sites, respectively. Unfortunately, the majority of western U.S. MOS sites are located at valley airports and are of limited utility for weather prediction in adjacent mountainous regions.

Primary forecasting responsibilities during the Games were shared by the National Weather Service (NWS) and television station KSL, the local National Broadcasting Company affiliate (Horel et al. 2002b). Five forecasters from other NWS Weather Forecast Offices (WFOs) and one forecaster from the National Severe Storms Laboratory were selected to supplement the NWS WFO at Salt Lake City. Among other responsibilities, including warning and forecast coordination and fielding media inquiries, the supplementary NWS forecasters issued forecasts of weather, wind, temperature, wind chill, and precipitation type and amount for the primary transportation corridors twice daily for 12 locations. As part of their training, supplementary NWS forecasters issued practice transportation corridor forecasts for the World Cup events held in northern Utah in February–March 2001. The KSL forecast team, responsible for forecasting at the five outdoor venues, was composed of 13 private-sector meteorologists with extensive forecasting experience in northern Utah (Horel et al. 2002b). Venue forecast parameters included sky cover, precipitation type and amount, air temperature, wind direction, wind speed, wind gusts, wind chill, visibility, humidity, and snow temperature. Assembled in 1999, venue forecasters received training during pre-Olympic and World Cup testing and training events held over two winters prior to the Olympics. Together, the supplementary NWS forecasters and KSL venue forecasters represented the primary forecasting arm of the Olympic Forecast Team.

This chapter describes and evaluates mesoscale-model-based MOS forecasts that were used for finescale weather prediction during the 2002 Olympic and Paralympic Winter Games. The MOS forecasts were provided for 18 low-, mid-, and high-elevation sites at Olympic venues and along transportation corridors. At some venues, such as the Snowbasin ski area, MOS forecasts were generated for multiple observing sites that were less than 1 km apart. This chapter compares the accuracy of MOS to that of its parent high-resolution mesoscale model and human-generated forecasts produced by the Olympic Forecast Team and provides insight into the variability of MOS performance across a wide range of local topographic influences and meteorological conditions.

The remainder of this chapter is organized as follows: The data and methods used to generate and verify the MOS forecasts are presented in the next section. Then, an evaluation of MOS performance, including a comparison with that of the high-resolution numerical model and manual forecasts, is described. A discussion of results then concludes the chapter.

## Data and Methods

MOS equations were generated for 18 sites in the Olympic region using three winters (November–April 1998/99, 1999/2000, and 2000/01) of observations and mesoscale model forecasts. The sites were located at outdoor Olympic venues and along major transportation corridors, ranged in elevation from 1288 to 2835 m, and, despite their close geographic proximity, sampled a wide range of meteorological conditions (Table 2.1; Fig. 2.1). Observations from the 18 sites were provided by the MesoWest cooperative networks, which collect, integrate, and disseminate weather observations from more than 70 organizations in the western United States (Horel et al. 2002a). Eight organizations

Table 2.1. MM5 MOS sites. See Fig. 2.1 for locations. The "T" indicates a transportation corridor forecast site, and "V" indicates an outdoor venue forecast site.

| Site | ID | Elev (m) | Site Type | Venue/Transportation Corridor |
|---|---|---|---|---|
| Deer Valley-Bald Eagle | DVB | 2591 | Mountain | Alpine/freestyle skiing |
| Deer Valley-Burns (V) | DVE | 2235 | Mountain | Alpine/freestyle skiing |
| Park City-Eagle | PCS | 2610 | Mountain | Alpine skiing/snowboarding |
| Snowbasin-Allens Pk | SBB | 2835 | Mountain | Alpine skiing |
| Snowbasin-John Paul | SBW | 2670 | Mountain | Alpine skiing |
| Utah Olympic Park-Bear Upper (V) | WBU | 2182 | Mountain | Ski jumping/luge/skeleton/ bobsled |
| Parleys Summit (T) | PSS | 2146 | Mountain | I-80 (mountain pass) |
| Park City-Base (T, V) | PCB | 2000 | Mountain valley | Alpine skiing/snowboarding |
| Park City Municipal Golf Course | SNC | 1937 | Mountain valley | Alpine skiing/snowboarding |
| Snowbasin-Base (V) | SBE | 1925 | Mountain valley | Alpine skiing |
| Soldier Hollow (V) | WMP | 1713 | Mountain valley | Nordic skiing/biathlon |
| Soldier Hollow Whales Tail | WM2 | 1689 | Mountain valley | Nordic skiing/biathlon |
| Bluffdale | UT7 | 1433 | Wasatch Front | I-15 |
| Hill AFB | HIF | 1459 | Wasatch Front | I-15/I-84 |
| Ogden-Hinkley Air-port (T) | OGD | 1362 | Wasatch Front | I-15/I-84 |
| Parleys Canyon Mouth | UT5 | 1498 | Wasatch Front | I-80 |
| Provo (T) | PVU | 1369 | Wasatch Front | I-15 |
| Salt Lake City Interna-tional Airport (T) | SLC | 1288 | Wasatch Front | Aviation hub |

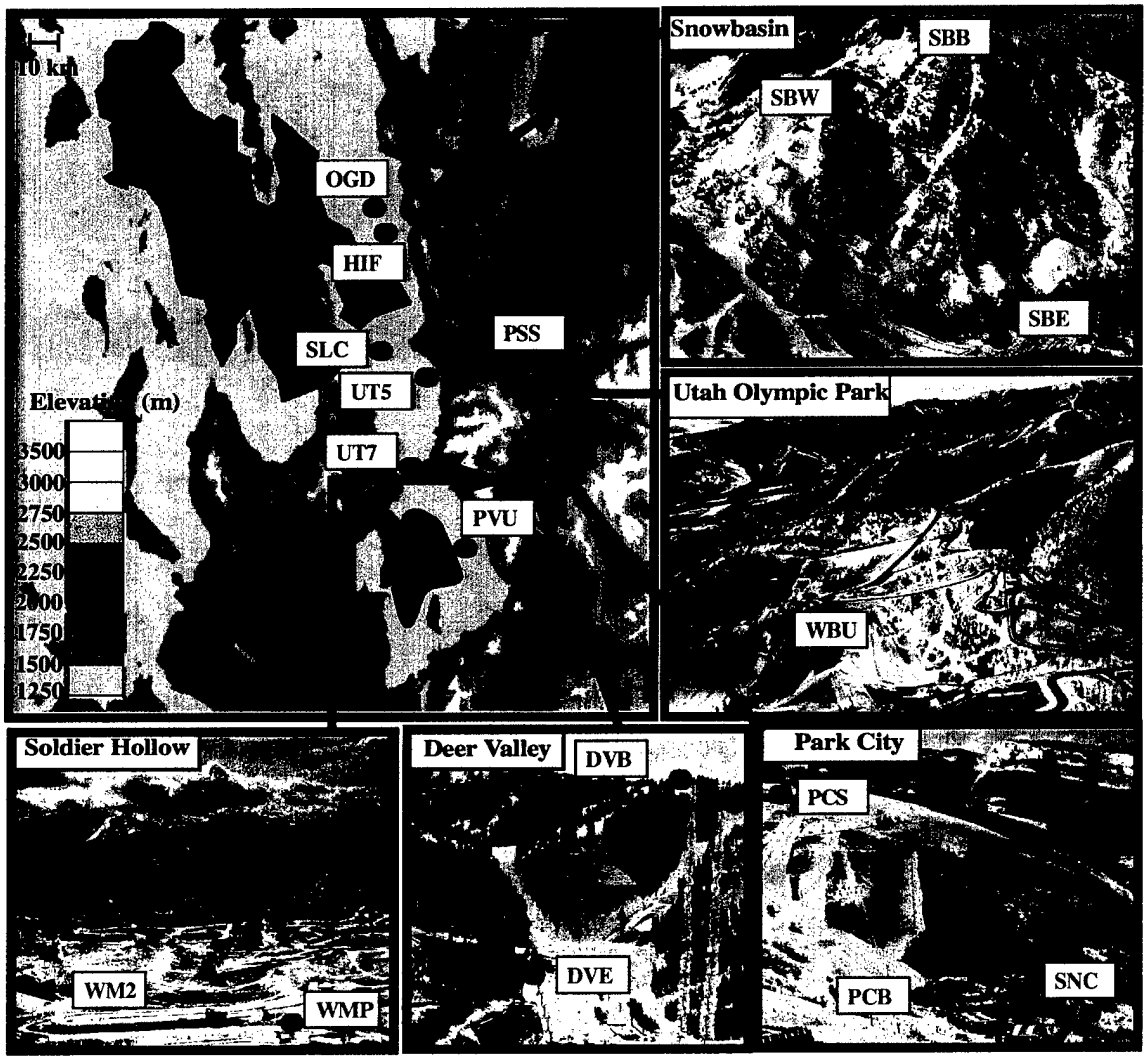



Figure 2.1. Geography of the Olympic region, outdoor Olympic venues, and MOS sites (green dots for mountain, red for mountain valley, and blue for Wasatch Front). Additional MOS site information is available in Table 2.1.

contributed the data used for this study, with the observations collected by a heterogeneous mix of sensors that use several different averaging and reporting intervals. For simplicity, all observations within 30 min of the top of each hour were averaged to generate a more homogeneous dataset for MOS development, implementation, and validation. A simple gross-range-check algorithm was used to eliminate clearly erroneous observations. For example, relative humidity was constrained to be from 0.1% to 100%, temperature from -70° to 135°F (-56.7° to 57.2°C), dewpoint from -90° to 100°F (-67.8° to 37.8°C), wind speed from 0 to 200 m $s^{-1}$, and wind direction from 0° to 360°. No additional quality checks were attempted. MOS equations were not developed for precipitation because a longer period of record was required to fully sample the large variability of precipitation events.

Mesoscale model forecasts were provided by a real-time version of the MM5 (Grell et al. 1994) that was run at the University of Utah (Horel et al. 2002b). During the three-winter MOS equation development period, the real-time MM5 featured two domains with 36- and 12-km grid spacing. Forecasts produced by the 12-km domain were used for MOS equation development. Although changes were made to the MM5 during the equation development period, no efforts were made to account for these changes in the MOS equations, as is done with updatable MOS (e.g., Wilson and Vallée 2002). Early tests suggested that traditional MOS produced adequate performance (Siffert 2001), in agreement with the findings of Wilson and Vallée (2002) who found that MOS forecast equations remained remarkably stable despite changes in the driving model. MM5 changes during the development period included an upgrade from MM5 version 2 to MM5 version 3, migration of the modeling system from an SGI Origin 2000 to a PC-based

Linux cluster, expansion of the 12-km domain, the addition of a 4-km domain, and incorporation of high-density MesoWest surface observations into the initial analysis provided by the NCEP Eta Model (Horel et al. 2002a; Lazarus et al. 2002). To limit the impact of these changes on MOS performance, the topography of the 12-km domain was not changed over the Olympic region, and feedback from the 4-km domain to the 12-km domain was not allowed. Throughout the MOS development and Olympic forecast periods, the real-time MM5 featured the convective, radiative, boundary layer, and microphysical parameterizations described by Kain and Fritsch (1993), Dudhia (1989), Hong and Pan (1996), and Dudhia (1989), respectively. Additional information concerning the real-time MM5 is provided by Horel et al. (2002b).

Following Glahn and Lowry (1972) and Glahn (1985), separate MOS equations were developed for each site, model cycle (0000 and 1200 UTC), variable, and forecast hour at 3-h intervals (e.g., 3, 6, 9,...36 h) using forward-stepwise multivariate linear regression. Hourly forecasts were then generated by weighting the 3-h coefficients in time. For example, a 7-h forecast equation reflected 2/3 influence from the 6-h forecast equation and 1/3 influence from the 9-h equation. Predictands (i.e., the variables that were being predicted) were temperature, dewpoint, relative humidity, wind speed, zonal wind, and meridional wind for the sensor level. MOS wind direction was derived from the zonal and meridional wind components. Note that MOS relative humidity was not derived from MOS temperature and dewpoint forecasts but was based on its own forecast equations since early work suggested such an approach was more accurate. Predictors (i.e., variables that may be related to each predictand) included MM5 model forecast variables (interpolated to the station location), geoclimatic predictors, and surface observations

(Table 2.2). For simplicity, only predictors at the forecast hour for which the equation was being developed were used in each equation. Geoclimatic predictors consisted of sine and cosine of the normalized day of the year. Surface observations included observations 3 h after the MM5 initialization time and observations, used as persistence predictors, valid at the time of the forecast. Since equipment or network failures can result in observational data outages, secondary equations that excluded prior surface observations were developed and used when prior observations were not available. A commercial software package, JMP (Sall et al. 2001), was used to perform the forward-stepwise multivariate linear regression. The default JMP stopping criteria, which are based on the Wilk's Lambda, Pillai's Trace, Hotelling-Lawley Trace, and Roy's Max Root statistical tests (Sall et al. 2001), were used to determine when to stop adding potential predictors. Using this approach, MOS equations typically consisted of about 15 predictors, similar to that of Nested-Grid Model (NGM) MOS (Jacks et al. 1990).

Unless otherwise indicated, results presented in this chapter are based on hourly forecasts and observations from 23 January to 24 March 2002, which encompassed the Olympic (8–26 February) and Paralympic (7–16 March) Games. For the purpose of this evaluation, the 18 verification sites were separated into three groups based on terrain characteristics: mountain, mountain valley, and Wasatch Front (Table 2.1; Fig. 2.1). Elevations at the seven mountain sites in the Wasatch Mountains ranged from 2146 to 2835 m. Five mountain valley sites, located in valleys to the east of the Wasatch Mountains, were characterized by elevations from 1689 to 2000 m. The remaining six locations, classified as Wasatch Front sites with elevations from 1288 to 1498 m, were located in the lowlands immediately west of the Wasatch Range. Raw model output from

Table 2.2. MOS Predictors.

| Predictors | Levels |
|---|---|
| **MM5** | |
| Temperature | $\sigma$ = 0.995, 800 hPa, 700 hPa, 600 hPa, 500 hPa |
| Dewpoint | $\sigma$ = 0.995, 800 hPa, 700 hPa, 600 hPa, 500 hPa |
| Relative Humidity | $\sigma$ = 0.995, 800 hPa, 700 hPa, 600 hPa, 500 hPa |
| Zonal Wind Speed | $\sigma$ = 0.995, 800 hPa, 700 hPa, 600 hPa, 500 hPa |
| Meridional Wind Speed | $\sigma$ = 0.995, 800 hPa, 700 hPa, 600 hPa, 500 hPa |
| Total Wind Speed | $\sigma$ = 0.995, 800 hPa, 700 hPa, 600 hPa, 500 hPa |
| Geopotential Height | 800 hPa, 700 hPa, 600 hPa, 500 hPa |
| Vertical Velocity | 800 hPa, 700 hPa, 600 hPa, 500 hPa |
| 3-h Accumulated Precipitation | |
| Surface-700-hPa Lapse Rate | |
| Mean Surface-500-hPa Relative Humidity | |
| **Geoclimatic** | |
| Sine of Normalized Day of Year | |
| Cosine of Normalized Day of Year | |
| **Observations** | |
| Most recent available prior surface observation at same time of day as forecast projection | |
| Surface observations at 3 h after initialization time | |

the MM5 12- and 4-km domains was verified based on interpolation of lowest half-sigma level ($\sigma = 0.995$) gridpoint forecasts to the positions of the observing sites.

Since the MOS forecasts during the Games were not archived, the results were based on applying the MOS equations to archived MM5 forecasts. In general, the resulting MOS forecasts should be the same as those available to Olympic forecasters, although it is possible that, because of the arrival of late observations, primary equations were used at a few times that the Olympic forecasters would have had access to a forecast generated by a secondary equation.

Accuracy measures used were the bias error (BE) and mean absolute error (MAE). The BE represents the model tendency to overforecast or underforecast the observed quantity (Wilks 1995) and is defined as

$$BE = \frac{1}{N} \sum_{i=1}^{N} (f_i - o_i), \tag{2.1}$$

where $N$ is the number of forecast/observation pairs in the sample, $f_i$ is the forecast, and $o_i$ is the observation. Negative (positive) BE values indicate underforecasting (overforecasting). The MAE indicates the average magnitude of the forecast error and is defined as

$$MAE = \frac{1}{N} \sum_{i=1}^{N} |f_i - o_i| . \tag{2.2}$$

Forecasts produced by the 12-km MM5, 4-km MM5, and MOS were also compared to those of the Olympic Forecast Team. This involved the use of a smaller set of data corresponding to the variables and periods for which the Olympic Forecast Team provided forecasts. For transportation corridors, temperature forecasts produced from 3 to 25 February and 7 to 16 March were examined for five locations (Table 2.1). Temperature,

relative humidity, and wind speed forecasts produced from 1 to 24 February were also considered for five outdoor venues (Table 2.1).

## Results

### Temperature Forecasts

Averaged over all locations, there was little difference in the performance of the 12- and 4-km versions of MM5, with temperature MAEs of 3.27° and 3.21°C, respectively (Fig. 2.2). MOS was substantially more accurate, with a mean absolute error of only 1.71°C, a reduction of over 40% compared to the direct model output. MAEs for the 12-km MM5, 4-km versions of MM5, and MOS were largest at mountain valley sites. MOS errors were smallest at mountain sites, whereas the 12- and 4-km MM5 performed best at Wasatch Front sites.

One possible contributor to MM5 errors was differences between model and actual elevation. In an effort to determine the impact of such errors, 12-km MM5 and 4-km MM5 temperatures were adjusted to the actual station elevation by assuming a *U.S. standard atmosphere, 1976* lapse rate of 6.5°C $km^{-1}$. Mass et al. (2002) used this approach in their evaluation of MM5 forecasts over the Pacific Northwest. While this adjustment improved the performance at higher-elevation mountain locations, it either degraded or provided little improvement at Wasatch Front and mountain valley locations (Fig. 2.2). Causes for this degraded performance and possible implications for producing graphical forecast products are discussed later in the chapter.

A major contributor to MM5 temperature errors during the validation period was an inability to simulate cold-pool strength and evolution properly. Cold pools are topographically confined layers of air that are colder than layers immediately above and

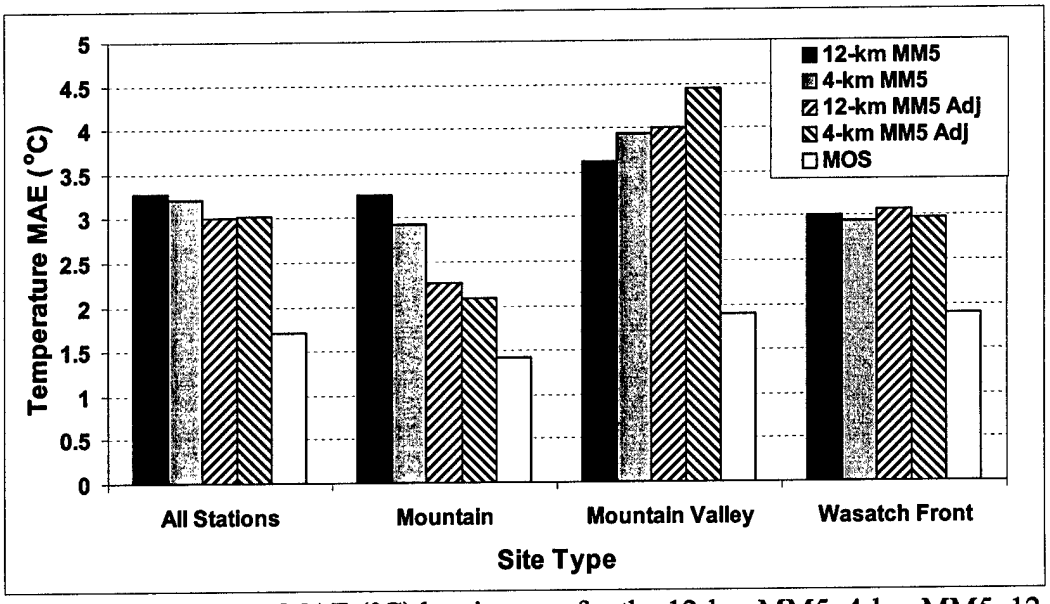



Figure 2.2. Temperature MAE (°C) by site type for the 12-km MM5, 4-km MM5, 12-km MM5 adjusted using a standard atmosphere lapse rate, 4-km MM5 adjusted using a standard atmosphere lapse rate, and MOS. MAE is averaged for all model runs and forecast hours (3–36).

can be characterized as either diurnal (i.e., nocturnal) or persistent (i.e., multiday) based on their duration (Whiteman et al. 2001). Persistent cold pools, characterized by unusually cold temperatures and limited boundary layer mixing, were present over the Wasatch Front from 28 January to 16 February, with a brief period of well-mixed conditions on 8 February accompanying the passage of a cold front. The temperature MAE time series[1] illustrates that consistently large errors were produced by the 12- and 4-km versions of MM5 during the cold-pool periods (Fig. 2.3a). In contrast, MOS errors were much smaller and typical of the Olympic and Paralympic periods as a whole. In general, MOS errors were more consistent throughout the 2-month period. A time series of the temperature bias error indicates that the 12- and 4-km versions of MM5 exhibited a large warm bias (~3°–4°C) during the persistent cold-pool events (Fig. 2.3b).

With the exception of two mountain sites, the 12-km MM5 and 4-km MM5 temperature MAEs were generally larger for stations with larger mean diurnal temperature cycles (Figs. 2.4a,b). Large contrasts between model and actual elevation were responsible for the anomalously high temperature MAEs at the two mountain sites. Adjusting model temperatures using a standard atmospheric lapse rate largely removed this error, moving the two mountain stations closer to a least squares fit line [the standard atmospheric lapse rate adjustment for temperature worked best for sites dominated by free-atmospheric temperatures (i.e., mountain sites)]. MOS temperature MAEs were smaller than those of the 12- and 4-km versions of MM5 (Fig. 2.2), particularly for stations with larger mean diurnal temperature cycles where MOS added the greatest value (Fig. 2.4c).

The 12-km MM5 and 4-km MM5 temperature MAEs were largest during the

[1] MAE and BE time series plots for the validation period are based on errors averaged for each model run at all forecast hours (3–36 every 3 h) for all 18 sites.

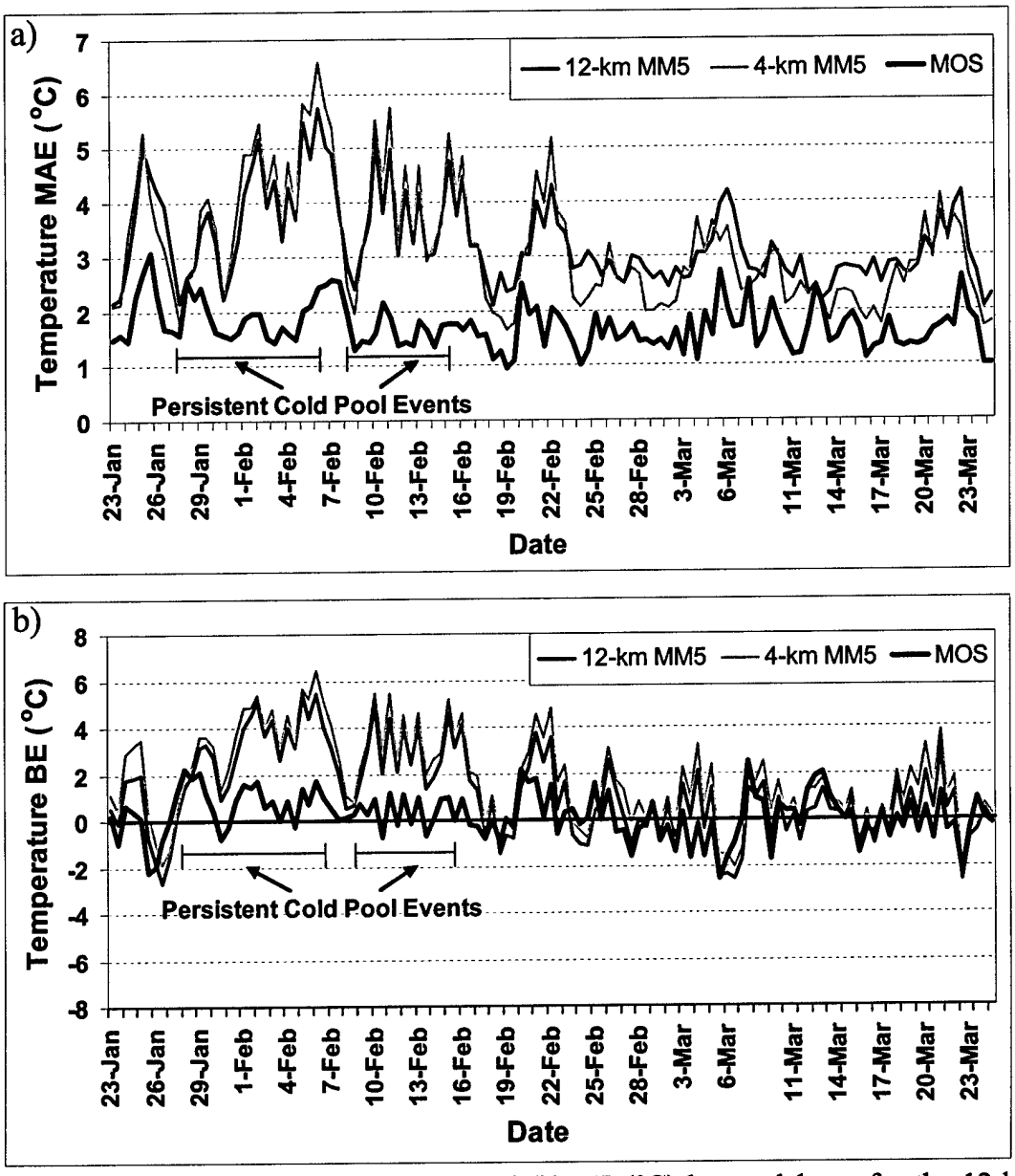



Figure 2.3. Temperature (a) MAE (°C) and (b) BE (°C) by model run for the 12-km MM5, 4-km MM5, and MOS. Both 0000 and 1200 UTC initialized model runs are included. MAE and BE for each model run are averaged for all forecast hours (3–36).

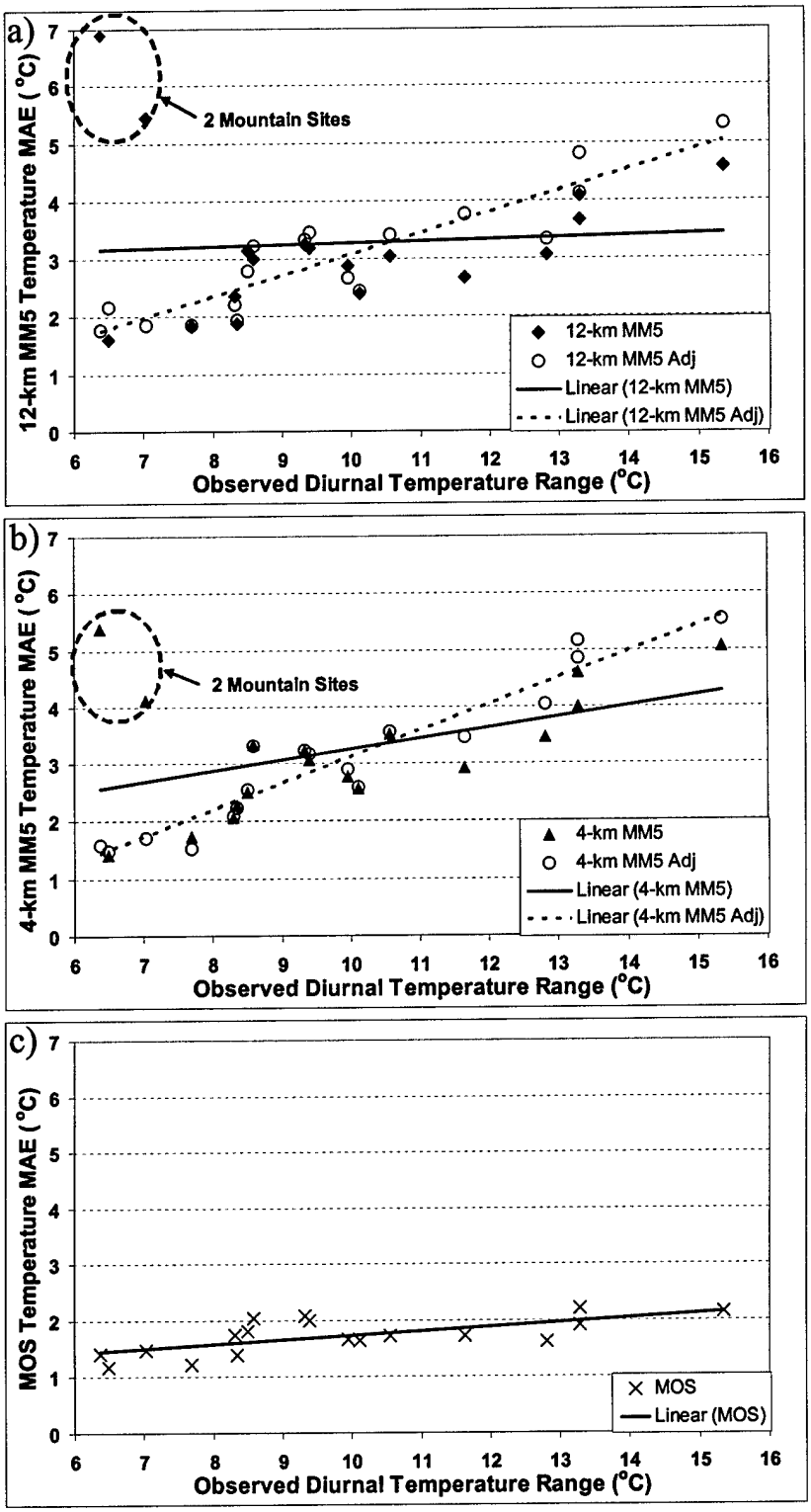



Figure 2.4. Temperature MAE (°C) vs observed mean diurnal temperature range (°C) for the (a) 12-km MM5, (b) 4-km MM5, and (c) MOS. Linear trend lines are displayed for MAEs based on model temperatures (solid), and model temperatures are adjusted to the actual elevation using a standard atmospheric lapse rate (dashed) in (a) and (b). MAE is averaged for all forecast hours (3–36) and model runs.

evening, overnight, and early morning hours (0000–1400 UTC) and smallest during the day (1400–0000 UTC) (Fig. 2.5a). As a result, MM5 errors did not increase monotonically with time, although for the 0000 (1200) UTC initialized MM5 forecasts, errors during the second nighttime (daytime) period were larger than the first, illustrating that model temperature errors grow with increasing forecast projection. MOS MAEs exhibited monotonic error growth at night and little or no growth during the day.

The 12-km MM5 and 4-km MM5 forecasts produced a large (2°–3°C) warm bias during the nighttime hours, suggesting that the model was unable to properly simulate the strength of nocturnal cold pools or stable layers (Fig. 2.5b). This warm bias was a major contributor to the large nighttime MAEs evident in Fig. 2.5a. During the day, relatively small (~1°C or less) 12-km MM5 and 4-km MM5 temperature bias errors were observed. The 12-km MM5 elevations were, on average, 100 m higher than the 4-km MM5 elevations at valley locations. As a result, the colder 12-km MM5 temperatures resulted in less bias than the 4-km MM5 since valley locations were often under the influence of nocturnal or persistent cold pools. The 12-km MM5 and 4-km MM5 nighttime warm biases were reduced from 2° to 3°C to 1°C by MOS. The large (2°–3°C) MM5 nighttime warm bias found in the present study contrasts with the smaller (~1°C) bias errors reported by Mass et al. (2002) in a verification study of a comparable resolution MM5 modeling system over the Pacific Northwest. Less cloud cover and more well-developed diurnal cold pools may help explain the larger nocturnal warm bias produced by the MM5 over northern Utah.

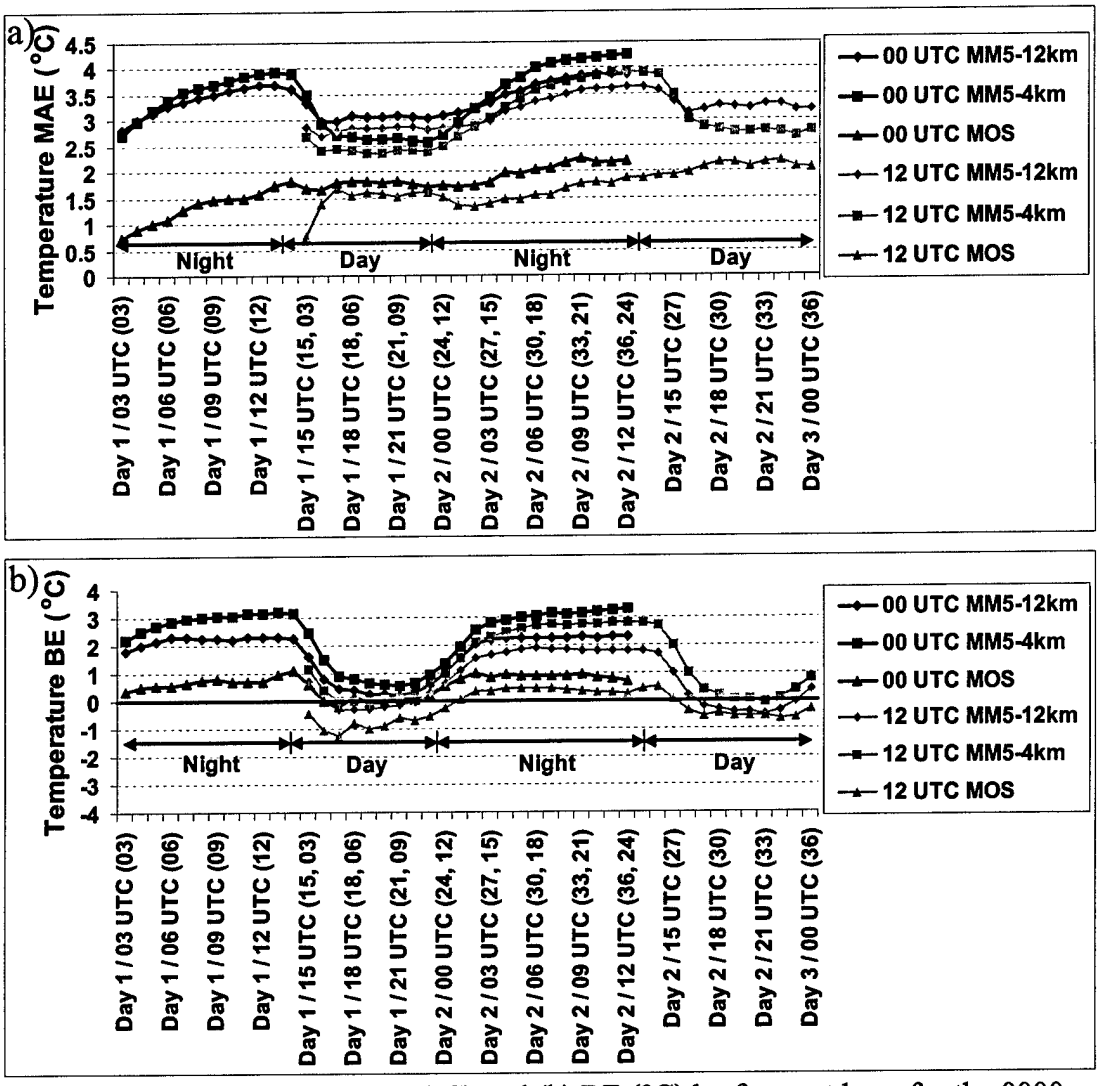



Figure 2.5. Temperature (a) MAE (°C) and (b) BE (°C) by forecast hour for the 0000 and 1200 UTC initialized 12-km MM5, 4-km MM5, and MOS.

## Dewpoint Forecasts

Unlike temperature, the 12-km MM5 and 4-km MM5 dewpoint MAEs during the persistent cold-pool episodes were typical of the Olympic and Paralympic periods (Fig. 2.6a). While MOS dewpoint errors were less than the 12- and 4-km versions of MM5 throughout the period, the value added by MOS was not substantially larger during the persistent cold-pool episodes. No significant BE was observed with the 12-km MM5 and 4-km MM5 dewpoint forecasts (0.27° and -0.38°C, respectively) (Fig. 2.6b). Overall, MOS overpredicted (higher than observed) dewpoint values for the 61-day period by 1.25°C. The largest dewpoint MAEs for the 12- and 4-km versions of MM5 occurred during the afternoon hours (1900–2300 UTC) (Fig. 2.7a). Once again, MOS errors were less than raw model errors at all hours. Most of the MM5 error during the afternoon was associated with a significant dry bias, with minimal dewpoint bias overnight (Fig. 2.7b).

## Relative Humidity Forecasts

Relative humidity MAEs for the 12- and 4-km versions of MM5 were 16% and 18%, respectively (Fig. 2.8). The comparatively worse performance of the 4-km MM5 may be related to temperature errors associated with model elevation errors as described earlier. MOS significantly outperformed the MM5 with a relative humidity MAE of ~10%. The greatest improvement by MOS over the MM5 was found at mountain valley locations, whereas the smallest improvement was for the mountain sites.

Like temperature, the 12-km MM5 and 4-km MM5 relative humidity MAEs were consistently higher during the two persistent cold pool events (Fig. 2.9). This might be expected since, during the cold events, neither the 12- or 4-km versions of MM5 accurately depicted the observed temperature (Fig. 2.3), which is used to determine the

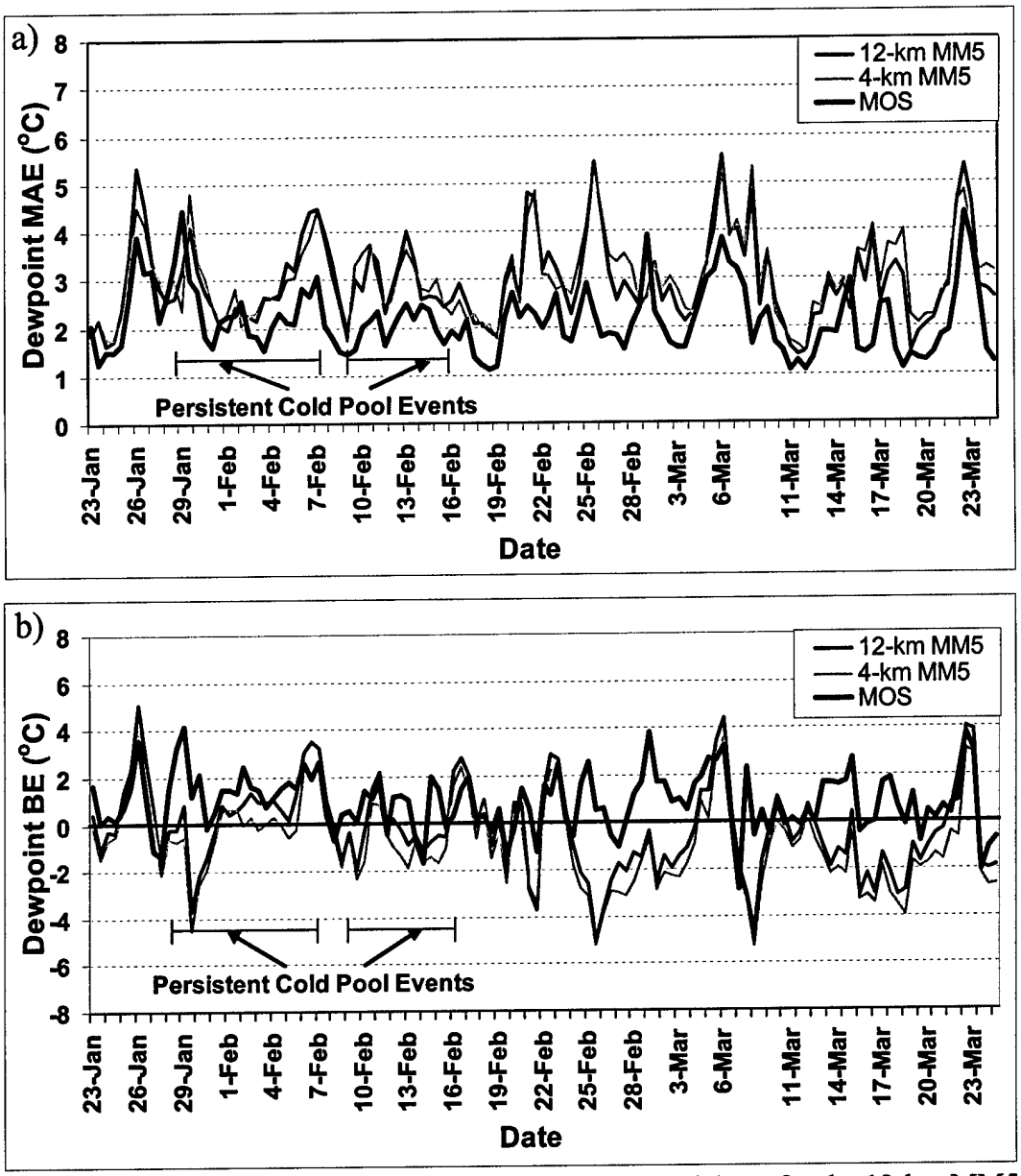



Figure 2.6. Dewpoint (a) MAE (°C) and (b) BE (°C) by model run for the 12-km MM5, 4-km MM5, and MOS. Both 0000 and 1200 UTC initialized model runs are included. MAE and BE for each model run are averaged for all forecast hours (3–36).

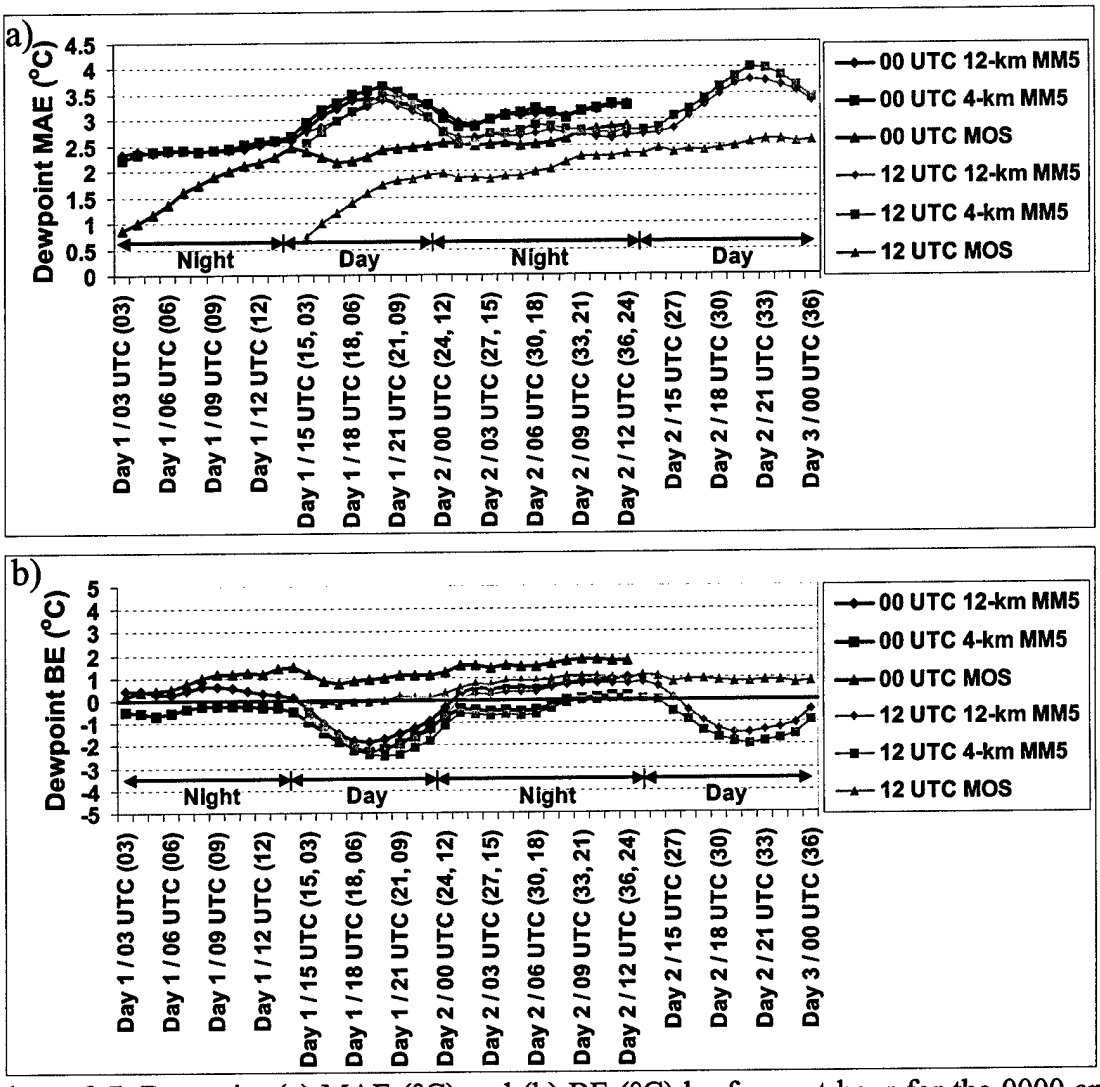



Figure 2.7. Dewpoint (a) MAE (°C) and (b) BE (°C) by forecast hour for the 0000 and 1200 UTC initialized 12-km MM5, 4-km MM5, and MOS.

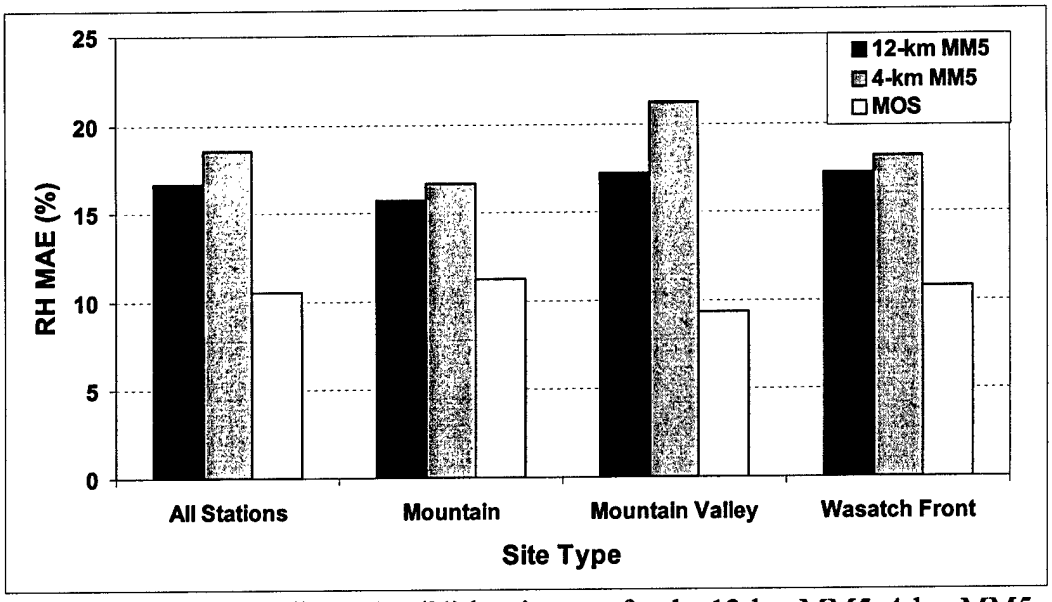



Figure 2.8. Relative humidity MAE (%) by site type for the 12-km MM5, 4-km MM5, and MOS. MAE is averaged for all model runs and forecast hours (3–36).

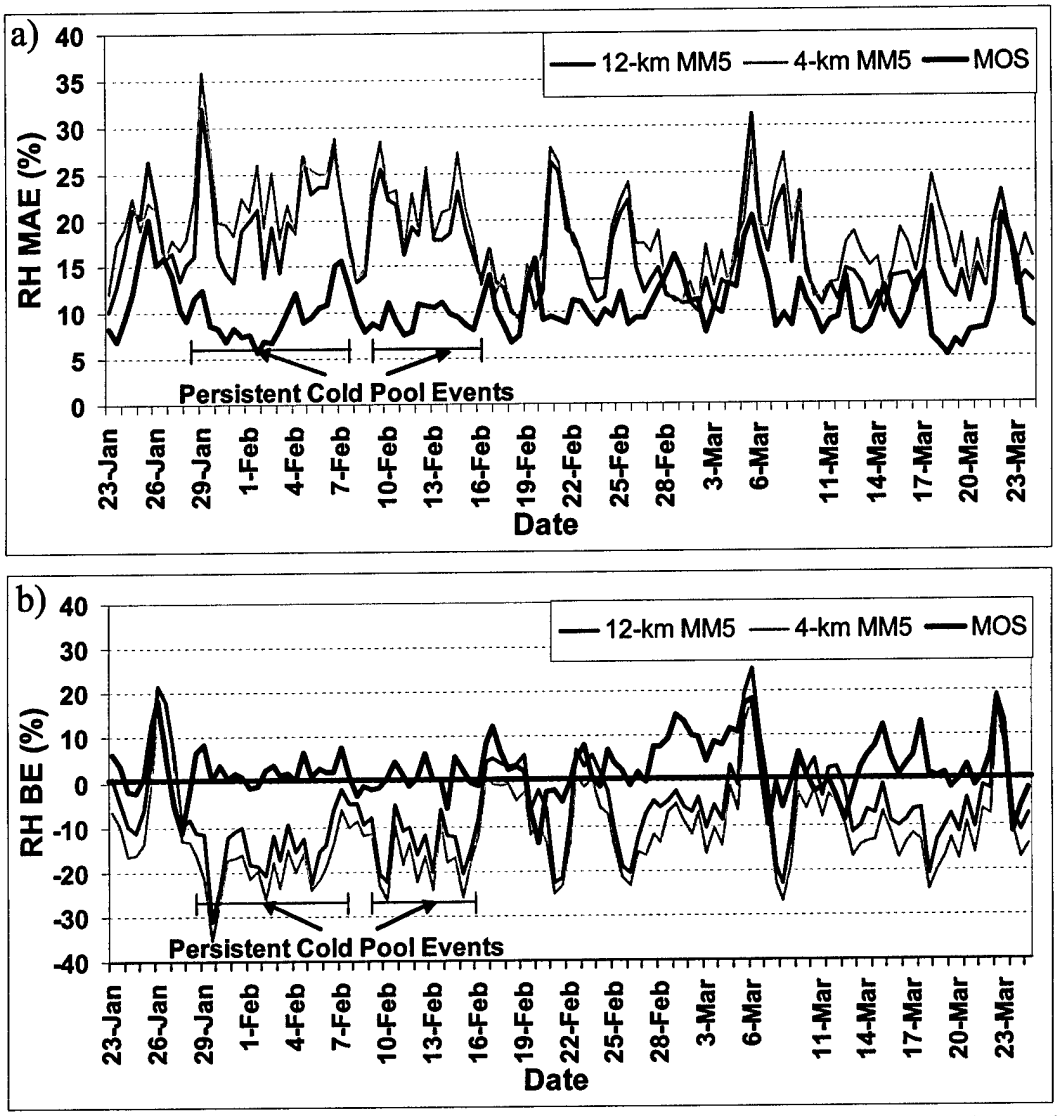



Figure 2.9. Relative humidity (a) MAE (%) and (b) BE (%) by model run for the 12-km MM5, 4-km MM5, and MOS. Both 0000 and 1200 UTC initialized model runs are included. MAE and BE for each model run are averaged for all forecast hours (3–36).

relative humidity. Episodes where the MM5 relative humidity MAEs were as large as those found during the persistent cold-pool events also occurred later in the Olympic period, and maxima in MOS relative humidity MAEs were also evident during these periods (e.g., 5–6 March and 22–23 March). Since MOS largely corrects for systematic model biases, poor model and MOS performance during these periods may be more a reflection of random model errors that arise from the poor placement and/or amplitude of synoptic and mesoscale systems.

The 12-km MM5 and 4-km MM5 relative humidity MAEs were largest at night, particularly from 0500 to 1500 UTC (Fig. 2.10a). MOS errors were less than the 12- and 4-km versions of MM5 at all hours and exhibited a more modest diurnal modulation. The relative humidity bias errors shown in Fig. 2.10b indicate that the 12- and 4-km versions MM5 were too dry, particularly at night. Since relative humidity is a function of both temperature and moisture, the strong overnight dry bias in relative humidity is most likely associated with errors in model temperature that result from the inability of the model to properly develop nocturnal cold pools. Causes for the larger dry bias in the 4-km MM5 may, again, be the result of the larger 12-km MM5 elevation errors that led to colder and, consequently, more accurate 12-km temperature forecasts during cold pool episodes. MOS bias errors were less than 5%. Overall, these results further illustrate the inability of the MM5 to properly simulate the structure of persistent and diurnal cold pools and stable layers.

## Wind Forecasts

Averaged over all stations, MOS wind speed MAEs were substantially lower (~1 m $s^{-1}$) than those of the 12- and 4-km versions of MM5 (3.3 and 3.2 m $s^{-1}$, respectively), a

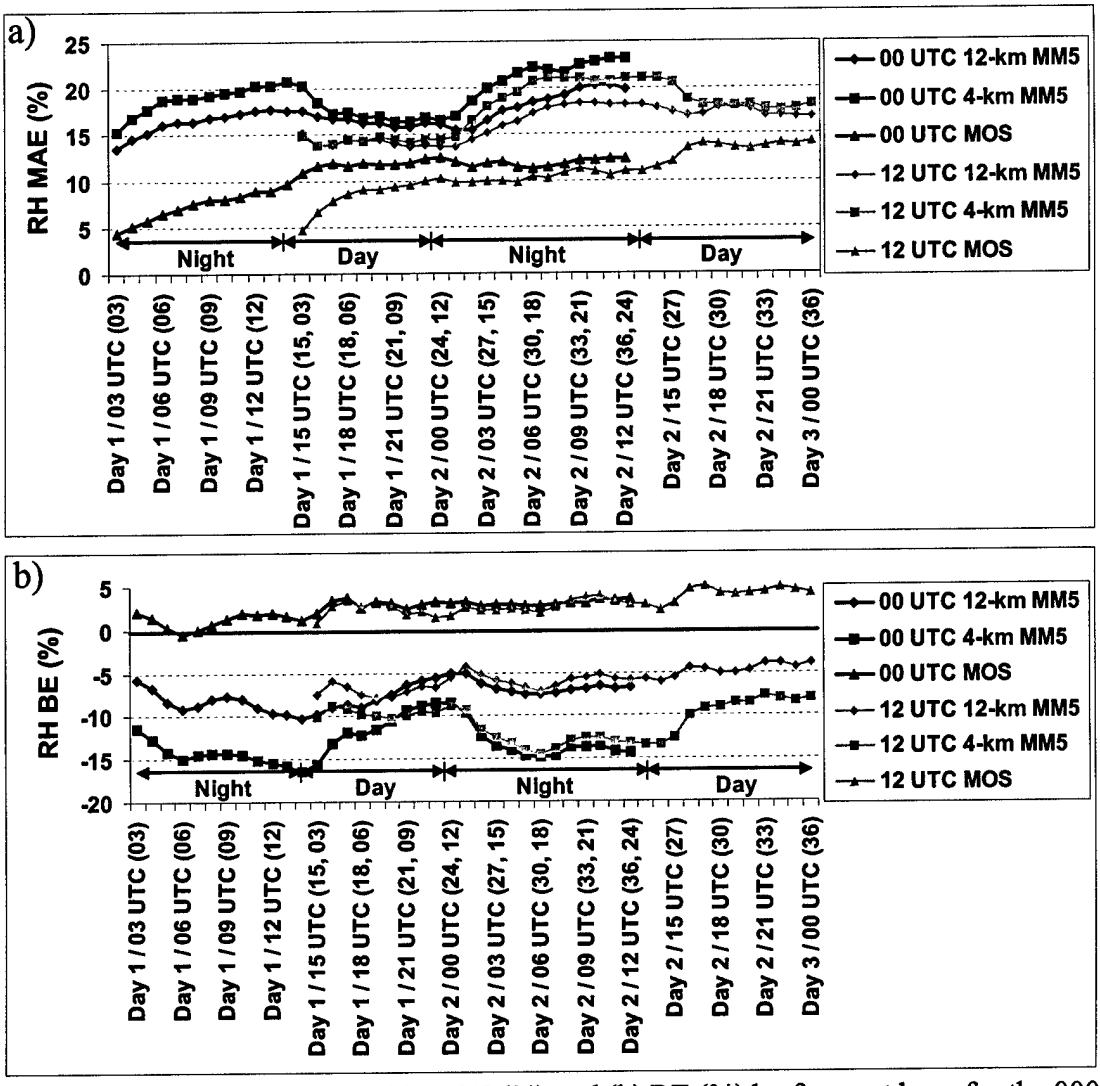



Figure 2.10. Relative humidity (a) MAE (%) and (b) BE (%) by forecast hour for the 0000 and 1200 UTC initialized 12-km MM5, 4-km MM5, and MOS.

reduction in error of over 65% (Fig. 2.11a). Although one might expect that there would be substantial improvement in wind speed skill with decreasing grid spacing, the 4-km MM5 forecast error was only marginally better than that of the 12-km MM5. At mountain valley locations, where the 12- and 4-km wind speed MAEs were the largest, MOS added the greatest value and featured very small MAEs. Wind direction MAEs for the 12- and 4-km versions of MM5, calculated for periods when wind speeds were at least 2.5 m $s^{-1}$, were 59° and 57°, respectively (Fig. 2.11b). MOS reduced these wind direction errors by 30% overall, with the greatest improvement found at mountain locations and only limited improvement evident at mountain valley and Wasatch Front sites.

There was a modest diurnal modulation of 12-km MM5 and 4-km MM5 wind speed MAE with the largest error during the night and early morning hours (Fig. 2.12a). This was likely related to the aforementioned problem simulating nocturnal cold-pool structure. The 12- and 4-km versions of MM5 overpredicted wind speeds, particularly during the overnight hours (Fig. 2.12b). There was no significant diurnal variability or overall bias, however, associated with wind speeds produced by MOS.

Of the variables examined in this study, wind speed and direction exhibit the greatest amount of spatial variability and are very sensitive to the topographic and landsurface characteristics of each observing site. This is particularly true of MesoWest observations, which for this study were provided by eight different networks designed for the needs of their operating agency or group. Some observing sites, including Deer Valley-Burns (DVE), Snowbasin-Base (SBE), Park City Golf Course (SNC) and Utah Olympic Park-Upper Bear (WBU), were installed to collect precipitation and were located in well-sheltered areas with limited wind exposure. As a result, wind speed magnitude and

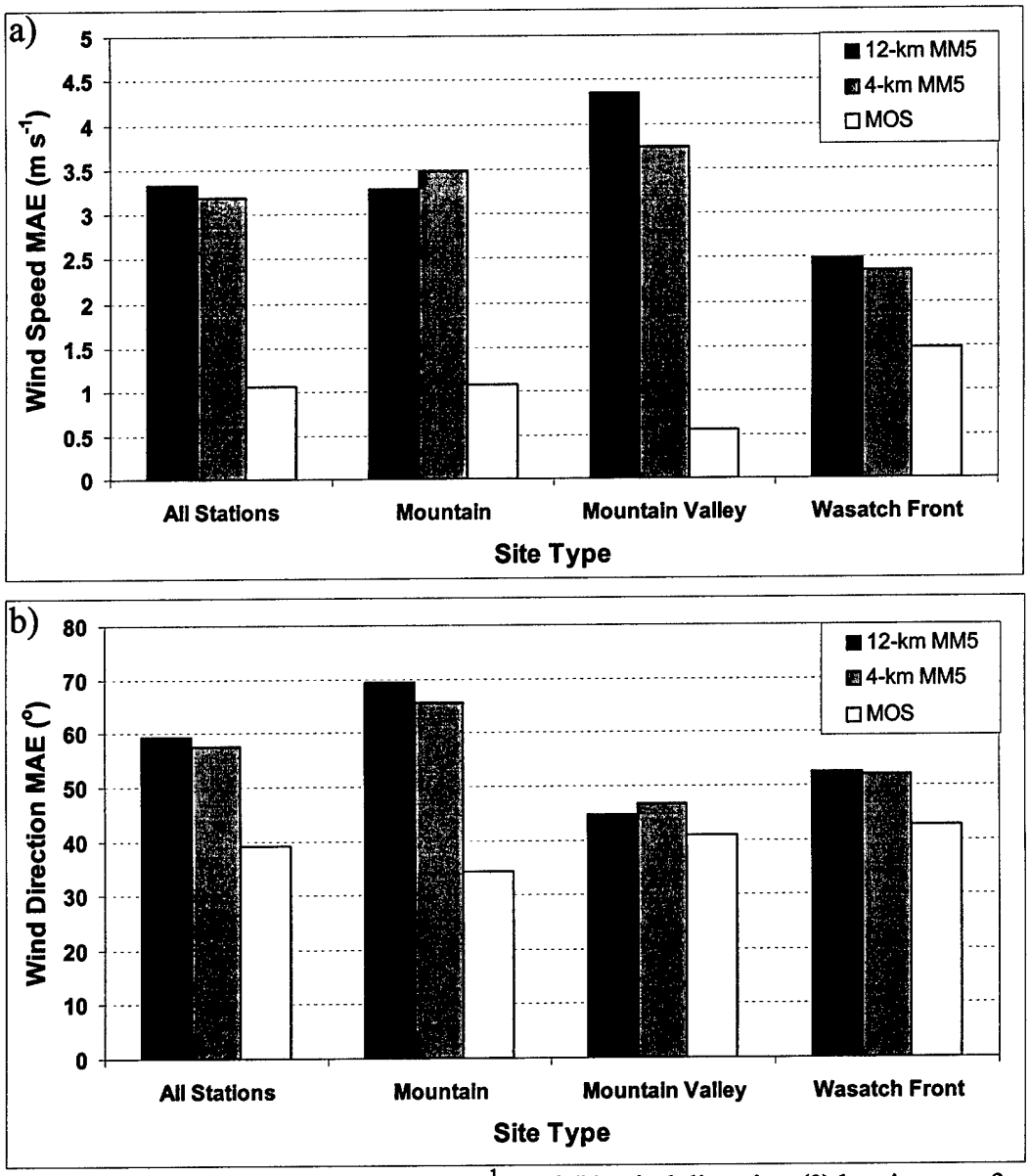



Figure 2.11. MAE for (a) wind speed (m $s^{-1}$) and (b) wind direction (°) by site type for the 12-km MM5, 4-km MM5, and MOS. MAE is averaged for all model runs and forecast hours (3–36).

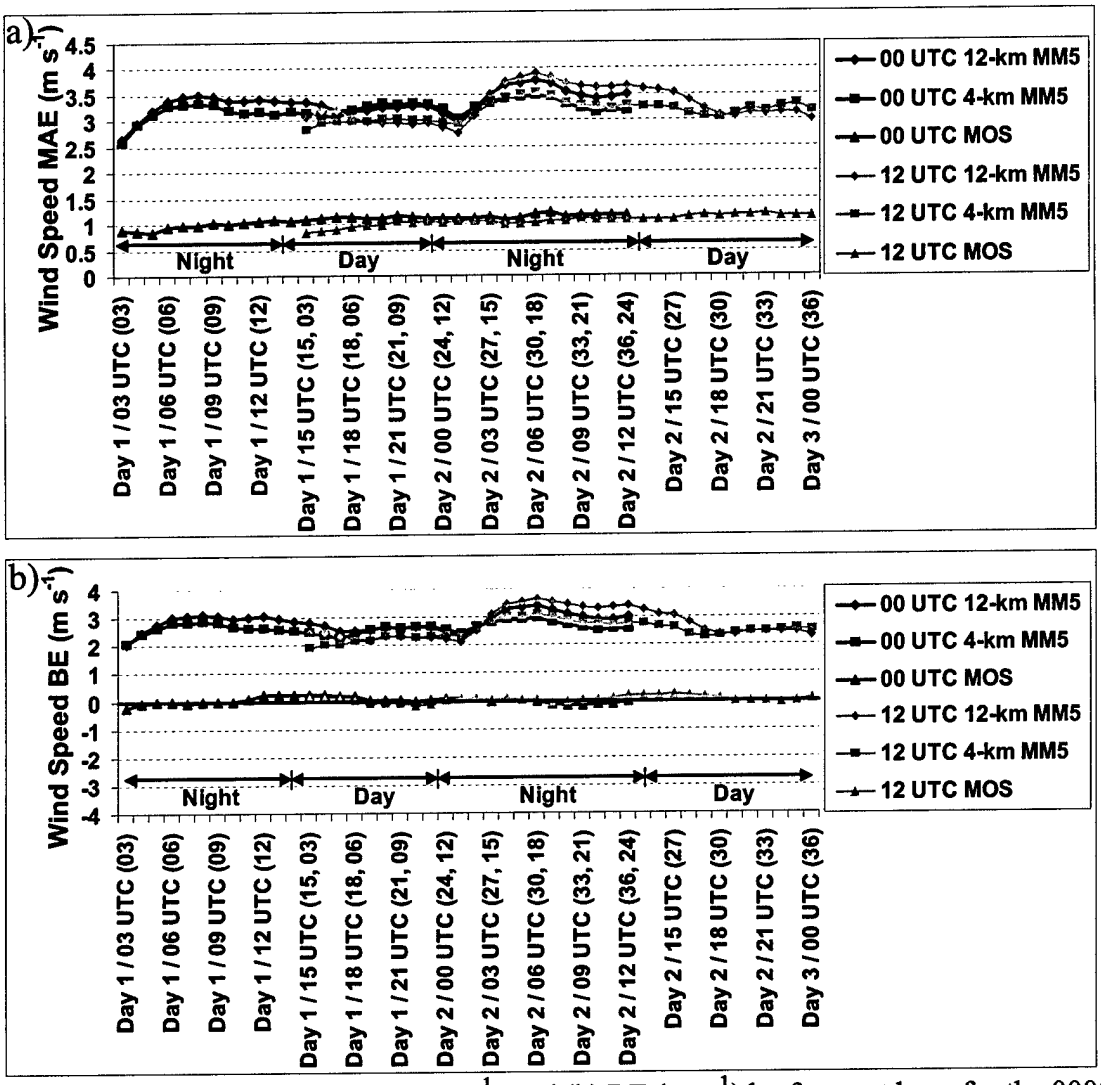



Figure 2.12. Wind speed (a) MAE (m $s^{-1}$) and (b) BE (m $s^{-1}$) by forecast hour for the 0000 and 1200 UTC initialized 12-km MM5, 4-km MM5, and MOS.

variability at these locations were typically lower than at other MesoWest sites.

The 12-km MM5 and 4-km MM5 wind speed MAEs were largest for sites with small observed wind speed standard deviations (Figs. 2.13a,b). At these sites, the grid cell averaged wind produced by MM5 was not representative of the unique local conditions at the observing sites. Smaller MAEs were found at more exposed sites that exhibited greater variability. In contrast, MOS MAEs grew with increasing wind speed standard deviation (Fig. 2.13c). In other words, as the variability of the observations increased, MOS performance decreased. At locations where there was less wind speed variability, MOS outperformed the 12- and 4-km versions of MM5 by a large margin because it corrected for the unique local conditions at these observing sites. At more exposed sites, such bias corrections were smaller and the larger MOS MAEs were a reflection of the greater variability and lower predictability of wind speed. These results also suggest that the influence of terrain and land-use heterogeneities on local wind observations should be considered when performing model verification studies, particularly in mountainous regions.

## Variations in Initial Conditions

The results described above were based upon MM5 simulations whose initial conditions were provided by the NCEP Eta Model initial analysis, which was enhanced at low levels based on high-density surface observations provided by the MesoWest cooperative networks (Horel et al. 2002a). During the Olympic and Paralympic periods, a second simulation using the NCEP AVN model for the initial-condition first guess was run at each initialization time. These simulations were referred to as MM5-Eta and MM5-AVN, respectively. The MOS equations developed for the MM5-Eta were also applied to

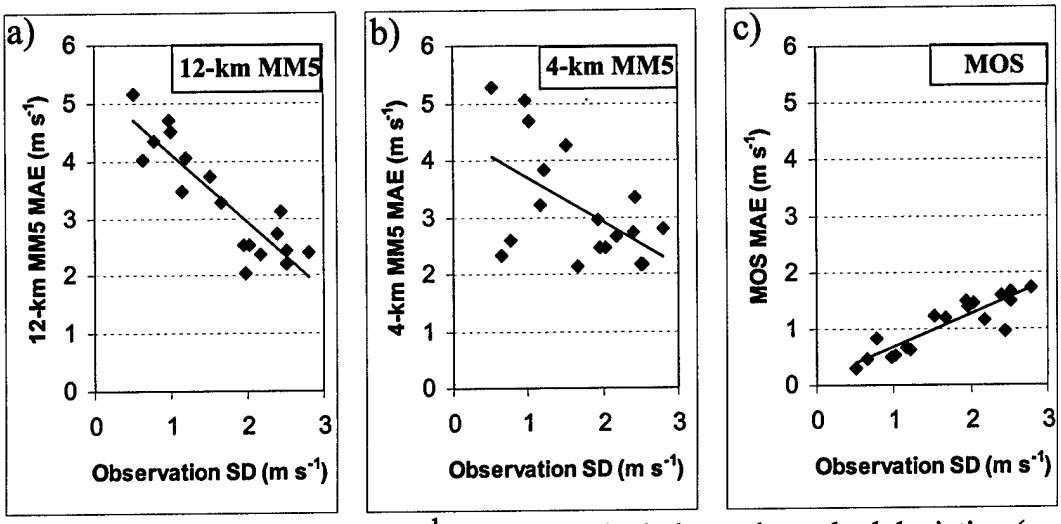



Figure 2.13. Wind speed MAE (m $s^{-1}$) vs observed wind speed standard deviation (m $s^{-1}$) for the (a) 12-km MM5, (b) 4-km MM5, and (c) MOS. MAE and observed standard deviation are averaged for all model runs and forecast hours (3–36).

the MM5-AVN, providing an opportunity to examine the variability of MOS performance produced by differing initial conditions.

Figure 2.14 presents the 36-h temperature, relative humidity, and wind speed MAEs from the MM5-Eta MOS, MM5-AVN MOS, and the forecast obtained by averaging the MM5-Eta and MM5-AVN MOS predictions (consensus MOS). The 36-h forecasts were considered independently to allow for the greatest potential of divergent solutions. There was remarkably little sensitivity of MOS errors during this period to differences in the initial and lateral conditions provided by the NCEP Eta and AVN models (Fig. 2.14). Strong local forcing mechanisms resulting from the complex terrain of northern Utah may have limited the effect of lateral boundary condition errors, or limited-area model simulations based on small perturbations in initial conditions may not have produced solutions that diverge like those of an unbounded model (Warner et al. 1997). The consensus MOS performance exhibited no significant improvement over the individual performance of MM5-Eta MOS or MM5-AVN MOS (Fig. 2.15). Similar results were obtained for relative humidity and wind speed (not shown). In contrast, Vislocky and Fritsch (1995) found that the consensus MOS forecasts calculated from two different models [e.g. NGM and Limited-Area Fine-Mesh Model (LFM)] improved performance over MOS forecasts from individual models. In the present study, perhaps the limited sensitivity of the MM5 to Eta and AVN initial conditions limits the value added by consensus MOS.

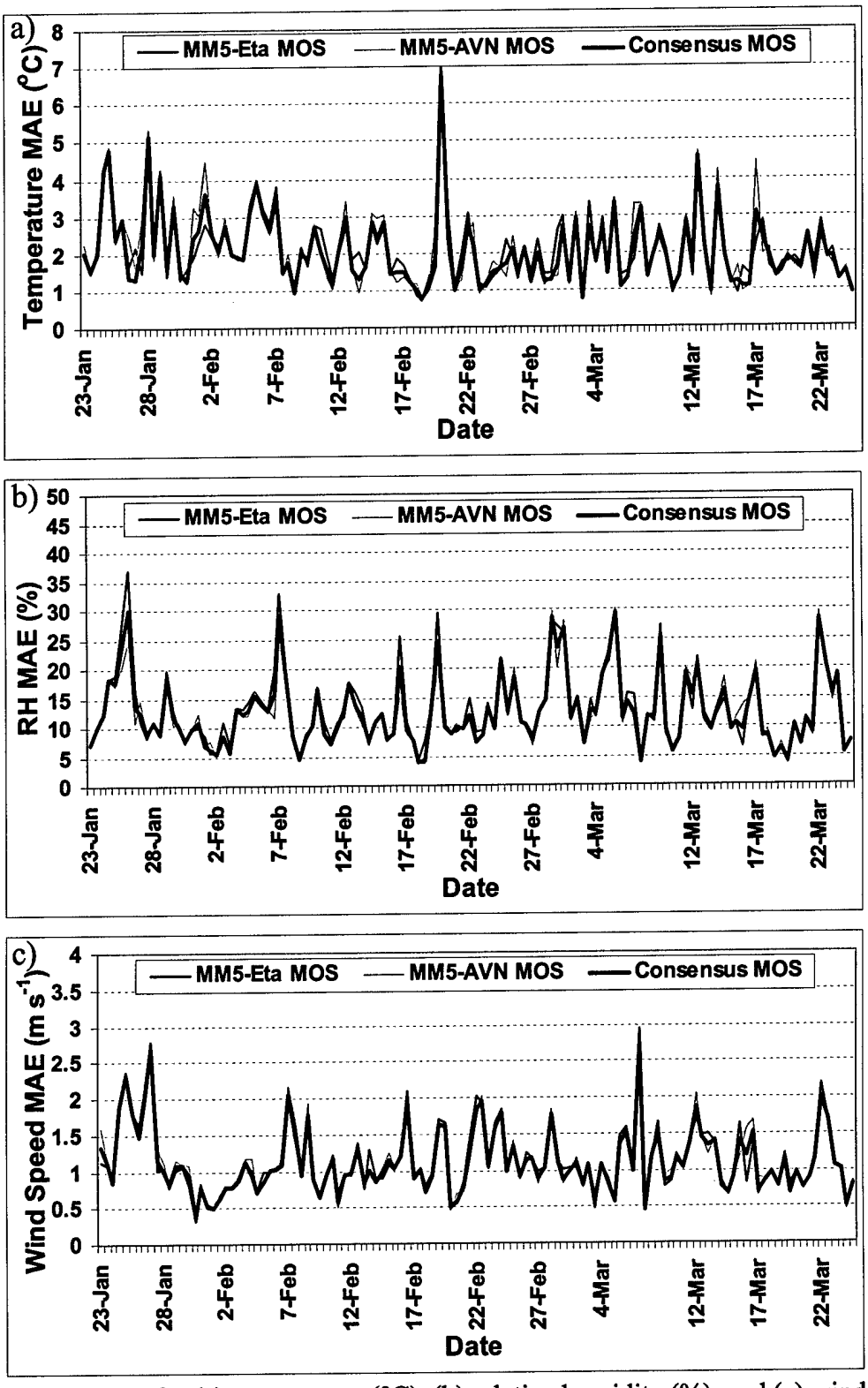



Figure 2.14. MAE for (a) temperature (°C), (b) relative humidity (%), and (c) wind speed (m $s^{-1}$) by model run for the 12-km MM5, 4-km MM5, and MOS. Both 0000 and 1200 UTC initialized model runs are included. MAE and BE for each model run are averaged for the 36-h forecast projection.

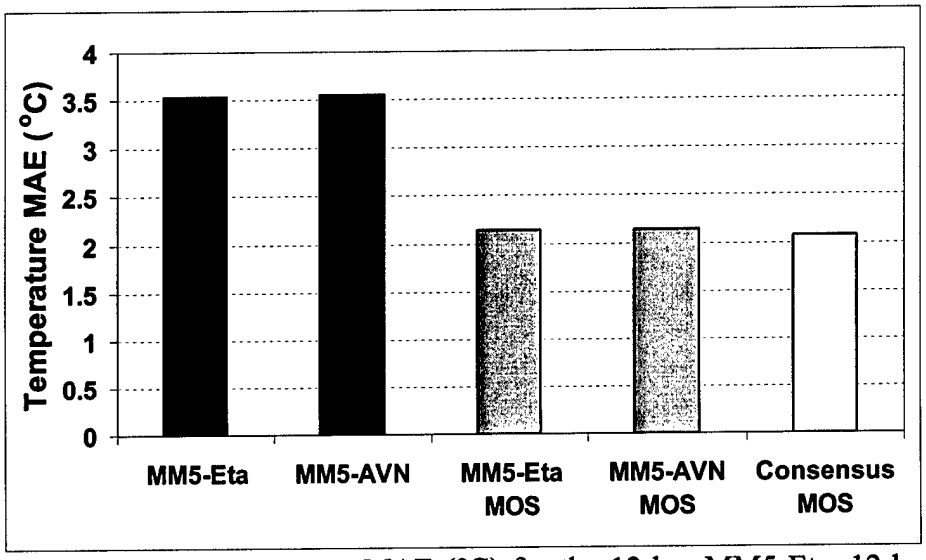



Figure 2.15. The 36-h temperature MAE (°C) for the 12-km MM5-Eta, 12-km MM5-AVN, MM5-Eta MOS, MM5-AVN MOS, and consensus MOS.

## Comparison with Olympic Forecast Team Transportation Route and Outdoor Venue Forecasts

The Olympic Forecast Team issued transportation corridor forecasts twice daily at ~2300 (the P.M. forecast) and ~1200 UTC (the A.M. forecast), including temperature forecasts valid at 1500, 1900, and 2300 UTC for the five MOS sites identified in Table 2.1. MM5 and MOS errors are presented for the model run used by the Olympic Forecast Team in producing their forecast (e.g., the 1200 UTC initialized MM5 for the P.M. forecast) and the model run corresponding to the time the Olympic Forecast Team issued their forecast (e.g., 0000 UTC for the P.M. forecast) (Fig. 2.16). Forecast projections for the 1200 and 0000 UTC model runs compared for the P.M. forecast issuance were 27/31/35 h and 15/19/23 h, respectively. For the P.M. forecast, the Olympic Forecast Team temperature MAEs (2.9°C) were comparable to those of the 12- and 4-km versions of MM5 for both the model run available to forecasters (1200 UTC) and the model run initialized at the time they issued their forecast (0000 UTC). MOS errors were, however, considerably smaller (~2°C), about 1°C less than those of the Olympic Forecast Team.

For the A.M. forecasts, a more complex picture emerges. Forecast projections for the 0000 and 1200 UTC model runs compared for the A.M. forecast issuance were 15/19/23 h and 7/11 h, respectively.[2] Olympic Forecast Team temperature MAEs (2°C) were about 1°C smaller than the 0000 UTC initialized 12- and 4-km versions of MM5 that were available for producing their forecast, and comparable to the 0000 UTC initialized MOS (~2°C). The improvement over the 1200 UTC 12-km MM5 and 4-km MM5, initialized at the time the Olympic Forecast Team forecast was issued, was smaller. The 1200 UTC

[2] Since 1200 UTC MOS uses the 1500 UTC temperature as a predictor, the A.M. forecast comparison for 1200 UTC products did not use the temperature forecast valid at 1500 UTC and is based only on the 7- and 11-h forecasts valid at 1900 and 2300 UTC.

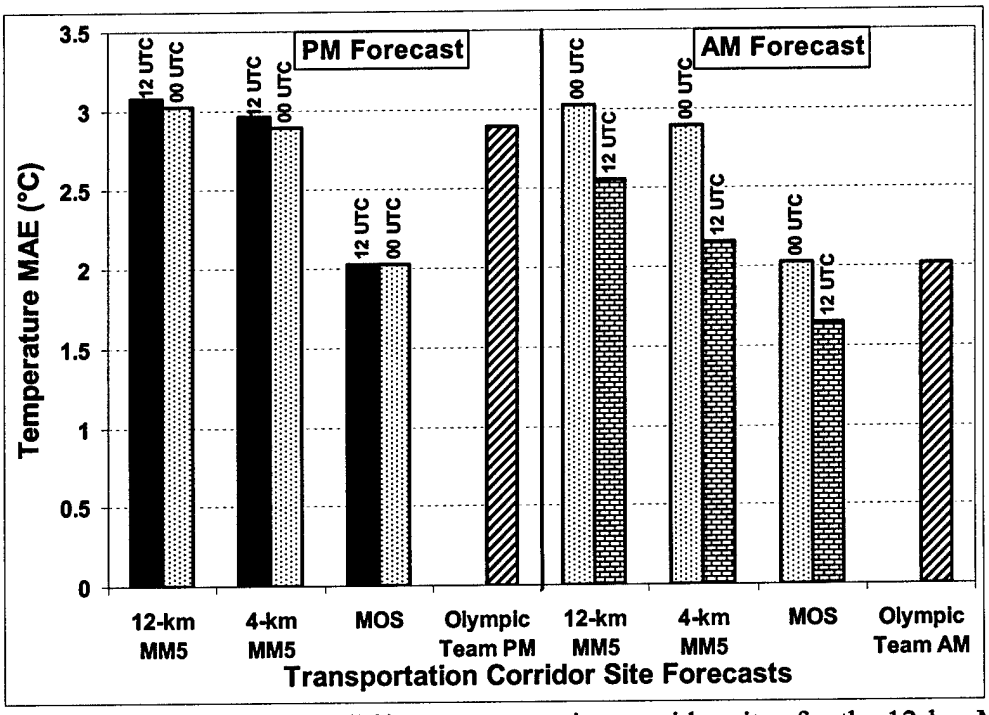



Figure 2.16. Temperature MAE (°C) at transportation corridor sites for the 12-km MM5, 4-km MM5, MOS, and Olympic Forecast Team forecasts issued at ~2300 (P.M. forecast) and ~1200 UTC (A.M. forecast). Validation periods are 3–25 Feb and 7–16 Mar 2002.

initialized MOS MAEs were 18% smaller than those of the Olympic Forecast Team.

From 1 to 24 February, the Olympic Forecast Team also issued forecasts for five Olympic venue locations (Table 2.1). For this period, Olympic Forecast Team forecasts of temperature, relative humidity and wind speed, issued for 3-h intervals through 36 h at ~0000 (the P.M. forecast) and ~1200 UTC (the A.M. forecast), were compared to the 12-km MM5, 4-km MM5, and MOS forecasts (Fig. 2.17). As with the transportation corridor forecasts, MM5 and MOS MAEs are presented for the model runs used by the Olympic Forecast Team in producing their forecast and the model run corresponding to the time that the Olympic Forecast Team issued their forecast. For the P.M. forecast, Olympic Forecast Team temperature MAEs (2.4°C) were significantly smaller than those of the 12- and 4-km versions of MM5 for both the model run available to forecasters (1200 UTC) and the model run initialized at the forecast issue time (0000 UTC) (Fig. 2.17a). MOS temperature MAEs for the model run available to forecasters (1200 UTC, 1.8°C) were smaller, however, than Olympic Forecast Team temperature MAEs. MOS temperature MAEs for the model run initialized at the time of forecast issue were even smaller (0000 UTC, 1.6°C). For the AM forecast, a similar result emerged. Olympic Forecast Team MAEs (2.1°C) were significantly less than those of both 12-km MM5 and 4-km MM5 runs, but larger than MOS temperature MAEs for the model run available to forecasters (0000 UTC, 2.0°C) and the model run initialized at the time of forecast issue (1200 UTC, 1.5°C), although the difference was not as large as found for the P.M. forecasts. Similarly, for both relative humidity (Fig. 2.17b) and wind speed (Fig. 2.17c), MOS MAEs were comparable or smaller than those of the Olympic Forecast Team.

The lower MAEs of MOS forecasts compared with those of the Olympic Forecast

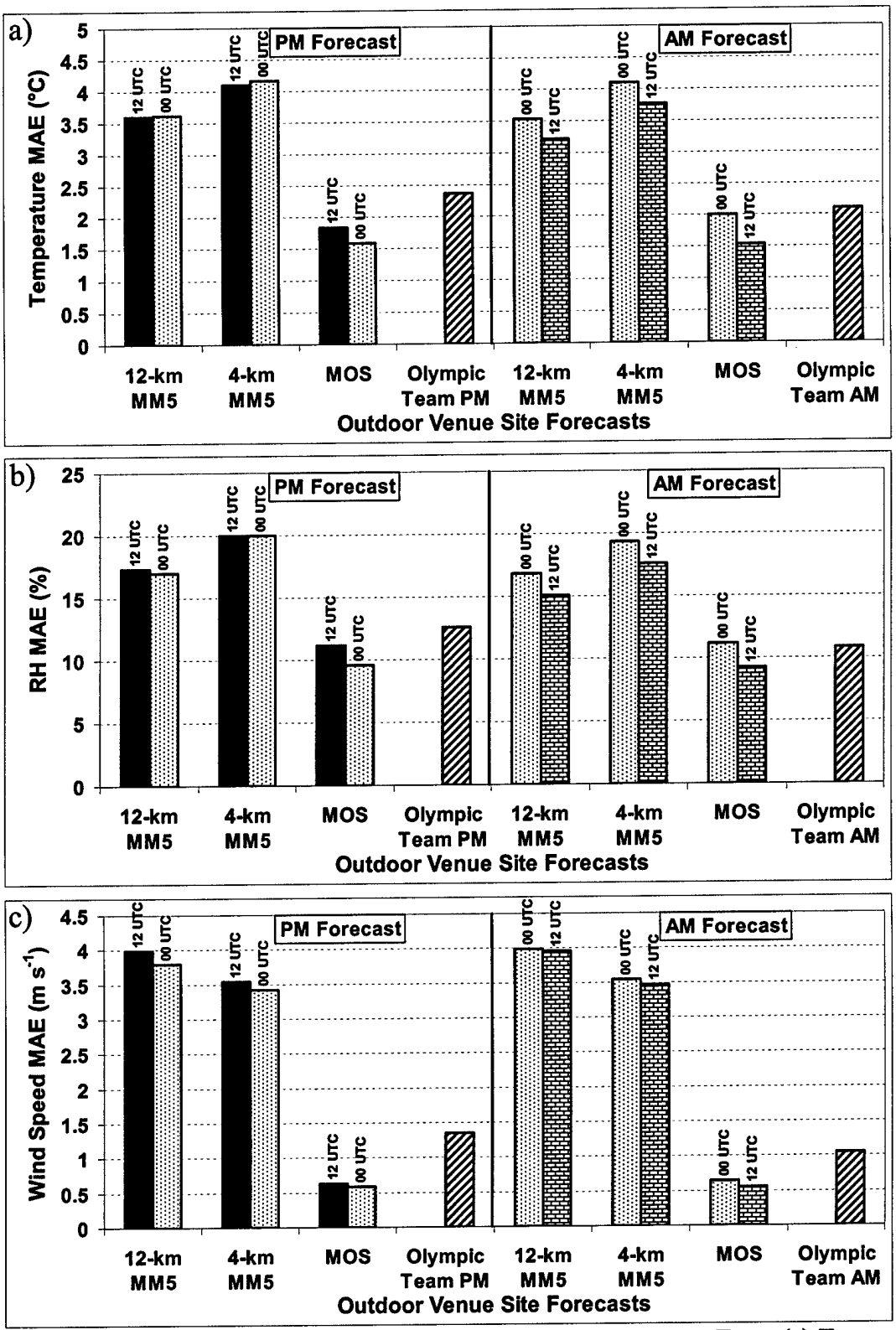



Figure 2.17. 12-km MM5, 4-km MM5, MOS, and Olympic Forecast Team (a) Temperature (°C), (b) relative humidity (%), and (c) wind speed (m $s^{-1}$) MAEs at outdoor venue sites for forecasts issued at ~0000 (P.M. forecast) and ~1200 UTC (A.M. forecast). Validation period is 1–24 Feb 2002.

Team may be surprising since many National Weather Service forecasters improve upon NGM and AVN MOS forecasts. For example, at the Salt Lake City forecast office from 1 December 2001 to 31 March 2002, human forecasts of maximum and minimum temperature at the Salt Lake City International Airport (SLC) and Cedar City Airport (CDC) featured an MAE of 1.99°C, compared with 2.36° and 2.17°C for NGM MOS and AVN MOS, respectively (M. E. Jackson, 2003, personal communication). The Olympic forecasts, however, featured considerably more detail than maximum and minimum temperature forecasts for two sites and included predictions for multiple variables at frequent time intervals, many of which are not presented in this study. In the case of the Olympic venue forecasts, temperature, relative humidity, and wind speed predictions were verified at 3-h intervals. Thus, the lower skill of the Olympic Forecast Team forecasts compared with MOS may be a reflection of the difficulties of predicting a large number of variables at frequent intervals. It is also possible that the relatively brief Olympic period did not feature weather regimes where the Olympic Forecast Team forecasters could improve substantially upon MOS.

## Implications for IFPS and the Future Role of Human Forecasters

The results presented above have a number of implications for operational weather prediction over the western United States, particularly in an era where the National Weather Service will prepare and disseminate gridded forecast products based on the Interactive Forecast Preparation System (IFPS; Glahn and Ruth 2003). First, for the sites examined in this chapter, decreasing model grid spacing from 12 to 4 km produced only marginal improvement in the prediction of surface temperature, wind speed, and wind direction and actually degraded the surface relative humidity forecast. This suggests that

increasing the resolution of a numerical model to match the proposed 2.5-km grid spacing of IFPS forecasts issued by the National Weather Service Western Region, as suggested by Mass (2003), will likely provide little, if any, improvement over guidance provided by existing coarser-resolution models such as the 12-km Eta Model. Second, it appears that some simple downscaling techniques, such as adjusting model temperatures for model and actual elevation differences using a *U.S. standard atmosphere, 1976,* may produce only marginal forecast improvements and, in some situations, such as when persistent or diurnal cold pools are present, actually degrade the forecast. While the assumed lapse rate can be adjusted in IFPS on a case-by-case basis, the considerable spatial variability in cold-pool strength and depth will make this a daunting task to perform accurately over an entire county warning area.

Perhaps the most important result is the dramatic gain in forecast accuracy produced using traditional MOS. Clearly, the ability to account and correct for systematic biases produces substantially larger forecast improvements than can be obtained by decreasing the horizontal grid spacing of existing models. Although improvements in observing technologies, data assimilation, and model numerics and physics will result in more accurate numerical forecasts with smaller biases in the future, it will likely take many years until model forecasts can be accurately applied to the prediction of surface weather variables.

Although it is possible to produce a gridded forecast using NGM, AVN, or Eta MOS with IFPS, the limited number and predominant valley siting of western U.S. MOS sites limits the utility of such forecasts. There is, however, a new possibility–applying MOS or other statistical techniques using high-density observations provided by the

MesoWest cooperative networks and other nonconventional networks.

Presently, MesoWest collects data from more than 3500 stations in the western United States (Horel et al. 2002a). Observations span all elevations because of contributions from organizations with meteorological sensors in valley locations (e.g., air quality and transportation agencies) and at high elevations (e.g., Bureau of Land Management and ski resorts). In some regions, data density is very high. For example, over northern Utah, data are collected from more than 250 sites (Horel et al. 2002a). A substantial historical record of many sites is available and could be applied for MOS equation development based on the Eta Model and the Global Forecast System. Other postprocessing approaches, such as Kalman filters (e.g., Homleid 1995; Majewski 1997), neural networks (Hsieh and Tang 1998), and group method of data handling (GMDH) algorithms (e.g., Abdel-Aal and Elhadidy 1995), could also be used and could possibly take advantage of output from ensemble modeling systems.

The need to utilize MesoWest or other high-density observing networks to produce more accurate gridded forecasts is further illustrated by the fact that MOS temperature, relative humidity, and wind speed MAEs were comparable to or smaller than those of the Olympic Forecast Team. This illustrates how difficult it is for humans to produce forecasts for a large number of variables at high temporal resolution and suggests that the development of MesoWest-based MOS would result in more accurate gridded first-guess fields for forecaster manipulation, and, in turn, better forecasts for the western United States. In addition, a more accurate first-guess field would also allow the forecaster to focus on other important components of the forecast process, including short-term forecasts, subjective evaluation of model analyses and forecasts, and communicating the

forecast to the public (Mass 2003).

One significant obstacle that remains, however, concerns how to use irregularly spaced, point-specific, MesoWest-based MOS to produce a gridded forecast. In particular, how could an accurate forecast be produced for data-sparse areas? Given the wide range of terrain geometries, how can one best interpolate the point forecasts to the grid? Should anisotropic weighting functions be used, as proposed by Lazarus et al. (2002)? Finally, for fields such as wind speed and direction, how will local terrain and land-use heterogeneities be considered? In addition, the small number of precipitation sites limits the ability to apply MesoWest observations for statistical correction of model precipitation forecasts. Clearly, issues remain for applying IFPS in the most efficient and effective manner.

## Conclusions

This chapter has examined the performance of a mesoscale-model-based MOS system that provided forecasts for 18 sites over northern Utah during the 2002 Olympic and Paralympic Games. MOS forecast accuracy was compared with direct output of the 12- and 4-km MM5 forecasts, as well as with that of the Olympic Forecast Teams.

Decreasing model grid spacing from 12 to 4 km produced marginal or no improvement in the prediction of surface temperature, relative humidity, wind direction, and wind speed. In contrast, MOS provided substantially better temperature, wind, and moisture forecasts. Large MM5 temperature forecast errors were produced at valley locations because of the poor handling of nocturnal inversions and persistent cold-pool events. These temperature forecast errors led to a dry relative humidity bias during the same periods. MOS reduced these errors by correcting for MM5 warm biases during such

stable periods. MOS added more value to wind forecasts at locations where there was the lowest observed wind speed variability. Variations in model initial and lateral boundary conditions (e.g., Eta or AVN) did not significantly affect raw model or MOS performance. MOS temperature, relative humidity, and wind speed MAEs at Olympic outdoor venue and transportation corridor sites were comparable to or smaller than those of the Olympic Forecast Team, highlighting the human challenges associated with forecasting multiple variables for high temporal resolutions.

These results suggest that statistical techniques can account for and correct systematic biases and improve upon purely dynamical predictions, even at fine grid spacings. With the development of high-density surface observation networks across the western United States (e.g., MesoWest), where detailed characteristics of local topography and microclimates remain unresolved, traditional MOS or other statistical forecast techniques should be incorporated into IFPS and other applications to produce more accurate gridded forecasts than can be obtained by downscaling direct model output to high-resolution grids.

# CHAPTER 3

# IMPROVEMENTS IN MODEL SKILL WITH DECREASING HORIZONTAL GRID SPACING OVER FINESCALE INTERMOUNTAIN TOPOGRAPHY DURING THE 2002 OLYMPIC WINTER GAMES

## Introduction

The Intermountain West has experienced rapid population growth during the last 10 years and, as a result, winter storms, severe winds, and, during quiescent synoptic regimes, air pollution associated with persistent cold pools in this region have a growing socioeconomic impact (e.g., Schultz et al. 2002). A primary goal of the United States Weather Research Program (e.g., Smith et al. 1997; Fritsch et al. 1998) is to improve short-term forecasts of weather phenomena that significantly affect society, but this poses significant challenges in mountainous terrain where meteorological conditions exhibit dramatic spatial variability due to local topographic effects. Forecast skill in the Intermountain West lags other regions in the country due to upstream data void regions, limitations associated with in situ and remotely sensed data and, ultimately, weaknesses in conceptual models of air flow interaction with complex terrain (Schultz et al. 2002).

Smith et al. (1997) identified several challenges associated with forecasting in mountainous areas including the sensitivity of rainfall patterns to errors in the large scale ambient wind, validation challenges presented by nonuniform precipitation patterns, the simulation of cold-pool events in valleys and larger basins, and the collective and multi-

scale effects of complex terrain. Whether topography enhances mesoscale predictability (e.g., Paegle et al. 1990) or exaggerates large scale model errors (e.g., Onton and Steenburgh 2001) to limit the predictability of mesoscale circulations (e.g., Davis et al. 1999) remains, in large part, unknown.

Previous studies have verified quantitative precipitation and, to a lesser extent, other surface weather parameters in complex terrain. In an effort to improve the understanding and prediction of cool-season precipitation processes in complex terrain, the Intermountain Precipitation Experiment (IPEX) was held in the vicinity of Salt Lake City, UT, from 31 January through 26 February 2000. Cheng (2001) documented the spatial structure and evolution of precipitation during six IPEX intensive observing periods (IOPs) and validated quantitative precipitation forecasts from the National Centers for Environmental Prediction (NCEP) Eta model (32-km grid spacing) and University of Utah mesoscale prediction system (36- and 12-km grid spacing), which was based on the fifth-generation Pennsylvania State University–National Center for Atmospheric Research (PSU–NCAR) Mesoscale Model (MM5). Cheng (2001) found that the 12-km MM5 nest produced more skillful forecasts than the Avn and Eta during the IPEX IOPs, but all three models were too coarse to accurately resolve the fine scale of the observed orographic precipitation structures. For example, the MM5 overforecasted precipitation in the Wasatch Mountains with the predicted maxima coincident with the highest model terrain, which was east of the actual crest. The MM5 also overforecasted precipitation amounts in valleys to the lee of the Wasatch Mountains, another apparent effect of inadequate model resolution. In the lower-elevation Wasatch Front there was no significant bias exhibited by the MM5.

Colle et al. (2000) presented a multimonth, high-resolution model verification of precipitation over the Pacific Northwest. In their study, which used a multiply-nested (36-, 12- and 4-km) version of the PSU–NCAR MM5, the 36-km domain produced broad precipitation maxima along the lower windward slopes of the Cascade Mountains with limited leeside precipitation shadowing. The orographic precipitation was more realistic at 12 km with distinct maxima over the southwestward facing windward slopes and more pronounced precipitation shadowing in the larger gaps and valleys. At 4 km, maxima were also produced over narrow ridges and to the lee of volcanic peaks. Nevertheless, while they noted improved quantitative precipitation forecast performance when the grid spacing was decreased from 36 to 12 km, the 4-km nest improved performance only during heavy precipitation events. Mass et al. (2002) found that mesoscale structures were better defined by the 4-km MM5, however, poorer objective verification resulted from space and timing errors. Numerical weather prediction studies over the Pacific Northwest (Colle et al. 1999, 2000; Mass et al. 2002), also reported noticeable improvements in 10-m wind, 2-m temperature and sea level pressure forecast skill as the grid spacing was decreased from 36 to 12 km, but additional decreases to 4 km had a limited effect on traditional verification scores. Overall, these studies suggest that model skill is not improving as grid spacing is decreased below 12 km, even in an area of complex terrain.

These studies, however, were carried out over terrain with broader horizontal scales (Colle et al. 1999, 2000; Mass et al. 2002) than the Intermountain West. The Cascades are characterized by a half width of ~50 km, whereas the mountains of northern Utah and the Intermountain West rise more abruptly and have half widths of ~5–10 km.

Storm and cloud characteristics also vary substantially between the Cascades and

interior ranges. In their study over northwestern Utah, Williams and Heck (1972) found that predictability was limited by the inherently small areal coverage of precipitation events in the mountainous west. More recently, Serreze et al. (2001) reported that large precipitation events are more spatially coherent in marine regions (e.g., Pacific Northwest) compared with drier, inland regions such as Utah. Other differences between the Pacific Northwest and Intermountain West are associated with the microphysical characteristics of clouds. Ice multiplication by rime splintering (Hallett and Mossop 1974), common in coastal regions with warmer cloud bases and larger cloud droplets, tends to be absent over inland regions (Cooper and Saunders 1980; Rauber 1987) where cloud bases are colder and cloud droplets are smaller. Large droplets and efficient glaciation processes over maritime orography lead to a more effective depletion of supercooled liquid water at lower levels than the continental regimes of inland ranges (Marwitz 1986, 1987).

Usually independent of the periods during which precipitation is observed, basin-and-range topography and generally lower humidities make the Intermountain West more susceptible to the formation and persistence of cold pool episodes. In their case study of deep stable layers (DSLs) in the Intermountain western United States, Wolyn and McKee (1989) found that DSLs initiated with the development of a deep layer of cold air over the region followed by rapid synoptic scale warming aloft from subsidence and/or horizontal warm advection. Strong infrared cooling associated with the dry, cloudless regimes common in the interior western United States can lead to fog, thereby increasing the reflectivity of solar radiation and prolonging inversion episodes (Wolyn and McKee 1989). In addition, the surrounding terrain acts to restrict the movement of the cold air near the basin surfaces, allowing cold pools to persist for several days or weeks in winter

(Wolyn and McKee 1989; Doran et al. 2002; Clements et al. 2003).

Due to recent advancements in model resolution, there has been limited opportunity for verification of high-resolution forecasts over the Intermountain West. Hart et al. (2004) found that there was only marginal or no improvement in the prediction of surface temperature and wind as grid spacing was decreased from 12 to 4 km, but this study used a limited number of stations and did not examine precipitation. Using a significantly expanded set of verification sites, the present study is motivated by the need to better understand the impact of model resolution improvements over finescale Intermountain topography, with the goal of identifying where, when, and why such resolution improvements affect model skill.

The remainder of this chapter is organized as follows: A description of the data and analysis methods used for verification are presented in the next section. Then, MM5 temperature, wind speed, wind direction, and precipitation forecast results at 12- and 4-km grid spacing are evaluated. A summary and conclusions are presented in the final section.

## Data and Methods

### Observational Data

The forecast verification was conducted from 23 January to 25 March 2002 which included the Olympic (8–26 February) and Paralympic (7–16 March) Games. Observational data were provided by MesoWest (Horel et al. 2002a), a diverse collection of cooperative networks that includes surface weather observations from more than 180 providers and 6000 stations nationally, with an emphasis on the western United States. MesoWest networks include data collected from a heterogeneous mix of sensors with different siting requirements, sensor heights and reporting intervals. Unlike the more

instantaneous observations from Automated Surface Observing System (ASOS) stations that were designed primarily to support aviation customers, MesoWest observations usually represent averages over the entire reporting interval, which can vary from 5 min to an hour (Horel et al. 2002a). Observations within 15 min of and nearest to the verification times were used for the validation.

The study was performed over the basin-and-range topography of northern Utah including the Great Salt Lake and Stansbury, Oquirrh, and Wasatch Mountains (Fig. 3.1). Nearly 200 stations over northern Utah were used to verify MM5 forecasts for temperature, wind speed, and wind direction (Figs. 3.2a,b) while nearly 100 stations were used to verify MM5 precipitation forecasts including MesoWest stations and observations reported by the National Climatic Data Center (NCDC) cooperative observing network (COOP) (Fig. 3.2c). COOP stations equipped with a Fischer-Porter weighing rain gauge to record hourly precipitation measurements along with stations that reported 24-h precipitation amounts within 1 h of 0000 UTC were used in this study. In addition to COOP stations, precipitation data were provided by several networks with a variety of instrument types including heated tipping buckets, ETI, Belfort, and Geonor weighing gauges, and large weighing gauges from the Natural Resources Conservation Service (NRCS) Snow Telemetry (SNOTEL) network. Precipitation observations from SNOTEL gauges, designed primarily for seasonal measurements, were recorded hourly at a resolution of 2.54 mm (0.10 in.) and run through a three-pass median filter to smooth out anomalous fluctuations. All other precipitation data were evaluated at hourly intervals of 0.254 mm (0.01 in.), even if the observations were available at up to 5-min intervals.

In addition to the variety of sensor types, the siting criteria for each location were

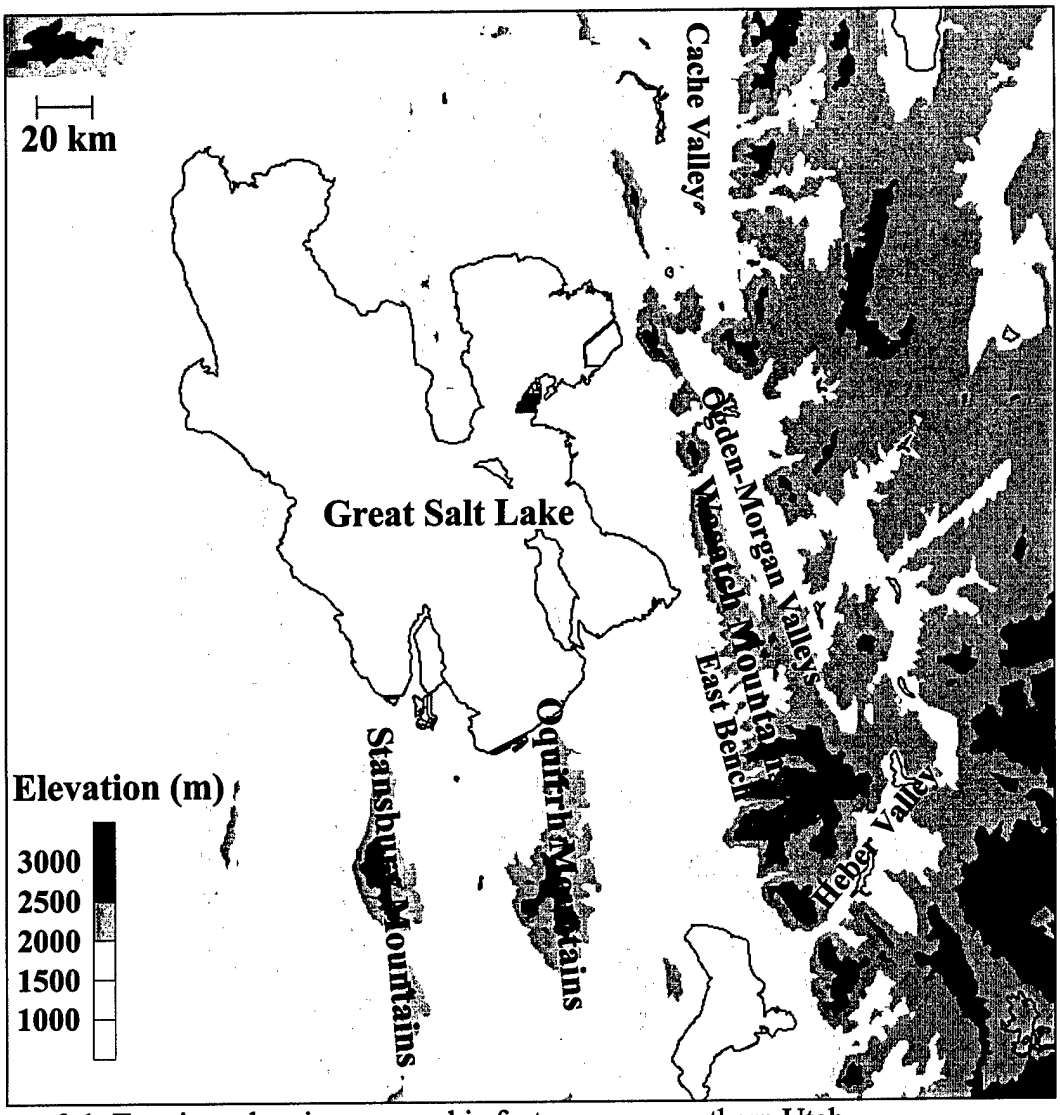



Figure 3.1. Terrain and major geographic features over northern Utah.

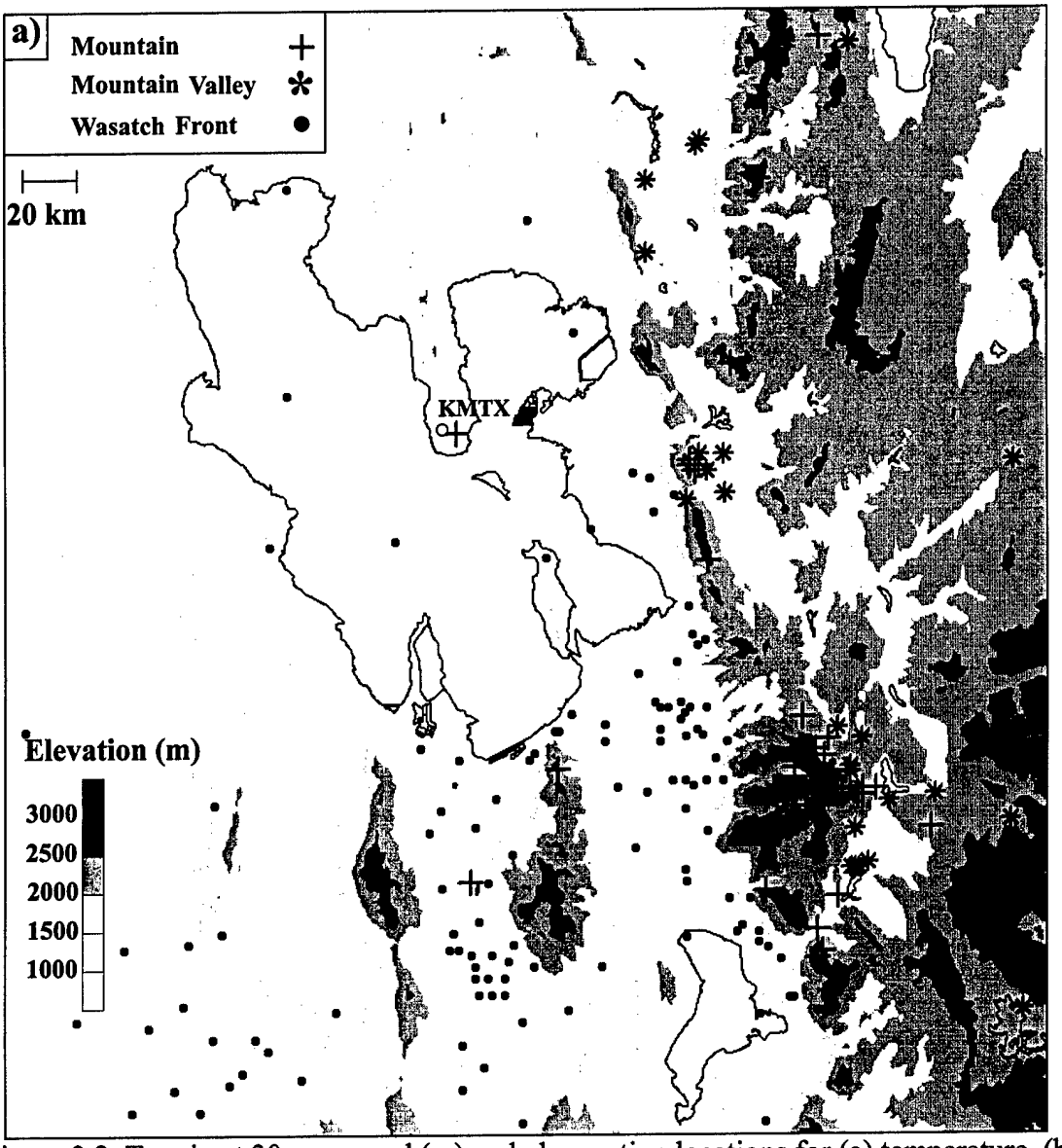



Figure 3.2. Terrain at 30 arcsecond (m) and observation locations for (a) temperature, (b) wind, and (c) precipitation.

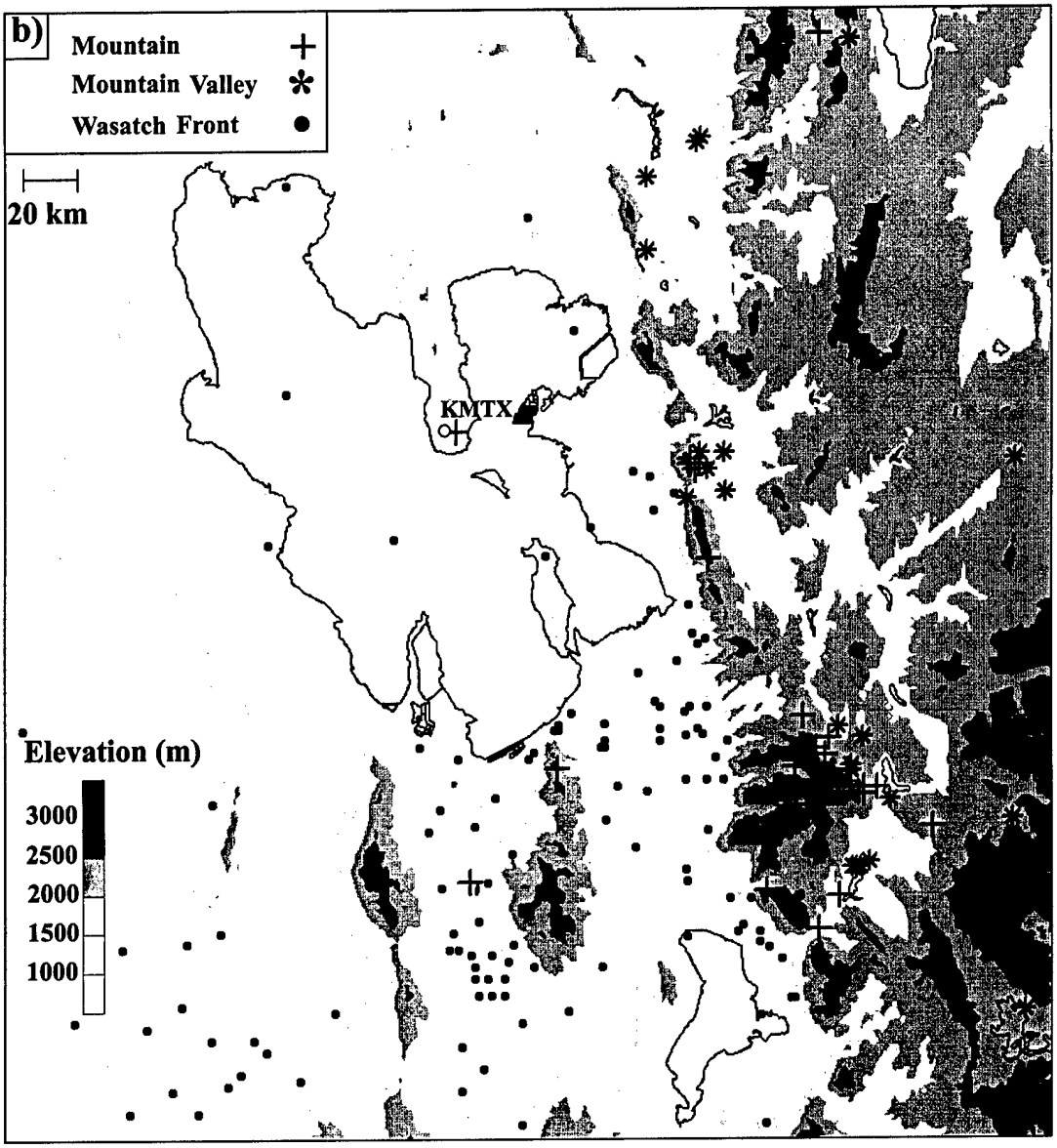



Figure 3.2. (continued)

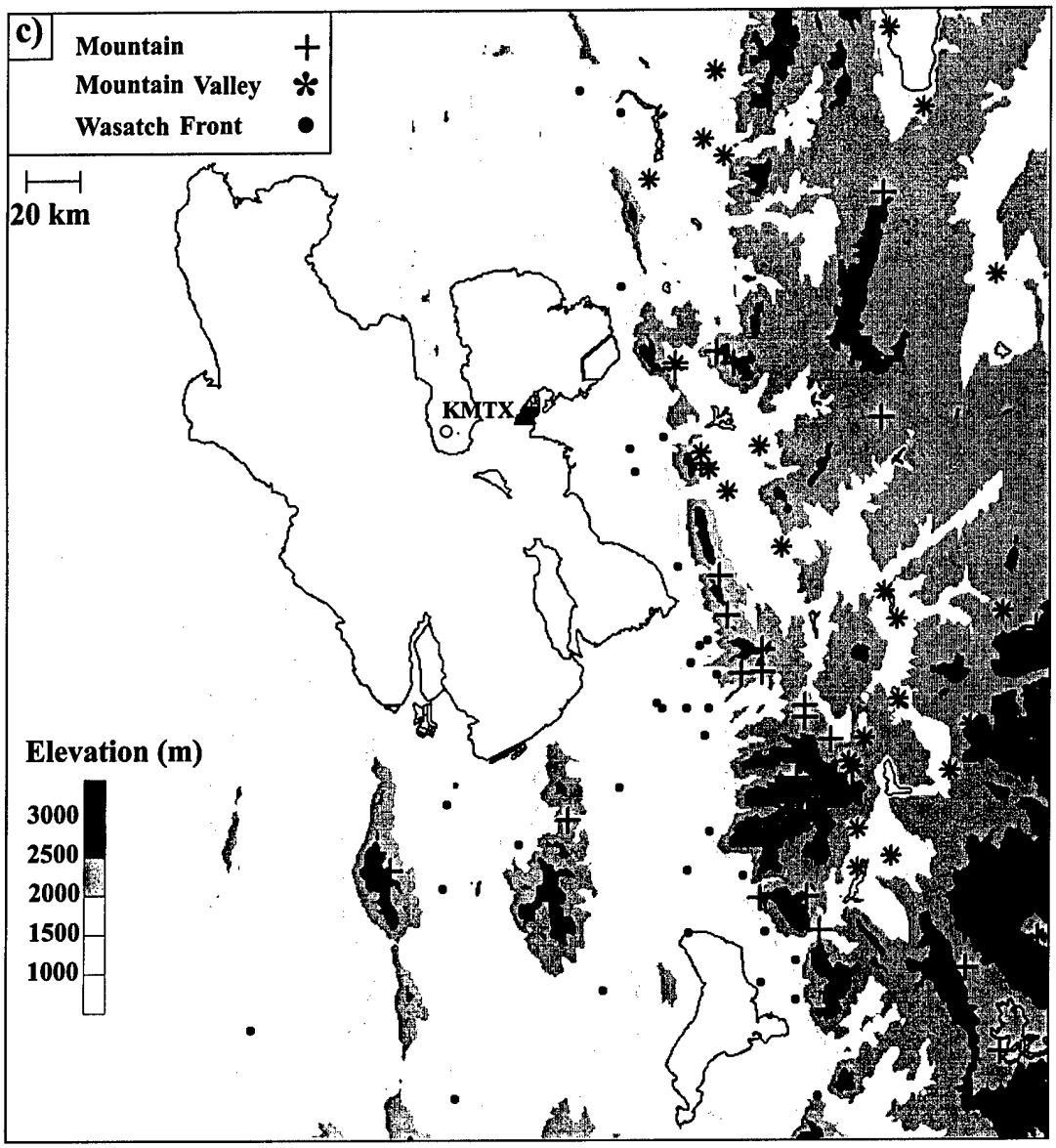



Figure 3.2. (continued)

determined by the owner of each respective network or group. For example, some sites designed to collect precipitation are installed in well-sheltered areas that may lead to underestimates of wind speed. It is assumed, however, that the observational errors are unbiased and uncorrelated with one another at least to the extent that any large biases between the model and observations reflect characteristics of the model more than the verifying data set.

Observation sites were separated into groups to facilitate the verification statistics. Locations were classified as either mountain, mountain valley, or Wasatch Front sites based on terrain characteristics (Fig. 3.2). Mountain sites are high-elevation locations in the Wasatch and neighboring mountains over northern Utah. Mountain valley sites are located in valleys east of the Wasatch Crest. Sites in the lowland region immediately west of the Wasatch Mountains are classified as Wasatch Front sites.

Temperature and wind observations were tested with a gross-range-check algorithm to remove erroneous observations (see Hart et al. 2004 for details). In addition, stations with fewer than 100 forecast/observation pairs were eliminated from the calculations. Precipitation observations were compared to lowest elevation (0.5°) reflectivity images from the Promontory Point WSR-88D (KMTX; see Fig. 3.2 for location) to remove precipitation recorded during nonprecipitating periods. Clearly erroneous precipitation gauge data, stations that did not report consistently for the entire period, and observations that reflected the obvious effects of bridging during frozen precipitation were removed. Factors including evaporative loss from heated tipping buckets and potential underestimation associated with undercatchment (Groisman and Legates, 1994) were not considered.

## Model Description

A real-time, multiply-nested version of the MM5 (Grell et al. 1994) was run four times daily (0000, 0600, 1200, and 1800 UTC) at the University of Utah with grid spacings of 36, 12 and 4 km (Fig. 3.3) (Horel et al. 2002b). The Olympic forecast team, responsible for weather prediction at each outdoor venue and along transportation corridors during the 2002 Olympic and Paralympic Games, had real-time access to the MM5 products. The nesting between the 36- and 12-km domains was two way, while no feedback was allowed from the 4- to 12-km domain. Observed terrain data (10-min resolution for 36-km domain and 30-s resolution for the 12- and 4-km domains) were bilinearly interpolated onto the MM5 grid and filtered with a two-pass smoother/de-smoother to provide model terrain. While more representative than the 36-km domain, the 12- and 4-km domains still do not fully represent the complex terrain over northern Utah. For example, the 12-km terrain depicts the high terrain of the Uinta Mountains in eastern Utah, but smooths out much of the basin-and-range topography of northern Utah including the Stansbury, Oquirrh, and Wasatch Mountains (cf. Figs. 3.1, 3.3b). The 4-km domain begins to resolve the aforementioned ranges, but fails to capture their crest heights and narrow profiles (cf. Figs. 3.1, 3.3c). All MM5 simulations used convective, radiative, boundary layer, and microphysical parameterizations described by Kain and Fritsch (1993), Dudhia (1989), Hong and Pan (1996), and Dudhia (1989), respectively. Initial and lateral boundary conditions were provided by the NCEP Eta Model initial analysis, with the ARPS Data Analysis System (ADAS) used to incorporate MesoWest surface observations into the near-surface analysis (Horel et al. 2002a; Lazarus et al. 2002). The

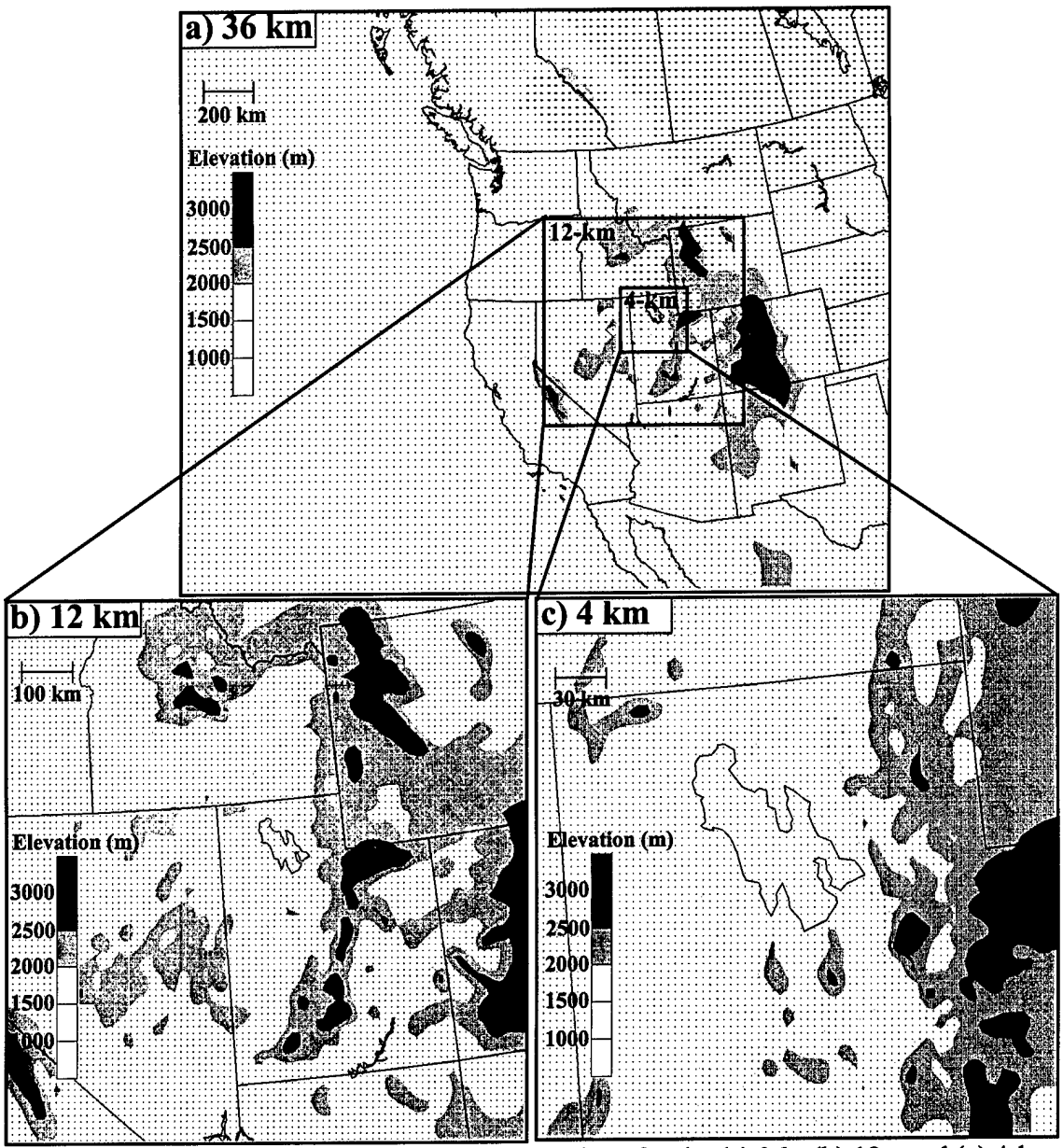



Figure 3.3. Model terrain (m) and grid point locations for the (a) 36-, (b) 12-, and (c) 4-km domains of the MM5.

verification concentrated on northern Utah (roughly encompassed by the 4-km domain), a region that included all of the outdoor venues for the 2002 Olympic and Paralympic Games.

## Verification Methods

Raw model temperatures and wind speeds from the MM5 12- and 4-km domains initialized at 0000, 0600, 1200, and 1800 UTC were verified based on a bilinear interpolation of lowest half-sigma level (~40 m) gridpoint forecasts to the observation locations. Results are presented for the lowest half-sigma level temperature (MM5 RAW) and an adjusted temperature (MM5 ADJ) that uses a simple approach to correct for differences between the model and actual topography and the lowest half-sigma level (40 m) and the observing sensor height (2 m). Specifically, following Mass et al. (2002), temperatures were adjusted to 40 m above the actual elevation using a *U.S. Standard Atmosphere, 1976* lapse rate of 6.5°C $km^{-1}$, then averaged with the model ground temperature to calculate MM5 ADJ temperatures. Similarly, 40-m wind speeds were reduced to 10 m using a logarithmic wind profile:

$$U(10m) = U(40m)\left(\frac{\ln\left(\frac{10m}{z_o}\right)}{\ln\frac{(40m)}{z_o}}\right), \tag{3.1}$$

where U represents wind speed and $z_0$ represents the roughness length depending on land use at the observation site (Mass et al. 2002; Colle et al. 2001). Roughness lengths for each observation location were calculated using an average of the roughness lengths for the surrounding MM5 grid points.

Bias error (BE) and mean absolute error (MAE) were used to assess temperature,

wind speed, and wind direction forecasts. The BE, associated with the model tendency to overforecast or underforecast the observed quantity (Wilks 1995), is given by

$$BE = \frac{1}{N}\sum_{i=1}^{N}(f_i - o_i), \tag{3.2}$$

where $N$ is the number of forecast/observation pairs in the sample, $f_i$ is the forecast, and $o_i$ is the observation. The MAE represents the average forecast error magnitude:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|f_i - o_i|. \tag{3.3}$$

MM5 precipitation forecasts were also bilinearly interpolated to each observation location. Precipitation verification scores were derived using a 2 x 2 contingency table (Fig. 3.4) approach (Wilks 1995), where each matrix element is incremented by one based on whether the observation and/or forecast reached or exceeded a given threshold during a specified forecast period (e.g., 6 or 24 h). While all model initialization times (0000, 0600, 1200, and 1800 UTC) were used for the 6-h precipitation statistics, only the 0000 UTC (0–24 h forecasts) and 1200 UTC (12–36 h forecasts) initializations were used for 24-h precipitation statistics. Bias scores, equitable threat (ET) scores, false alarm ratios (FAR), and probabilities of detection (POD) were calculated based on the contingency table. Bias score, the ratio of forecast to observed events, is defined as

$$Bias = \frac{A+B}{A+C}, \tag{3.4}$$

where A, B, and C are defined in Fig. 3.4. Overforecasting (underforecasting) results in bias scores that are greater than (less than) one with B = 1 for unbiased forecasts. The ET score

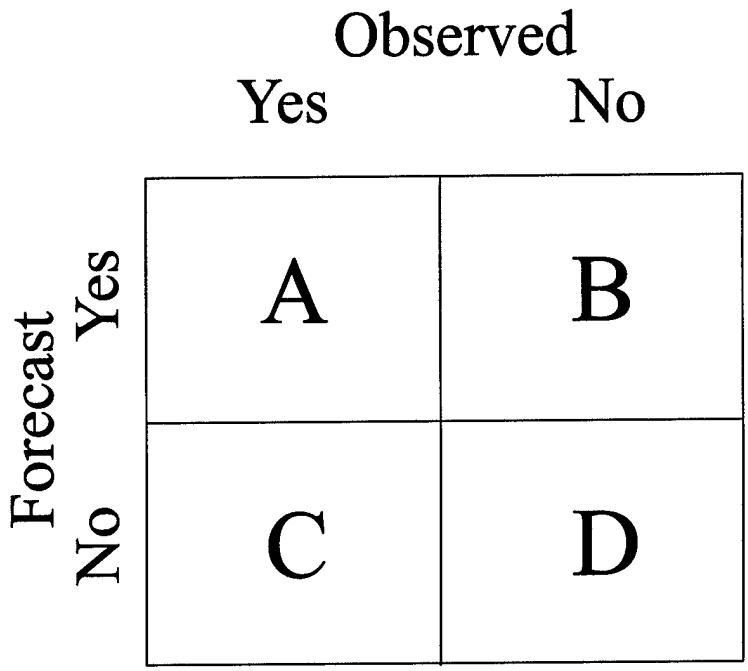



Figure 3.4. A 2 X 2 contingency table where A, B, C, and D represent the number of occurrences in which the observations and model meet or exceed a given precipitation threshold value (e.g., 0.254 mm, 2.54 mm,...).

represents the fraction of correct "yes" forecasts, or "hits", to the total number of occurrences in which the event was forecast and/or observed and is given by

$$ET = \frac{A-E}{A+B+C-E}, \tag{3.5}$$

where E represents the correction for expected forecast hits that occur by chance:

$$E = \frac{(A+B)(A+C)}{A+B+C+D}. \tag{3.6}$$

FAR, the fraction of forecasts that do not verify, is defined by

$$FAR = \frac{B}{A+B}. \tag{3.7}$$

POD, the fraction of observed events that were correctly forecast, is given by

$$POD = \frac{A}{A+C}. \tag{3.8}$$

Precipitation forecasts were also examined using the root-mean-square error, RMSE, where

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (f_i - o_i)^2}{N}}, \tag{3.9}$$

$f_i$ is the forecast precipitation, $o_i$ is the observed precipitation and N is the number of observations that reached a specified threshold.

The percentage bias score, $B_{station}$, for a given threshold was calculated by summing all of the forecast precipitation and dividing by the observed precipitation at each station where

$$Bstation = \frac{\sum_{i=1}^{N} f_i}{\sum_{i=1}^{N} o_i} x100, \tag{3.10}$$

$f_i$ is the forecast precipitation, $o_i$ is the observed precipitation and N is the number of observations reaching a specified threshold at that station.

## Results

### Temperature Forecasts

For all stations, differences between the 12- and 4-km MM5 RAW temperature MAEs were small with values of 3.4°C and 3.3°C, respectively (Fig. 3.5a). Although the MAE for the 4-km domain was somewhat smaller than that of the 12-km domain, the student t-test value (Panofsky and Brier 1958) for the difference between the means was small,[1] suggesting that *the null hypothesis (i.e., 12-km and 4-km domain forecasts are of equal skill) can not be rejected at the 95% confidence level.* For both domains, a significant portion of the RAW temperature MAE magnitude could be attributed to an overall warm bias (Fig. 3.5b). Adjusting for model-actual terrain and the 2-m sensor height (MM5 ADJ) produced only a slight reduction in temperature MAE (Fig. 3.5a) and virtually no change in bias (Fig. 3.5b). Therefore, simple corrections to account for the inaccurate model representation of the local topography and surface layer failed to make substantial forecast improvements.

---

[1] Student t-test sample sizes were based on the number of stations (e.g., 195 stations for temperature).

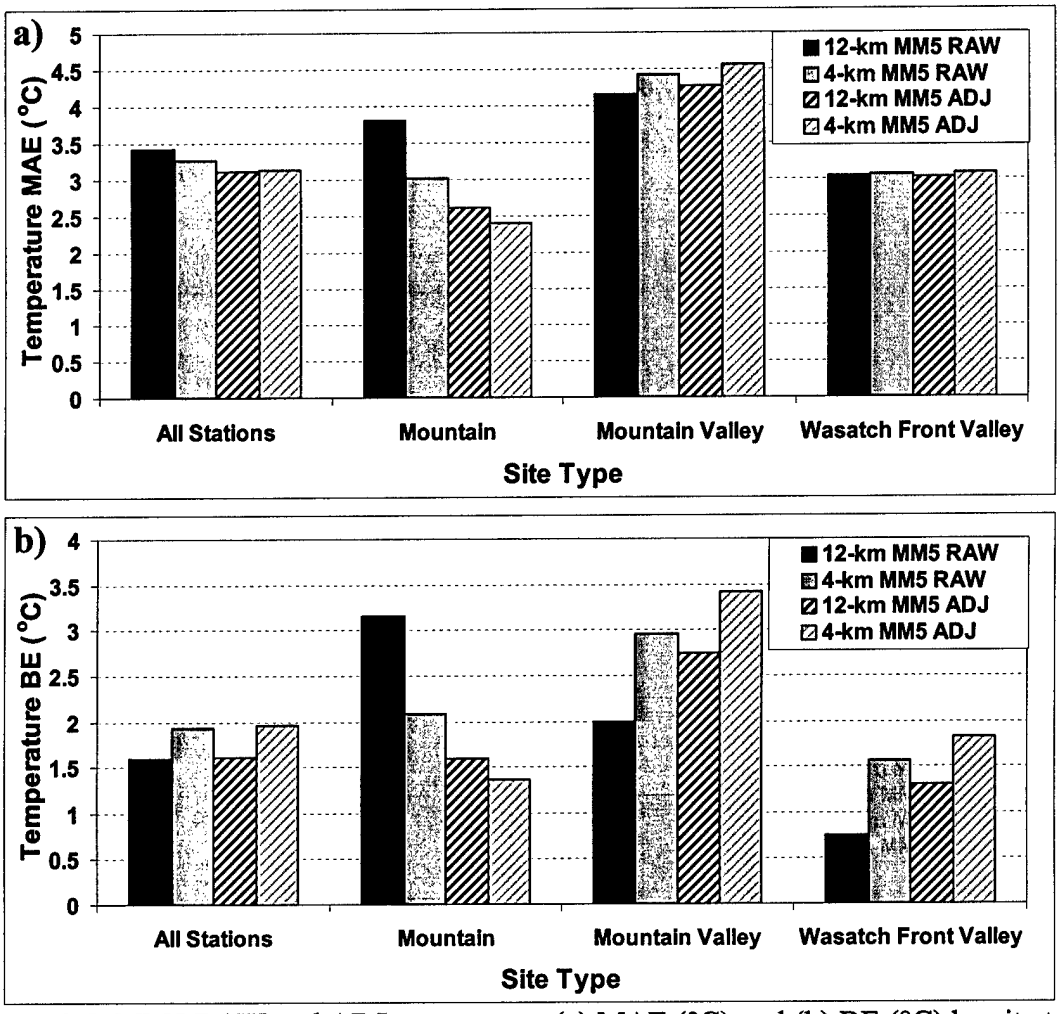



Figure 3.5. MM5 RAW and ADJ temperature (a) MAE (°C) and (b) BE (°C) by site type for the 12- and 4-km domains. MAE and BE are averaged for all model runs and forecast hours (0–36).

A closer inspection of the temperature MAEs shows that differences between the 12- and 4-km domain were more substantial at the mountain locations, where the RAW temperature MAE for the 4-km domain was much smaller than that of the 12-km domain (3.0° and 3.8°C, respectively, Fig. 3.5a). Although the MAE difference was smaller for MM5 ADJ, the 4-km domain still produced a better forecast in the mountains. Bias errors were also substantially lower for the 4-km domain (Fig. 3.5b). These results illustrate that *finer resolution and better terrain representation did improve surface temperature forecasts over the narrow, steeply sloped Wasatch Mountains.*

In contrast, at mountain valley and Wasatch Front locations, 4-km RAW and ADJ MAEs and BEs were equal to or higher than those of the 12-km domain. In addition, although one would expect that performing topographic and sensor height corrections (MM5 ADJ) would improve the forecast, at mountain valley and Wasatch Front sites this resulted in either no improvement or a degraded forecast, consistent with the findings of Hart et al. (2004). Therefore, finer resolution and terrain representation did not improve the MAE of temperature forecasts over the low-elevation Wasatch Front and within mountain valleys. In the latter case, the forecast was actually less accurate. For simplification, the balance of this section will focus on a comparison of MM5 RAW temperature results; however, similar results were found for MM5 ADJ temperatures (not shown).

Why does improved resolution not increase the forecast skill at Wasatch Front and mountain valley locations? Detailed analysis of the Olympic forecasts suggests that it is related to an inability of this version of the MM5 to accurately simulate the strength and evolution of persistent or nocturnal cold pools, which are most pronounced within low-

elevation basins and mountain valleys (Wolyn and McKee 1989; Whiteman et al. 2001). An excellent example of this inability was provided by the persistent cold pools that were observed from 28 Jan – 16 Feb (a cold-frontal passage on 8 Feb provided a brief period of mixing). The Salt Lake City sounding from 1200 UTC 4 February 2002 is typical of this period and features a stable layer that extends from the surface to 800 hPa (roughly mid-mountain) with temperatures well below freezing (Fig. 3.6). During this period, overall MM5 RAW temperature MAEs (Fig. 3.7a) were unusually large and the model warm bias was exaggerated (Fig. 3.7b). The largest temperature BEs were found at lower-elevation mountain valley and Wasatch Front locations (Fig. 3.7c) where the cold pools were entrenched. In contrast, the BEs at mountain sites, most of which were above the cold pool, remained consistent with values found during the entire 2-month verification period.

At the Wasatch Front and mountain valley sites, where large bias errors were observed, the MAE (not shown) of the 4-km domain was larger than that of the 12-km domain, whereas at the mountain sites that were above the cold pool the 4-km MAE was smaller. The differences between the 12- and 4-km domain temperature MAEs are shown in Fig. 3.8 to highlight the influence of cold pools on model performance. For example, when conditions were generally less stable (e.g., after 16 Feb), the 4-km often performed better than the 12-km domain for all site categories (Fig. 3.8a). Improvements by the 4-km over the 12-km domain were even more apparent if considering only late afternoon (e.g., 2300 UTC) observations when boundary layer lapse rates were, on average, less stable (Fig. 3.8b).

Similar model performance characteristics were observed at night when either persistent cold pools are strongest, or nocturnal cold pools develop. For example, MM5

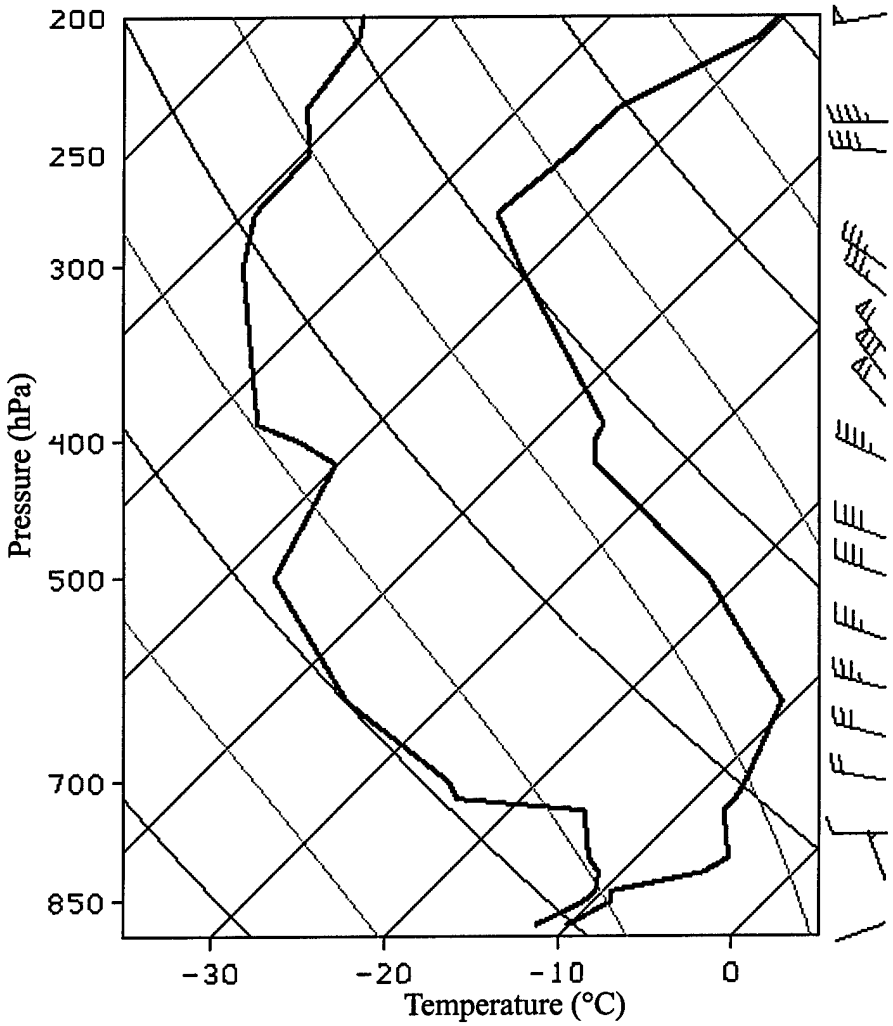



Figure 3.6. Skew *T*-log*p* diagram (temperature, dewpoint, and wind profiles) at Salt Lake City (SLC) taken at 1200 UTC 4 Feb 2002. Full and half barbs denote 5 and 2.5 knots, respectively.

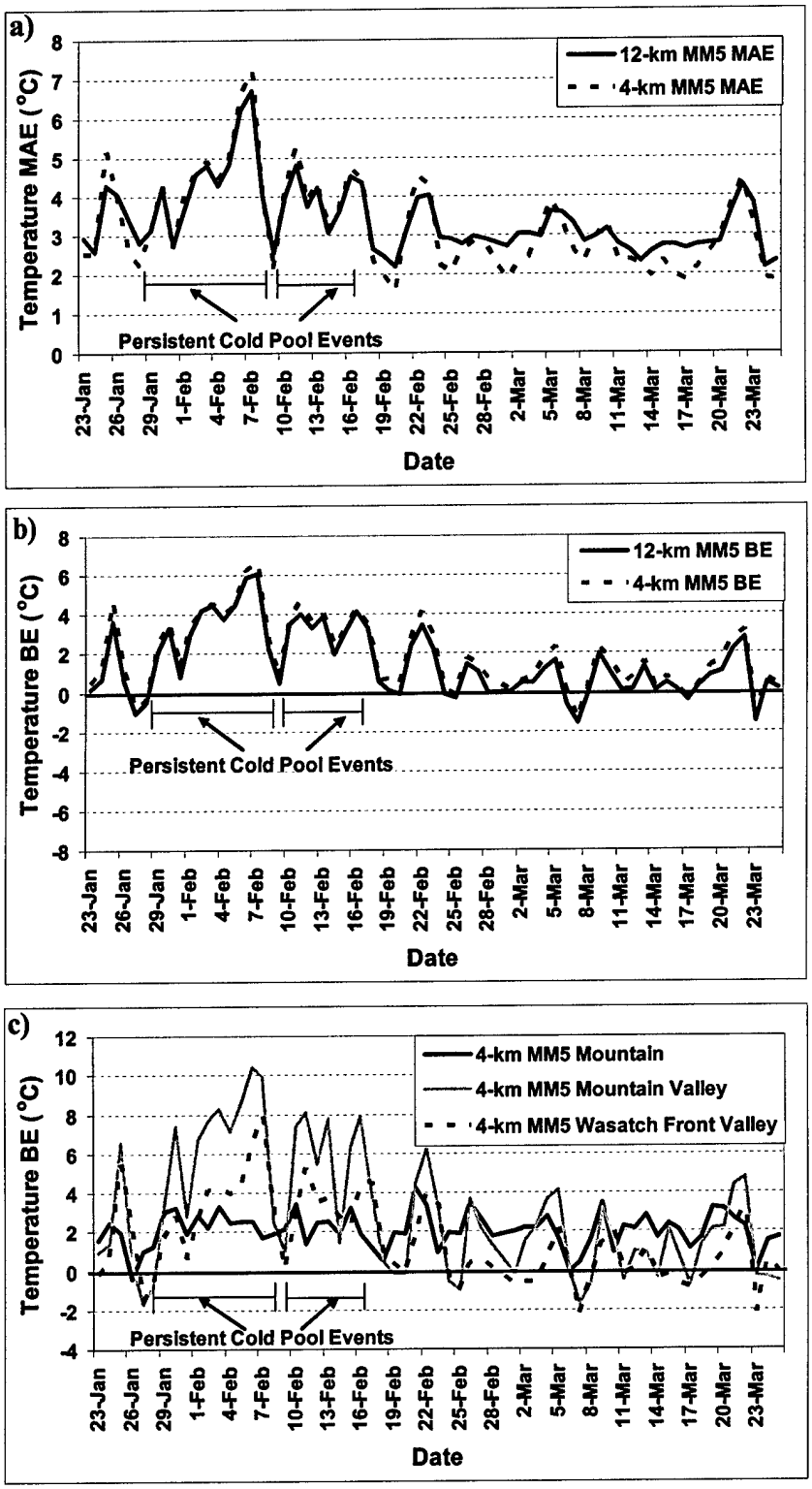



Figure 3.7. MM5 RAW temperature (a) MAE (°C) and (b) BE (°C) by day for the 12- and 4-km domains. (c) 4-km domain RAW temperature BE (°C) for mountain, mountain valley and Wasatch Front Valley sites. MAEs and BEs include all model runs (00, 06, 12, and 18 UTC) and forecast hours (0–36).

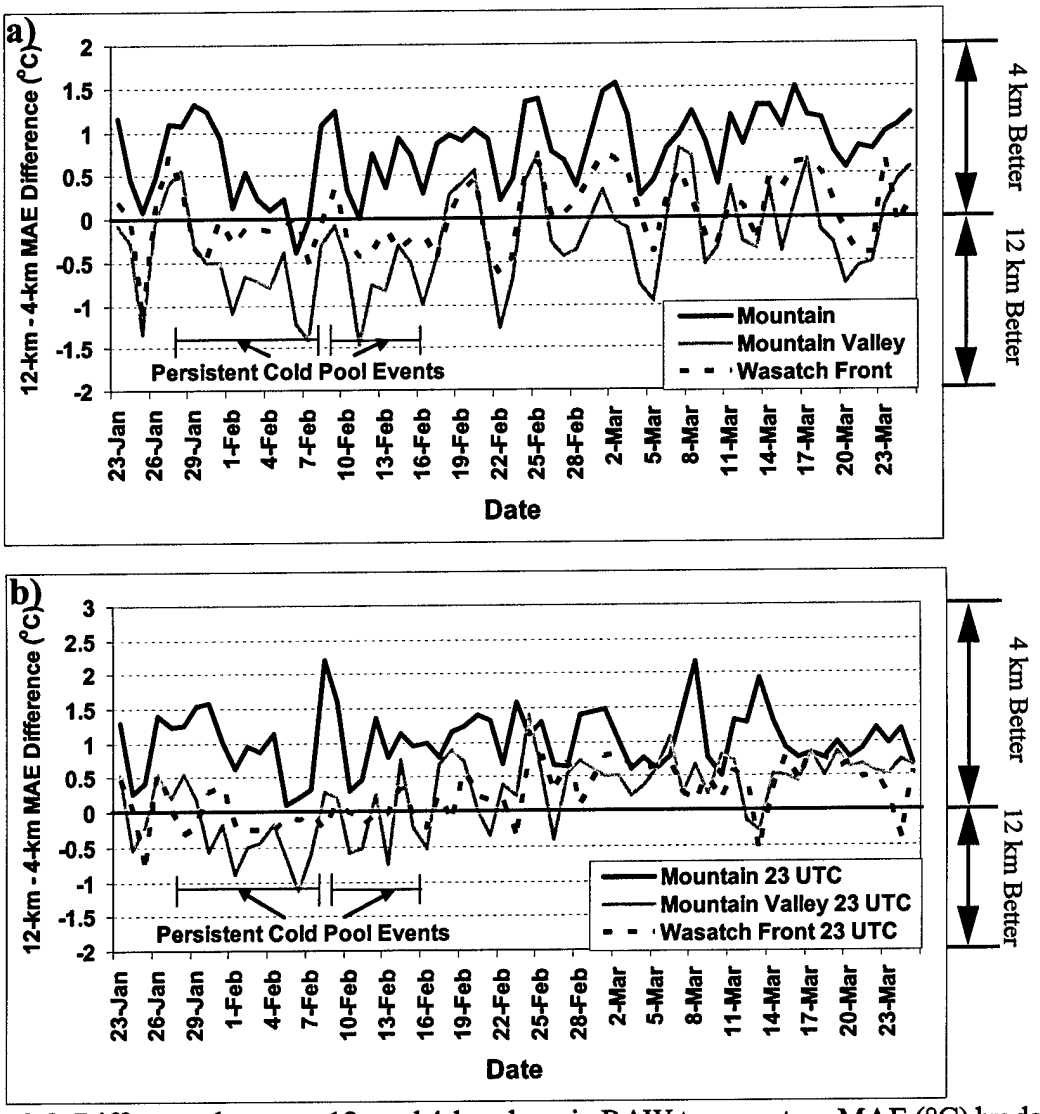



Figure 3.8. Difference between 12- and 4-km domain RAW temperature MAE (°C) by day for mountain, mountain valley and Wasatch Front locations averaged for (a) all hours and (b) only 23 UTC. MAE differences include all model runs (00, 06, 12, and 18 UTC) and forecast hours (0–36).

RAW temperature MAEs for the 4-km domain were larger than those of the 12-km domain overnight and during the early morning hours (Fig. 3.9a). In contrast, during the afternoon when the atmosphere may be well mixed, or the stability during a cold pool episode is typically weakest, the 4-km MM5 RAW MAEs were smaller than those of the 12-km. Most of the nighttime and early morning hour MM5 RAW temperature errors were associated with large warm biases of approximately 3.0°C and 2.5°C for the 4- and 12-km domains, respectively (Fig. 3.9b). These results illustrate that the finer resolution and better terrain representation of the 4-km domain improves model performance in Wasatch Front and mountain valley locations during periods that are well mixed or feature weak static stability. During persistent or nocturnal cold pools, however, when the static stability is high, 4-km MAEs are larger than those of the 12-km domain.

If the 4-km domain produces a better forecast than the 12-km domain during well-mixed conditions, why does it degrade the forecast during periods of strong static stability? This appears to be related to the topography of the two domains. For lower elevation sites (< 2000 m), the elevation of the 4-km domain is on average 100 m lower (and typically closer to the actual elevation) than the 12-km domain (Fig. 3.10). During well-mixed conditions, this results in temperatures that in the 4-km domain are ~1°C warmer and closer to those observed since the 4-km terrain more closely matches the actual terrain. In contrast, during nocturnal or persistent cold pools, the *observed* temperature over Wasatch Front and mountain valley locations typically *increases* with height. In both domains, the MM5 was unable to fully capture the intensity of these cold pools and, as a result, the model soundings, while stable, typically featured temperatures that *decreased* with height. Ironically, the higher elevation model terrain represented an

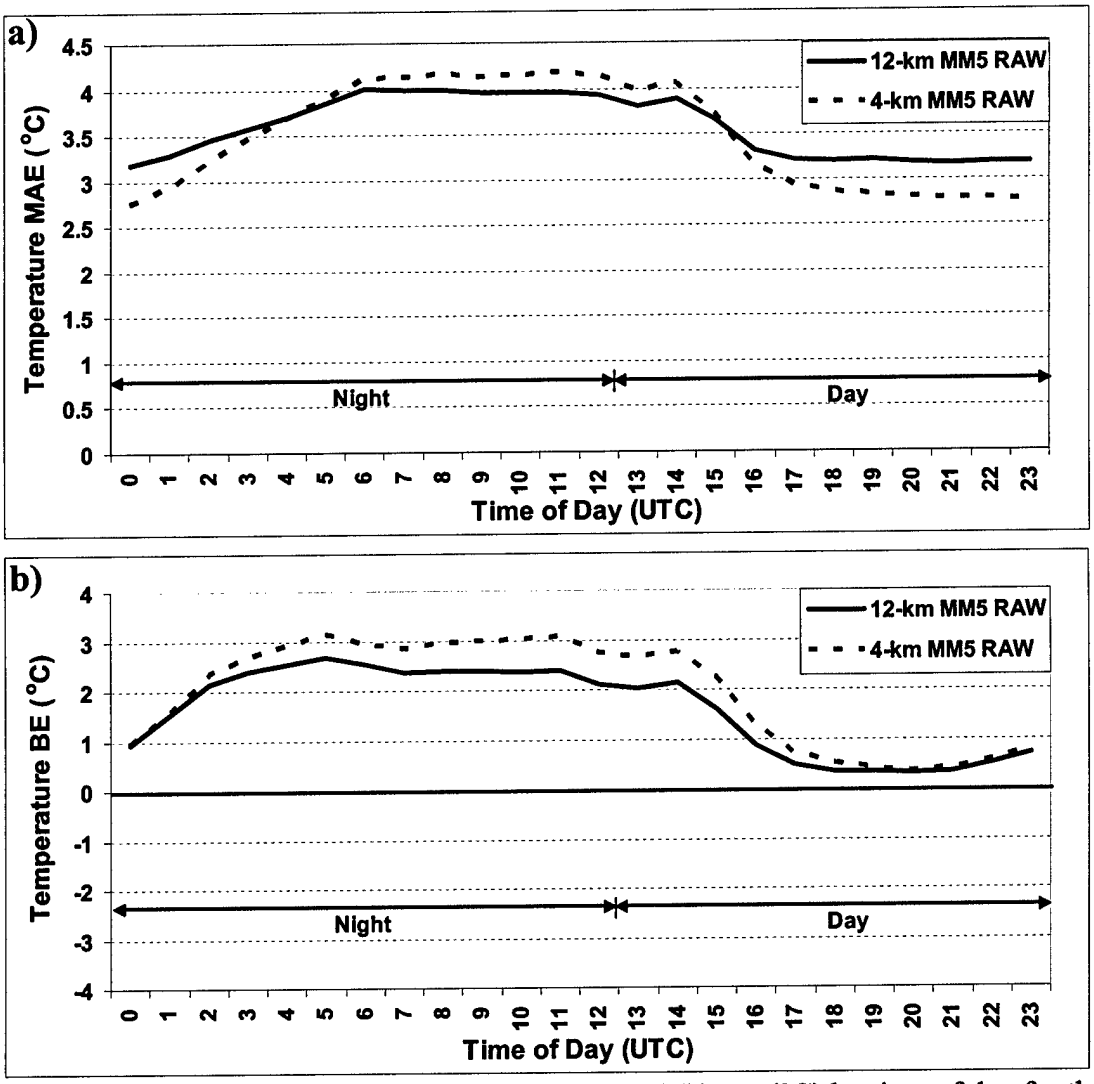



Figure 3.9. MM5 RAW temperature (a) MAE (°C) and (b) BE (°C) by time of day for the 12- and 4-km domains. MAE and BE are averaged for all model runs and forecast hours 12–36.

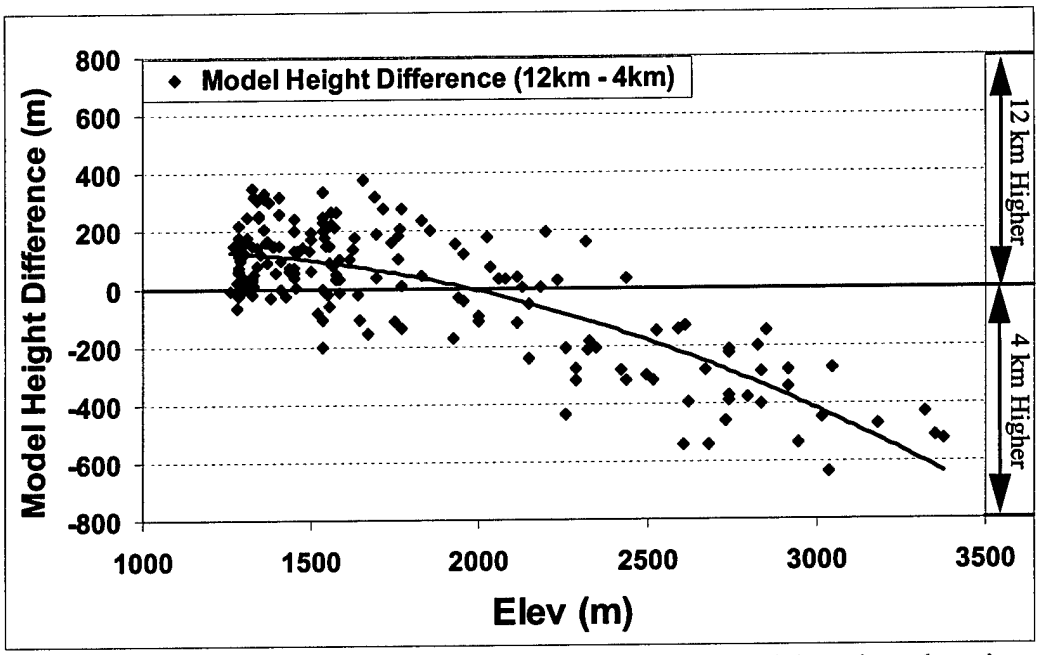



Figure 3.10. Difference between the 12- and 4-km domain model station elevation (m). Line represents a second-order polynomial fit.

advantage for the 12-km domain since it produced temperatures that were cooler (and hence closer to observed) than those of the 4-km domain. Thus, *the improved topographic representation in the 4-km domain, along with the inability of this version of the MM5 to capture the intensity of persistent or nocturnal cold pools, resulted in less accurate surface temperature forecasts.*

Figure 3.11 summarizes the impact of decreased grid spacing on model skill as a function of site elevation. For high elevation sites (> 2000 m), the 4-km domain produced dramatically improved temperature forecasts. These forecast improvements were related to better representation of the narrow high-elevation Wasatch Mountains and the fact that upper-elevation sites are typically above the nocturnal or persistent cold pools that are found in mountain valleys and Wasatch Front locations. In contrast, the 4-km domain produced similar or even somewhat poorer forecasts than the 12-km domain at low-elevation sites (< 2000 m). At these locations, the improved topographic representation of the 4-km domain combined with the model's inability to simulate the intensity of nocturnal or persistent cold pools degraded the overall surface temperature forecast performance of the high resolution nest. These results suggest that improved initialization and boundary layer parameterization is needed in order for the benefit of high-resolution nesting to improve forecasts over the Intermountain West. High-resolution nesting might also provide greater forecast improvements during the warm season when persistent cold pools are rare and well-mixed conditions are more common.

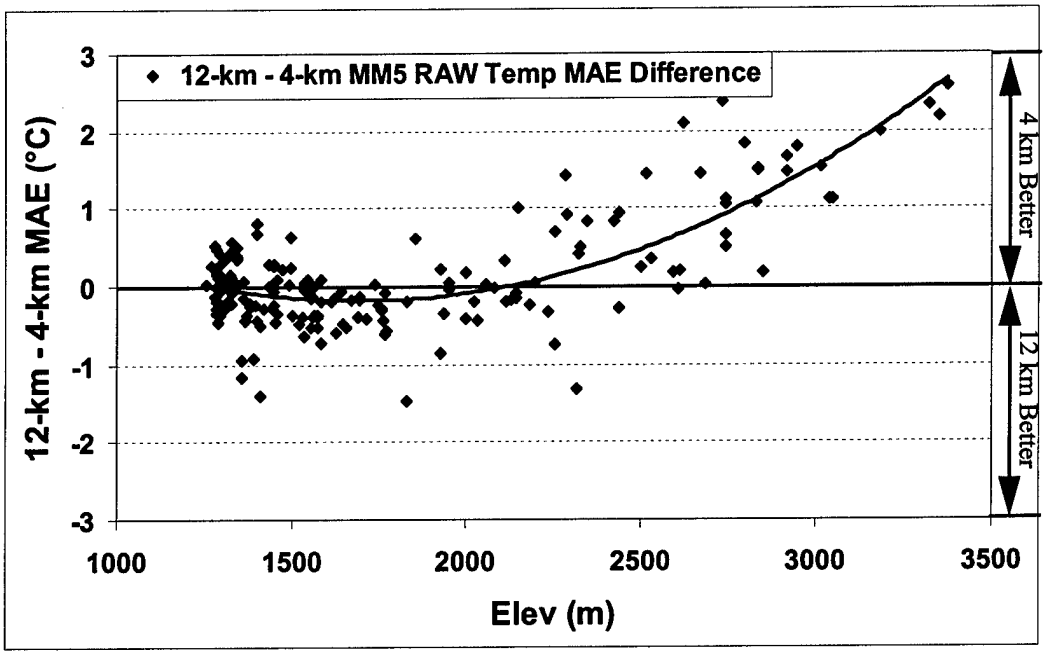



Figure 3.11. Difference between the 12- and 4-km domain RAW temperature MAE (°C) at each station vs elevation (m). Line represents a second-order polynomial fit.

## Wind Forecasts

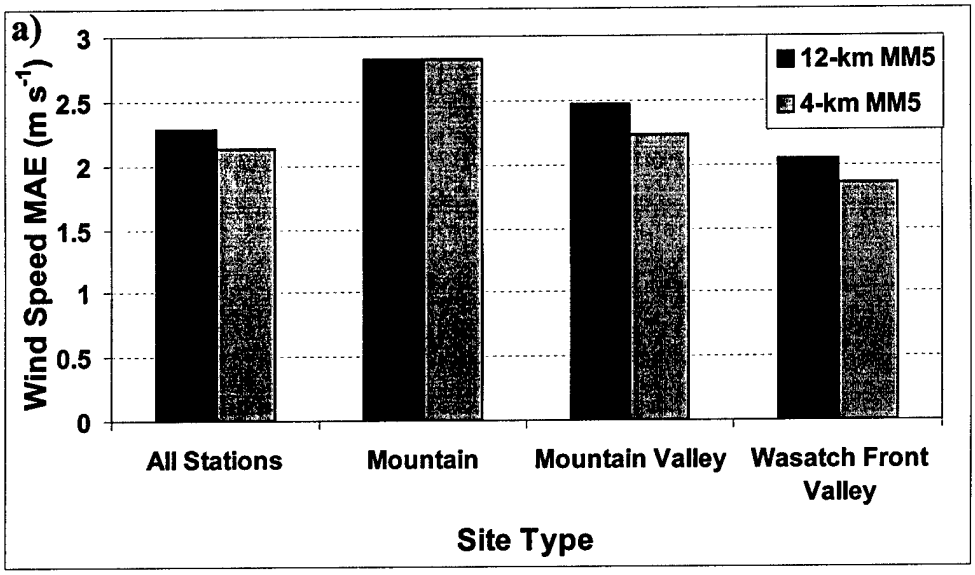
Wind speed MAEs and BEs were also compared for the 12- and 4-km MM5 domains (Figs. 3.12a,b). For all stations, when the horizontal resolution was decreased from 12 to 4 km, 10-m wind speed forecast MAEs improved from 2.3 to 2.1 m $s^{-1}$. Based on a student t-test evaluation, the null hypothesis could be ruled out at the 5% level suggesting that the two means were statistically different at the 95% confidence level. Thus, *measurable improvements in wind speed forecasts were produced by the use of the 4-km nest.* By site type, the 4-km MM5 produced lower MAEs and BEs at the lower-elevation locations, classified as mountain valley or Wasatch Front valley sites.

Before discussing wind speed errors as a function of time of day, it is worth examining the diurnal evolution of observed mean wind speeds by site type. For valley locations,[2] wind speed decreases during the evening transition period following sunset (0000–0200 UTC) (Fig. 3.13a). Average wind speeds remain relatively low until the morning transition period (1600–1800 UTC) when wind speeds increase as the convective boundary layer grows and entrains higher momentum air from aloft. In contrast, the diurnal change is much weaker at mountain locations where winds are slightly stronger at night and weaker during the day.

During the night, model forecasts by the 12- and 4-km domains overestimated 10-m wind speeds at valley locations (Figs. 3.13b,c). Much of the nighttime wind speed forecast errors of approximately 2.4 and 2.1 m $s^{-1}$ at valley sites for the 12- and 4-km domains, respectively, were associated with overprediction biases of 1.5 and 1.0 m $s^{-1}$.

---

[2] Here, valley locations represent both Wasatch Front and mountain valley sites.



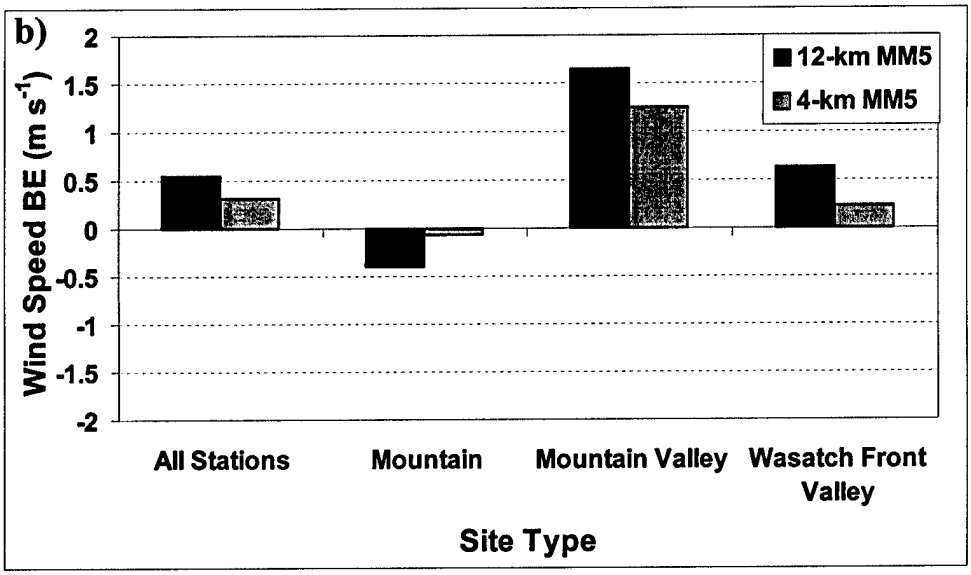




Figure 3.12. Wind speed (a) MAE (°C) and (b) BE (°C) by site type for the 12- and 4-km domains. MAE and BE are averaged for all model runs and forecast hours (0–36).

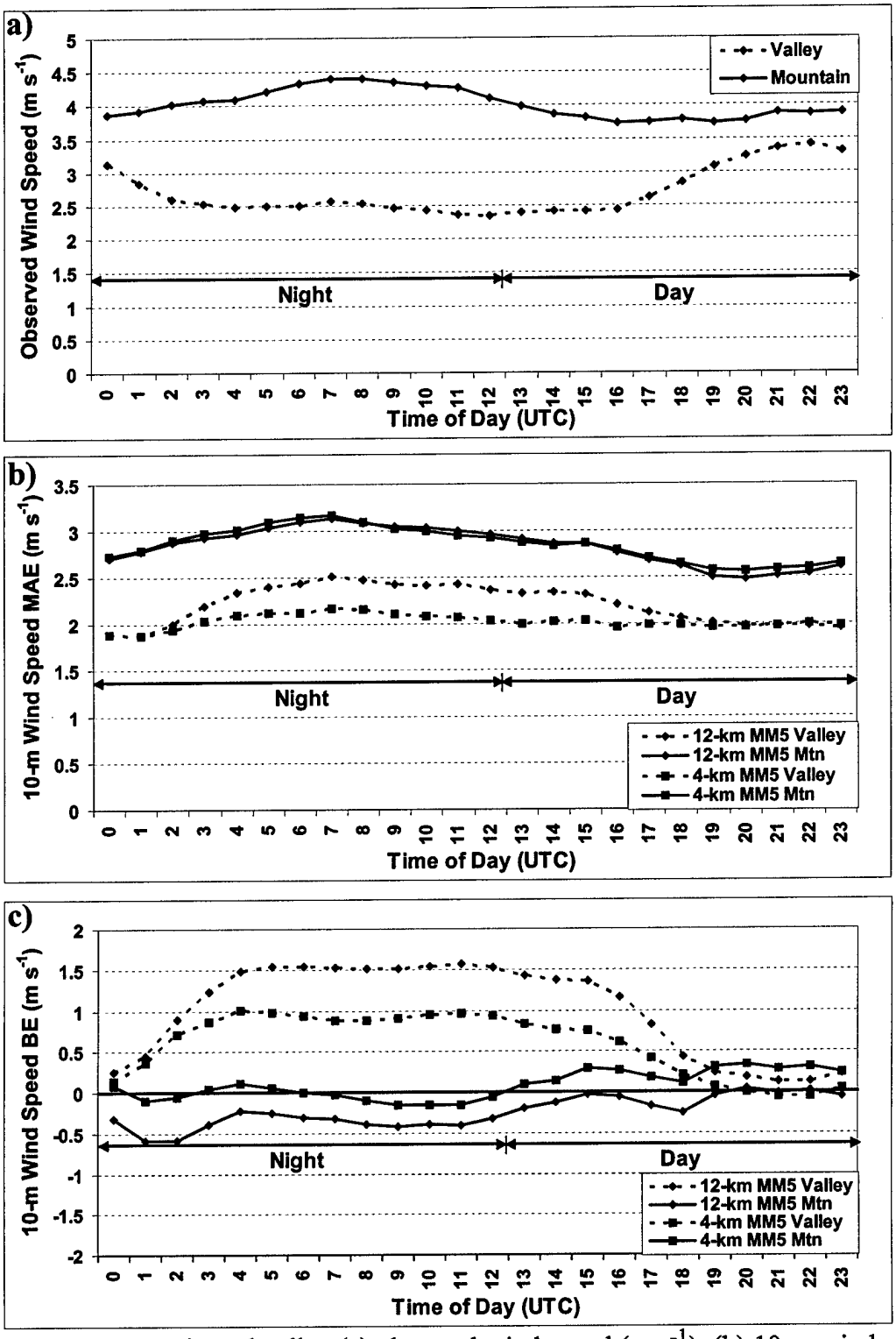



Figure 3.13. Mountain and valley (a) observed wind speed (m $s^{-1}$), (b) 10-m wind speed MAE (m $s^{-1}$), and (c) 10-m wind speed BE (m $s^{-1}$) by time of day. MAE and BE are averaged for all model runs and 12–36-h forecasts.

The overprediction biases could be the result of excessive momentum mixing, possibly related to the inability to fully develop the nocturnal inversion as noted in the prior section, or a lack of observation representativeness during light wind conditions at night. Wind speed MAEs and BEs for the 12- and 4-km domains were smaller during the afternoon hours when the boundary layer environment was less stable. Improvements in MAE and BE by the 4-km domain were largest during the overnight hours, but were more limited during the day (Figs. 3.13b,c). In contrast, differences in MAE between the 12- and 4-km domains were small at all hours at mountain locations. The limited improvements during the day may be related to the stronger momentum mixing and influence of winds aloft, which typically reduces the day-to-day wind consistency (Whiteman 1990; Stewart et al. 2002). Less consistent and more variable wind speeds during the day may result in lower predictability compared to at night.

Wind direction MAEs for all stations (Fig. 3.14a) showed that the 4- and 12-km domains featured MAEs of 50 and 55 degrees, respectively, a difference that was statistically significant at the 95% confidence level. The difference was largest at mountain sites where wind direction MAEs for the 4-km domain were 23% lower than those of the 12-km domain. Similar to wind speed forecasts, most of the skill added by the 4-km domain wind direction forecasts occurred during the nighttime hours (Fig. 3.14b) when there was a reduction of wind direction MAE at valley sites (mountain valley and Wasatch Front sites combined) of about 10% from the 12- to 4-km domains. Furthermore, wind direction MAEs at the mountain sites during the nighttime hours were lowered from about 62 to 43 degrees, a reduction of over 30%. At higher elevations, improved terrain representation of the 4-km domain may have increased exposure to free atmosphere winds

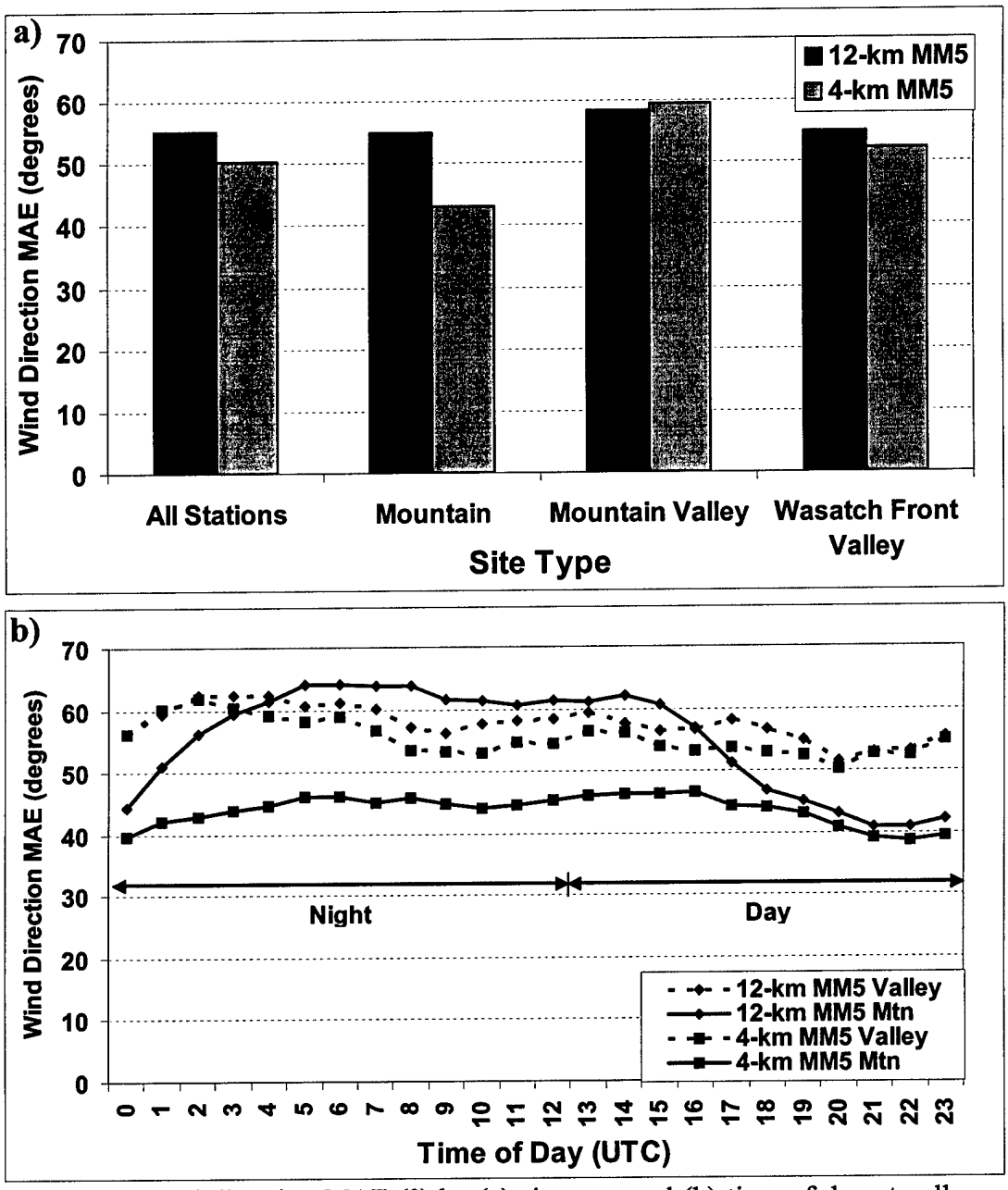



Figure 3.14. Wind direction MAE (°) by (a) site type and (b) time of day at valley and mountain sites for the 12- and 4-km domains. MAE by site type averaged for all model runs and forecast hours (0–36). MAE by time of day averaged for all model runs and 12–36-h forecasts.

and mitigated wind direction errors associated with the lower average model heights of the 12-km domain.

In summary, both wind speed and direction MAEs were lower for the 4-km domain, illustrating one advantage of continuing to decrease grid spacing over the fine scale intermountain topography. Detailed analysis of the MAE as a function of site type and time of day showed that the improvement for wind speed was produced mainly by the improved simulation of nocturnal flows in the lowland regions by the 4-km grid. For wind direction, the most substantial gains were at mountain locations at night, which were likely a reflection of the fact that the 4-km grid is beginning to resolve the mountain ridges and better simulates their exposure to free atmosphere winds above the stable nocturnal boundary layer.

## Precipitation Forecasts

During the study period, most of the precipitation recorded over northern Utah occurred during nine events, although some events were largely restricted to the high-elevation mountain locations. Precipitation totals were largest at the mountain locations (Fig. 3.15), and there were more events toward the end of the period (e.g., after 5 March).

Bias scores, ET scores, FARs, and PODs were calculated for 24-h forecasts of the 0000 UTC (0–24-h forecast period) and 1200 UTC (12–36-h forecast period) initialized model runs (Fig. 3.16). It should be noted that the effects of undercatchment have not been considered. As a result, a "perfect" forecast would likely be associated with bias scores higher than 1.0, however, similar to Colle et al. (2000), bias scores of 1.6 (0.8) still likely reflect overforecasting (underforecasting) by the model of true precipitation amounts. Bias

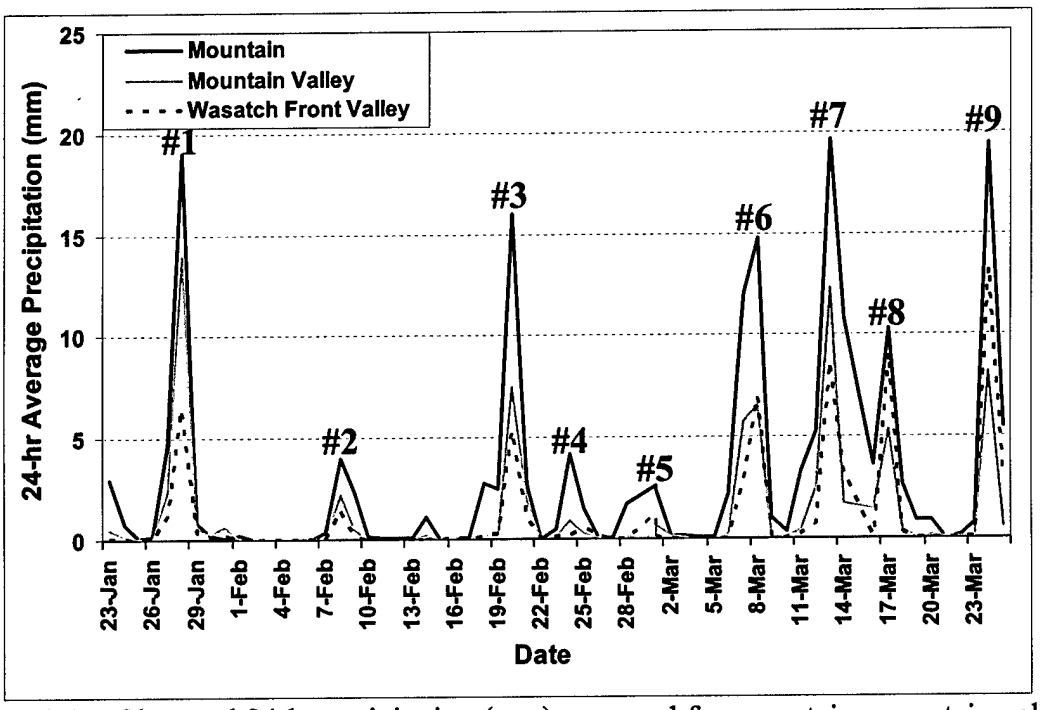



Figure 3.15. Observed 24-h precipitation (mm) averaged for mountain, mountain valley and Wasatch Front Valley locations.

scores for thresholds of 0.254, 2.54, and 6.35 mm were comparable for both the 12- and 4-km domains and likely associated with at least marginal overprediction at thresholds of 0.254 and 6.35 mm where scores were 1.5 or more (Fig. 3.16a). At thresholds greater than 2.54 mm, 4-km domain bias scores were larger than those of the 12-km domain, particularly at the heaviest thresholds, but most likely closer to the actual precipitation when considering the effects of undercatchment. In contrast, at the 19.05-mm and 25.4-mm thresholds, 12-km domain bias scores of 0.8 and 0.5, respectively, were likely associated with significant underprediction.

The 24-h ET scores were higher, indicative of greater skill, for the 4-km domain at all thresholds (Fig. 3.16b). More specifically, the 4-km domain ET scores were 5–10% larger for the smaller 24-h thresholds (0.254, 2.54, and 6.35 mm); however, at higher thresholds (19.05 and 25.4 mm) threat scores were 50–100% higher for the 4-km domain. FARs for the 4-km domain were similar to those of the 12-km domain for lighter thresholds and smaller at higher thresholds (Fig. 3.16c), suggesting that threat score improvements did not necessarily come at the expense of more false alarms. Probabilities of detection generally decreased with increasing thresholds for both domains (Fig. 3.16d), an indication that higher threshold events were more difficult to forecast. Based on the probabilities of detection, the 4-km domain forecasts were as much as twice as likely to verify for events with at least 12.7 mm observed in 24 h. These 24-h precipitation results suggest that the improved terrain representation of the 4-km domain results in better simulations of larger precipitation events that are likely associated with orographic forcing.

The contingency table scores used above do not reflect the magnitude of

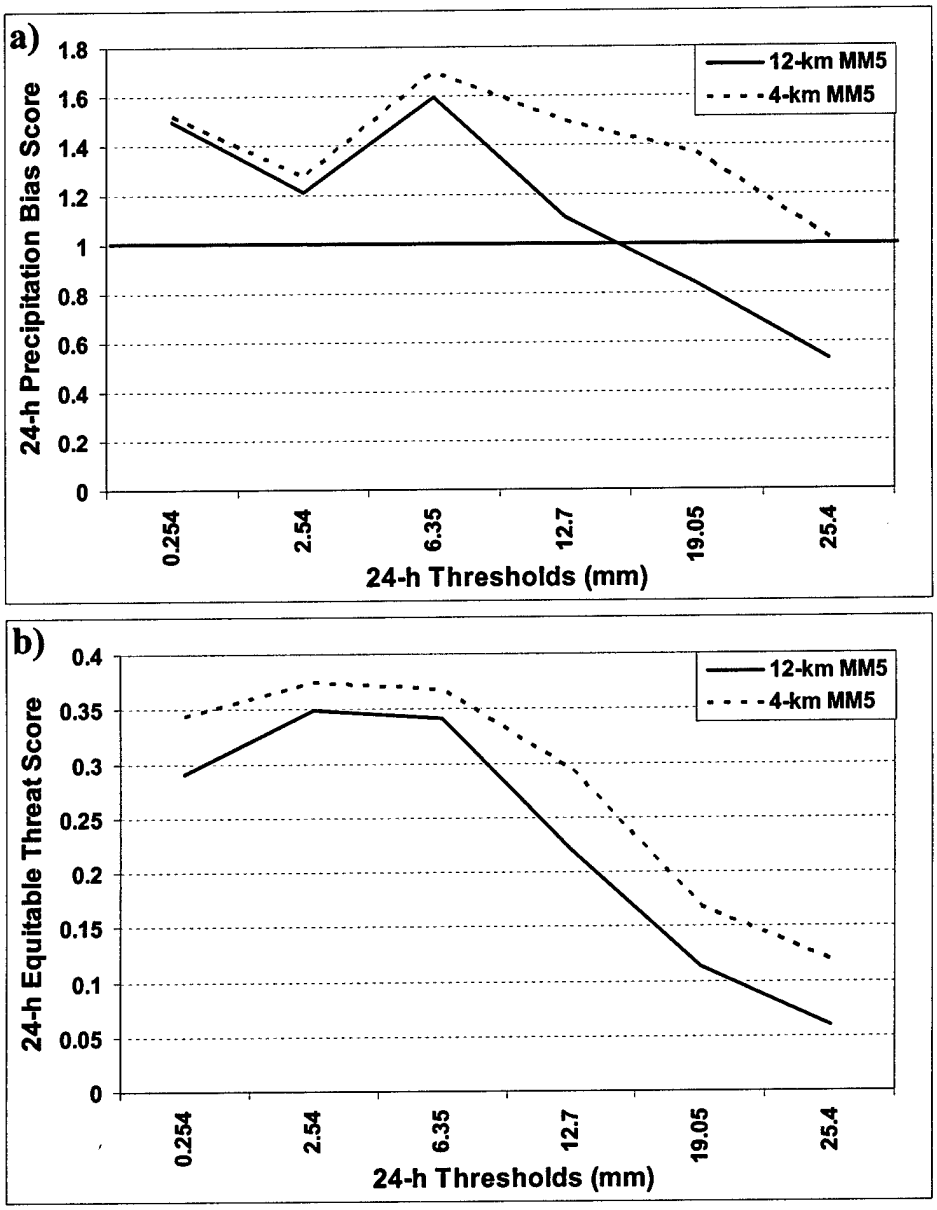



Figure 3.16. The 24-h precipitation (a) bias scores, (b) equitable threat scores, (c) false alarm ratios, and (d) probabilities of detection for the 12- and 4-km domains. Scores are averaged for the 0000 UTC and 1200 UTC initialized model runs for forecast hours 0–24 and 12–36, respectively.

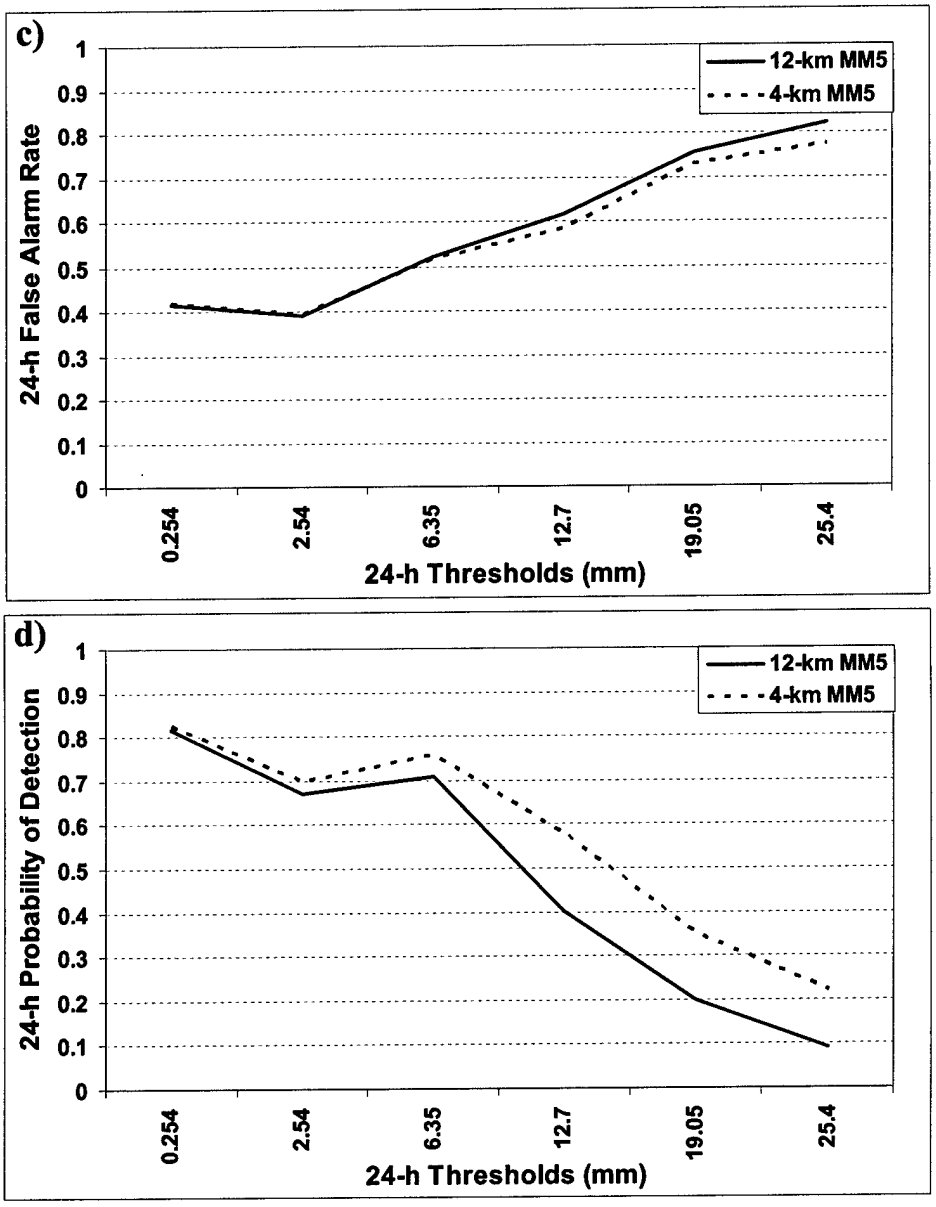



Figure 3.16. (continued)

precipitation forecast errors. In an attempt to represent 12- and 4-km domain 24-h precipitation forecast errors quantitatively, root-mean-squared errors were calculated for several thresholds (Fig. 3.17). At thresholds lower than 6.35 mm, 4-km domain rms errors were not appreciably different than those from the 12-km domain. At higher thresholds (≥ 12.7 mm in 24 h), however, 4-km domain rms errors were 10–15% lower than the 12-km domain, suggesting again that the improved terrain representation associated with the higher resolution 4-km domain added forecast skill for the larger precipitation events.

Bias scores were also computed at each observation site to better illustrate spatial model performance by calculating the ratio of total 24-h forecast precipitation to observed precipitation for the 0000 UTC and 1200 UTC runs and then multiplying by 100 to produce percentages. Collectively, the 12- and 4-km domains produced bias scores of 100% or more for all site types (Fig. 3.18). The largest bias scores were observed at mountain valley locations where the 12- and 4-km domains produced 241% and 185% of the observed precipitation, respectively. At the lower-elevation Wasatch Front sites, the 4-km domain forecasted 175% of the observed precipitation compared to 133% for the 12-km domain. Since variations in instrument catch efficiencies were not considered, some of the overprediction values may be attributed to undercatchment. However, bias scores ≥ 160% are likely associated with model overprediction.

The standard deviations in Fig. 3.18 indicate that there was considerable variability in the station bias scores, particularly at mountain valley locations. A closer inspection of the spatial characteristics of bias scores illustrates several smaller-scale model precipitation features (Fig. 3.19). For example, the 12-km domain underforecasted precipitation over the Stansbury, Oquirrh and portions of the northern and southern

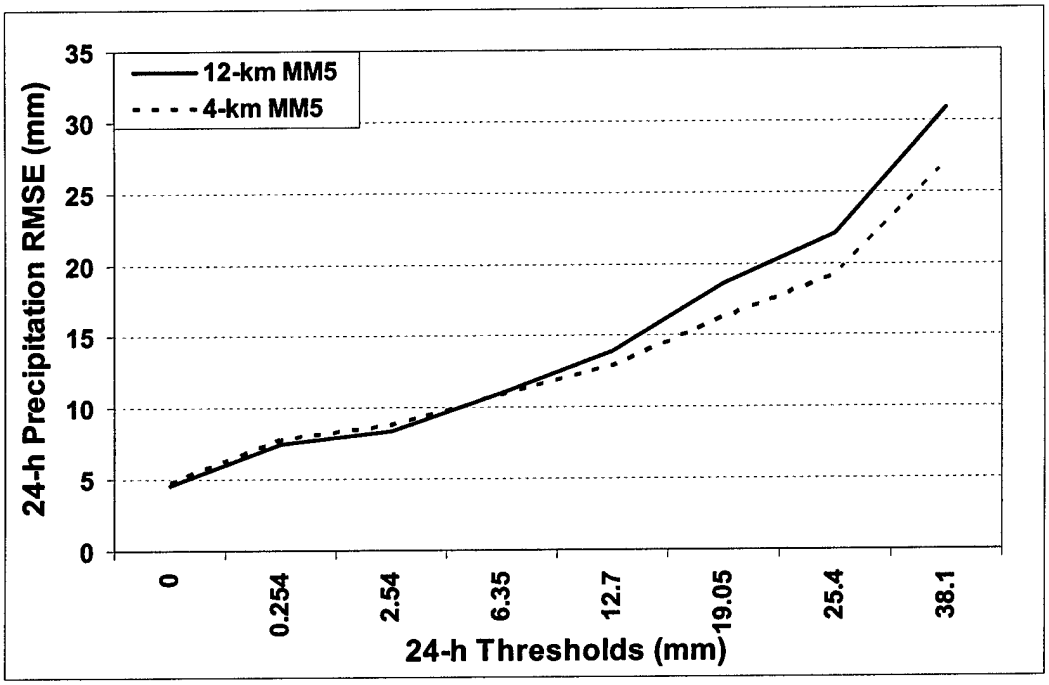



Figure 3.17. The 24-h rms error (mm) for the 12- and 4-km domains. Errors are averaged for the 0000 and 1200 UTC initialized model runs for forecast hours 0–24 and 12–36, respectively.

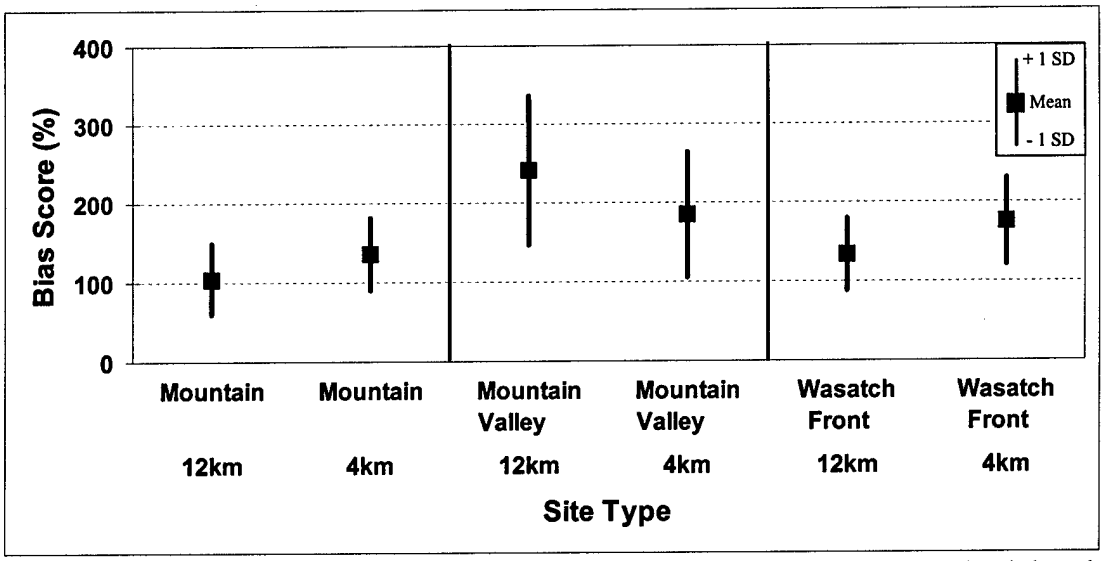



Figure 3.18. Mean precipitation bias scores (x100) ± 1 standard deviation (SD) by site type.

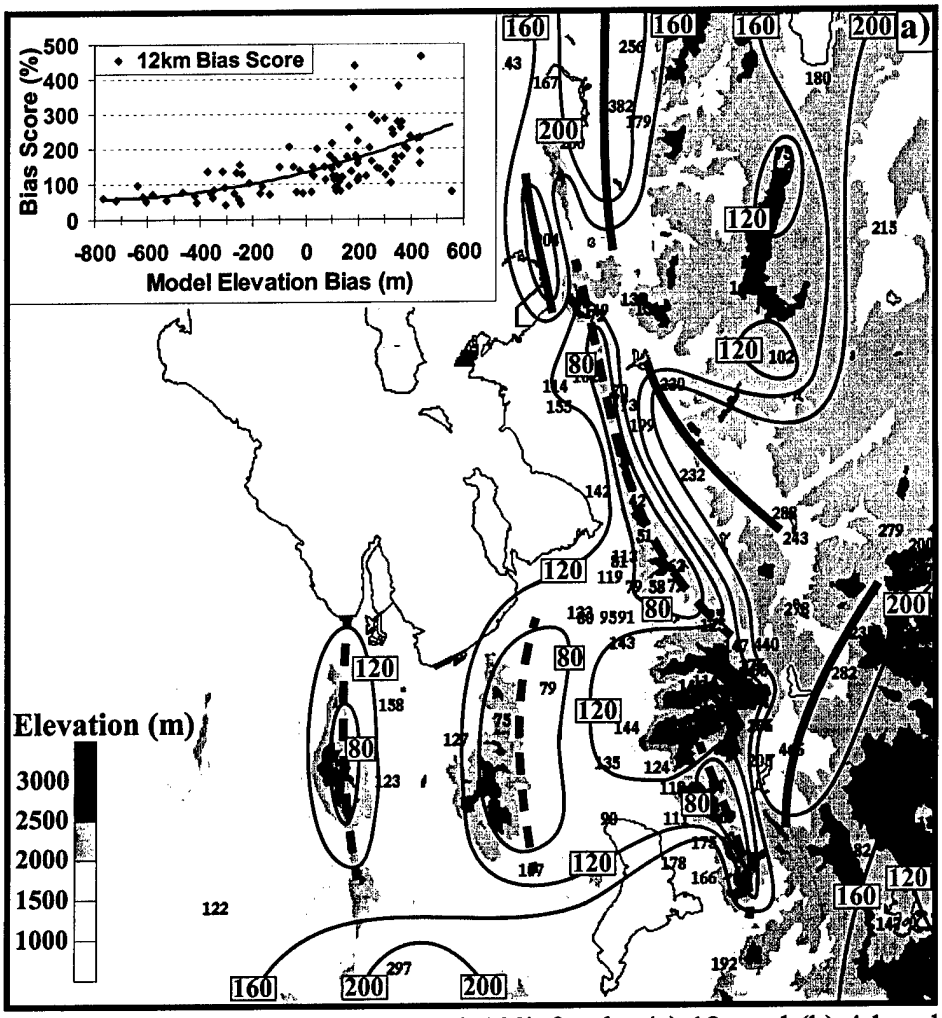



Figure 3.19. 24-h precipitation bias scores (x100) for the (a) 12- and (b) 4-km domains. Contours drawn at 40% intervals from 80% to 200%. Bold lines indicate local maxima (solid) and minima (dashed). 24-h precipitation bias scores (x100) vs model elevation bias (m) (inset).

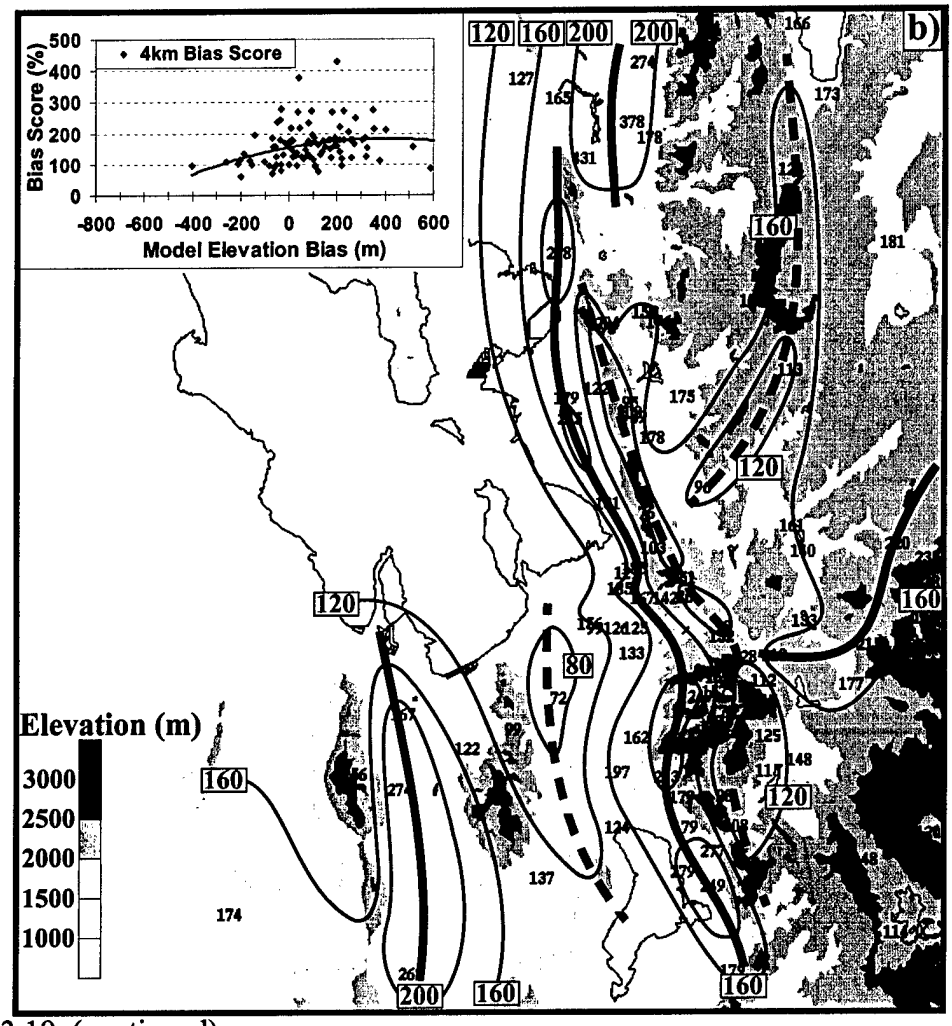



Figure 3.19. (continued)

Wasatch Mountains (Fig. 3.19a). The 12-km domain significantly overforecasted precipitation (≥ 200%) at mountain valley sites east of the Wasatch Crest including those in the Cache Valley near the Utah/Idaho border and the Heber and Ogden-Morgan Valleys (see Fig. 3.1 for locations). The lack of narrow mountain ridges (i.e., the Stansbury, Oquirrh, and Wasatch Mountains) and valleys in the 12-km domain led to a strong association between these bias scores and model elevation errors (Fig. 3.19a). For example, at most mountain locations where the model elevation is lower than observed, the 12-km domain significantly underforecasted precipitation amounts. In contrast, at sites where the model elevations were higher than observed, most notably at mountain valley locations east of the Wasatch Crest, the 12-km domain produced too much precipitation.

Although mountain ranges in the 4-km domain are still lower and broader than observed, the 4-km domain bias scores were closely tied to topographic features. The 4-km domain bias scores reflected local minima along and immediately to the lee of the major mountain crests (Fig. 3.19b). In addition, the 4-km domain overpredicted precipitation (local bias score maxima) along the eastern (windward) bench locations where the windward slope is displaced upstream from its actual location. While the 4-km domain produced local minima in bias scores at the high-elevation sites (along and immediately to the lee of major mountain crests), more precipitation was produced relative to the 12-km domain. Better terrain representation in the 4-km domain mitigated the strong correlation between bias scores and model elevation errors observed with the 12-km domain (Fig. 3.19b).

Two remaining unanswered questions concern how the model forecasts over shorter-time periods may be sensitive to errors in the timing of large-scale weather

systems and how model errors change with time. Therefore, bias scores, ET scores, FARs, and PODs were also calculated at 6-h intervals for thresholds of 0.254, 2.54, and 6.35 mm from 6 through 36 h (Fig. 3.20). For the 0.254-mm threshold, bias scores were only slightly higher for the 4-km domain. For larger precipitation amounts, the bias scores were about 20–30% larger for the 4-km domain. While some precipitation events were restricted to portions of the study domain (e.g., mountains), it should be noted that 6-h precipitation amounts averaged for the entire area only exceeded 2.54 mm during seven events. Similar to the results of Colle et al. (1999, 2000) over the Pacific Northwest, bias scores increased steadily through 18 h, particularly for the larger 6.35-mm threshold, suggesting that the model, initialized with no hydrometeors, was spinning up precipitation throughout this period.

Equitable threat scores generally decreased through 36 h (Fig. 3.21), however, the largest ET scores occurred at 12 h for the 2.54- and 6.35-mm thresholds which again may be related to the model spin-up of precipitation (Colle et al. 1999, 2000). For the 0.254-mm threshold, ET scores for the 4-km domain lagged those of the 12-km, but for the 2.54- and 6.35-mm thresholds, threat scores were 20–30% higher for the 4-km domain. These results suggest that while the improved terrain representation did not improve accuracy for the very light precipitation amounts, the 4-km domain produced better results for higher precipitation amounts when orographic forcing mechanisms may be more important.

False alarm ratios were similar for the 12- and 4-km domains at thresholds of 0.254, 2.54 and 6.35 mm (Fig. 3.22), illustrating that the improved topographic representation of the 4-km domain did not increase false alarms. The fraction of observed precipitation that was correctly forecast, or probabilities of detection, increased through

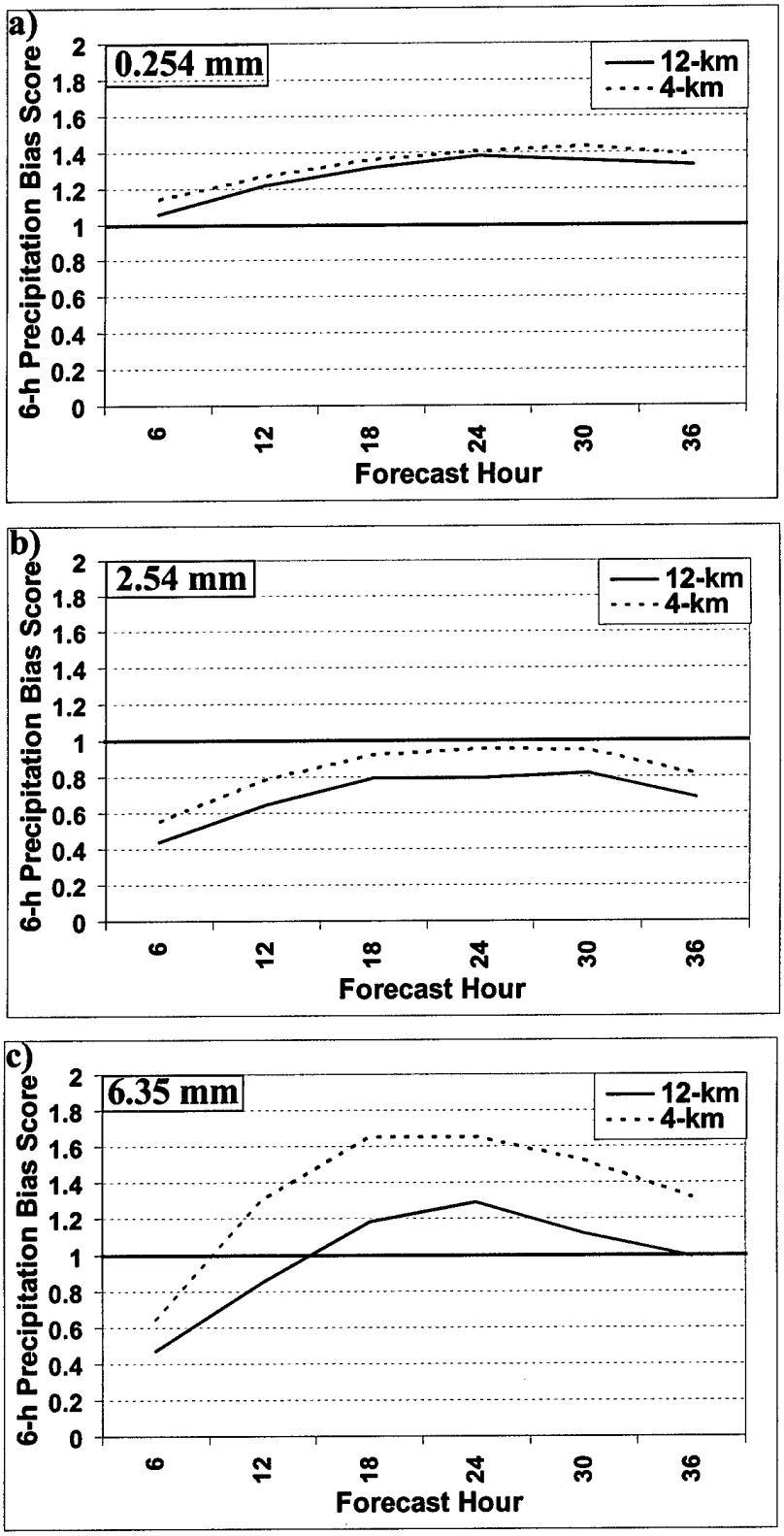



Figure 3.20. The 6-h precipitation bias scores for the (a) 0.254-, (b) 2.54-, and (c) 6.35-mm thresholds for the 12- and 4-km domains.

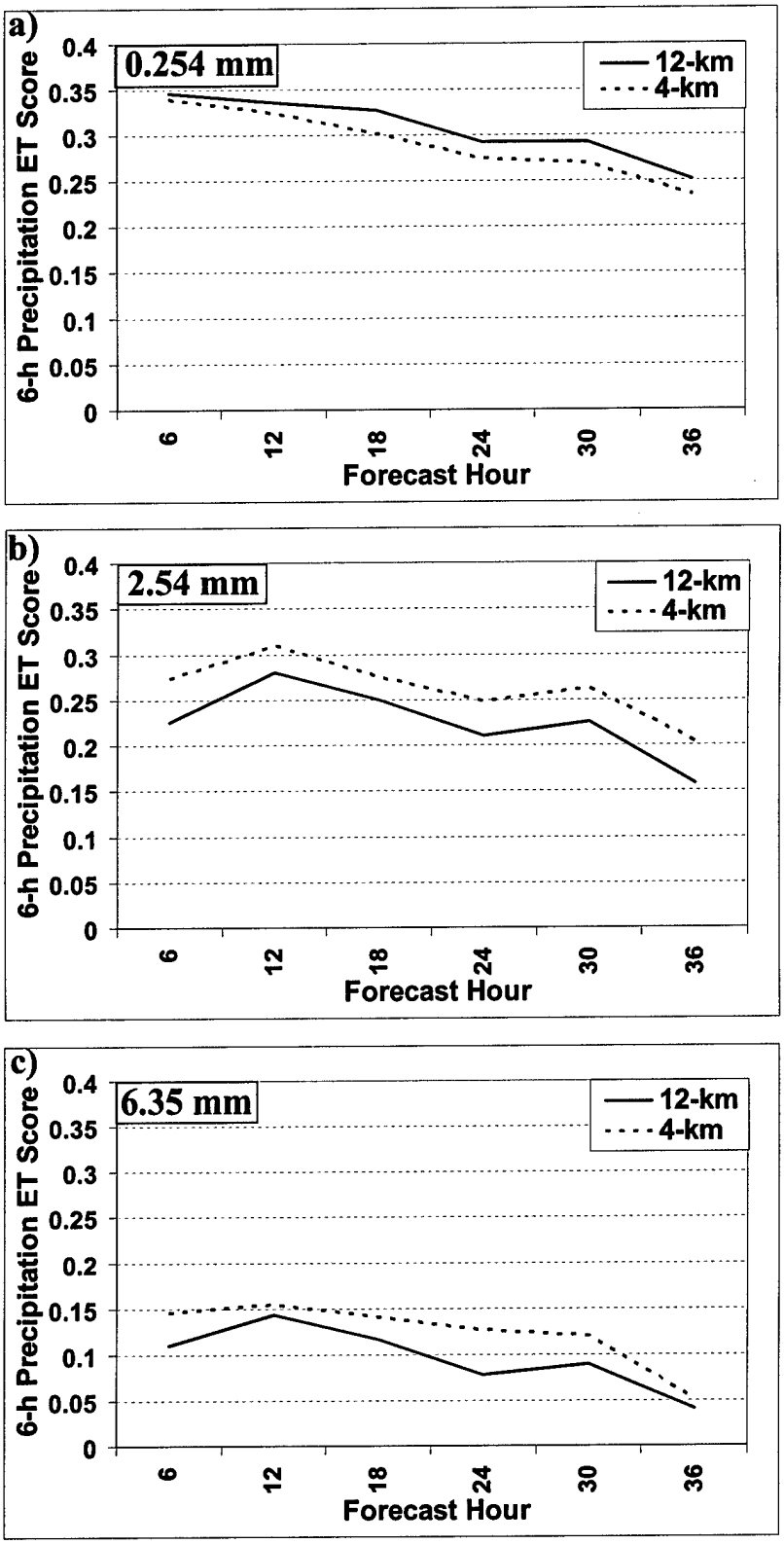



Figure 3.21. The 6-h equitable threat scores for the (a) 0.254-, (b) 2.54-, and (c) 6.35-mm thresholds for the 12- and 4-km domains.

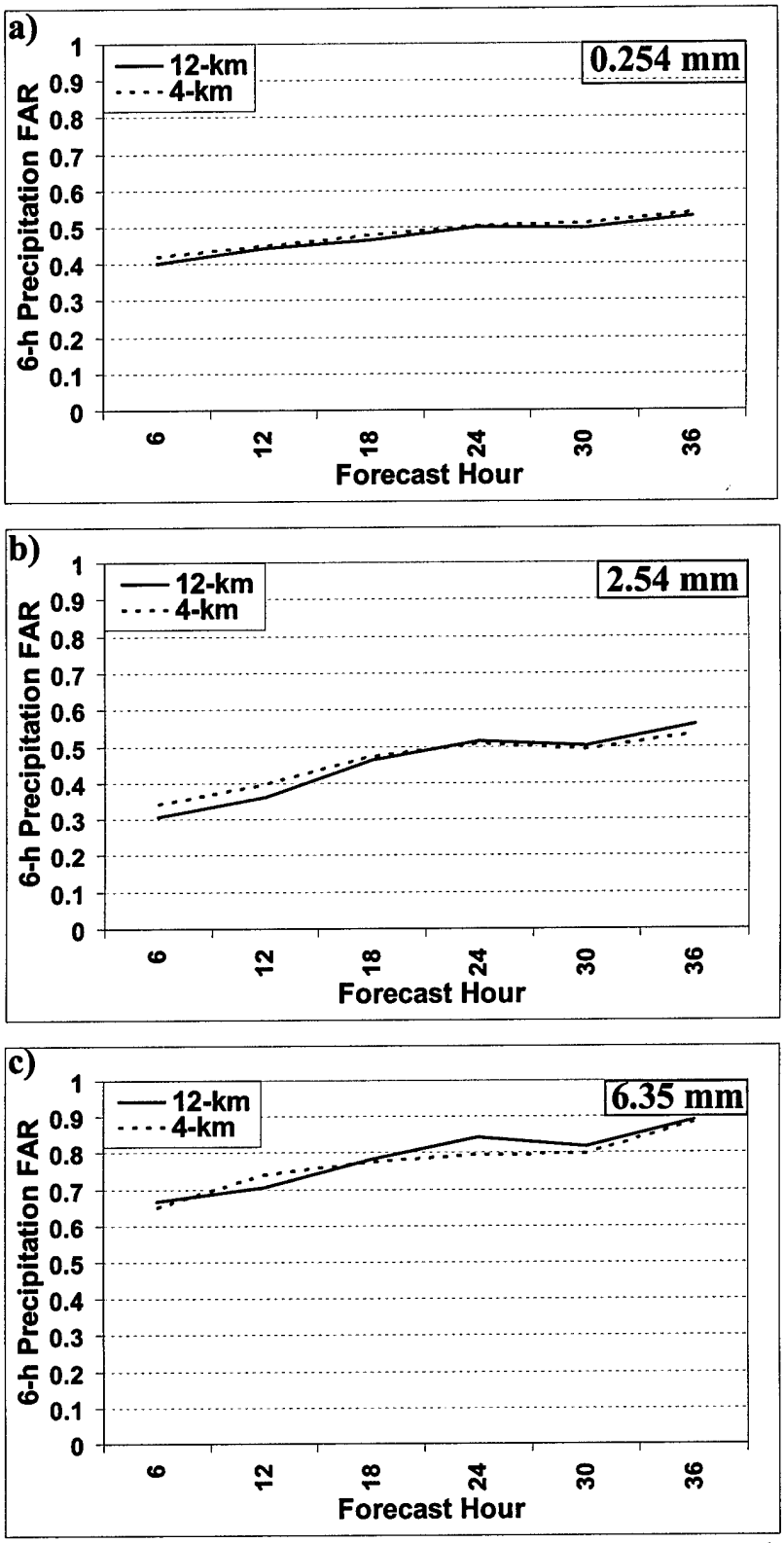



Figure 3.22. The 6-h false alarm ratios for the (a) 0.254-, (b) 2.54-, and (c) 6.35-mm thresholds for the 12- and 4-km domains.

18 h, especially at the 6.35-mm threshold where the 4-km domain added the most skill (Fig. 3.23).

## Summary and Conclusions

This chapter has examined temperature, wind, and precipitation forecast performance from the 12- and 4-km domains of the MM5 over northern Utah from 23 January 2002 through 25 March 2002, a period that included the Olympic and Paralympic Games. Nearly 200 sites reporting temperature and wind and 100 precipitation sites were used to determine if, when, and where the 4-km domain was more skilful than the 12-km domain.

Although Hart et al. (2004) found minimal improvement as model grid spacing was decreased from 12 to 4 km, the present study, which is based on a much larger number of stations and also features precipitation verification, has found that 4-km domain wind and precipitation forecasts were more accurate. There were statistically significant overall improvements in wind speed and wind direction forecasts when the horizontal grid spacing was decreased from 12 to 4 km, illustrating at least one possible advantage to using high-resolution nesting over the Intermountain West. More specifically, the 4-km domain produced better simulations of the nocturnal flows in lowland regions which led to better wind speed forecasts at these locations. In addition, as a result of improved resolution of the terrain at mountain locations, the 4-km domain was able to better simulate the exposure of these sites to the free atmosphere and, thereby, produced better wind direction forecasts. These results imply that higher-resolution simulations could be valuable for fire weather prediction and transport and dispersion simulations, two critical

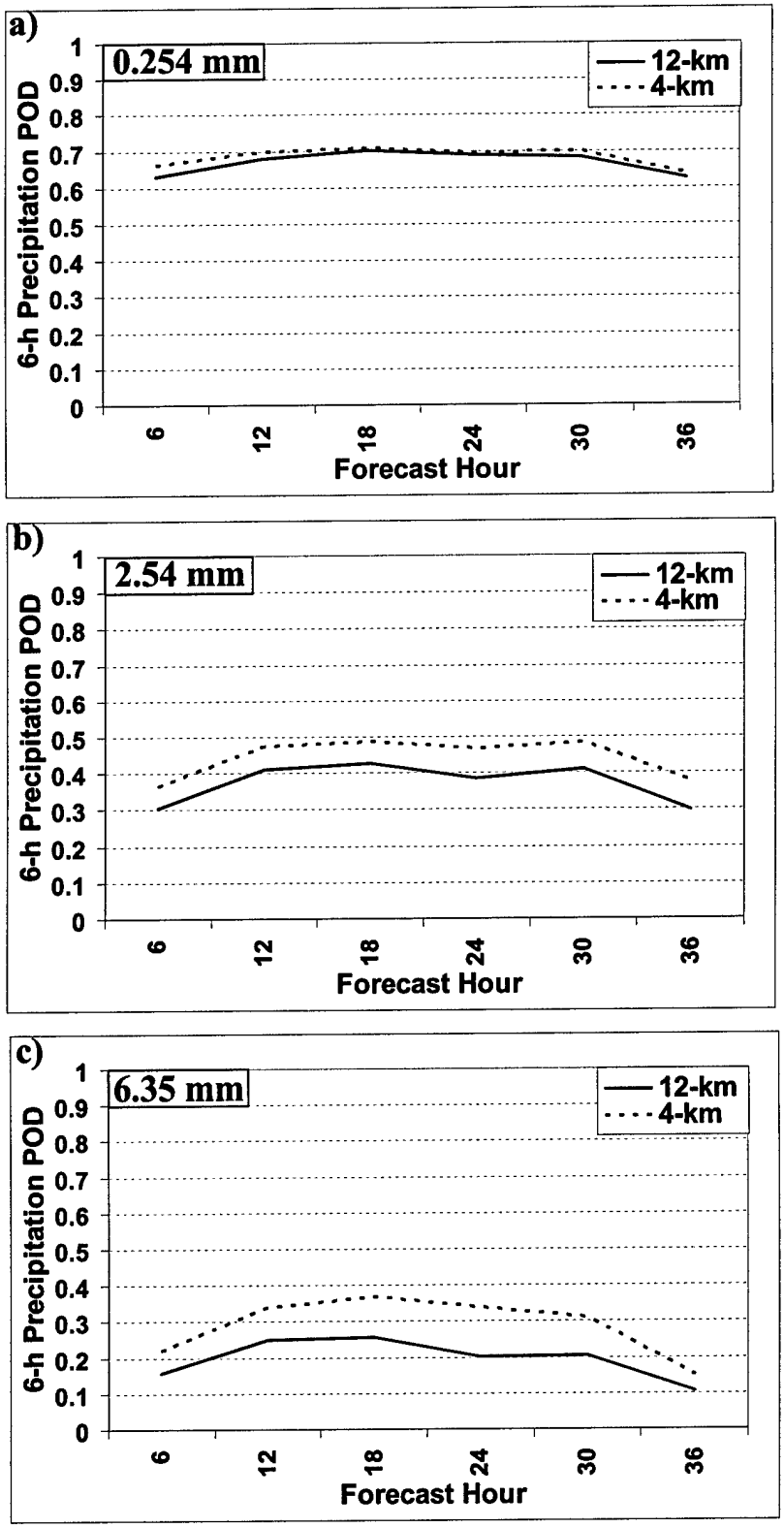



Figure 3.23. The 6-h probabilities of detection for the (a) 0.254-, (b) 2.54-, and (c) 6.35-mm thresholds for the 12- and 4-km domains.

forecast challenges in regions with heavily populated valleys, generally low humidities, and summer precipitation minima. For example, spot forecasts performed for prescribed burns and wildfires are strongly dependent on accurate wind forecasts. In addition, clearing index forecasts, used by air-quality decision makers, are very sensitive to winds. However, overall improvements in wind speed and wind direction forecasts associated with decreasing grid spacing from 12 to 4 km were not as substantial as those found by Hart et al. (2004) using MOS based on the 12-km domain of the MM5. Based on the results of the present study, mesoscale-model-based MOS using the higher-resolution 4-km domain should be explored for potential forecast improvements.

Precipitation verification at 24- and 6-h intervals suggest that the 4-km domain was more accurate for larger precipitation thresholds (> 6.35 mm for 24-h and ≥ 2.54 mm for 6-h intervals). Larger precipitation amounts are often associated with orographic precipitation processes that are more accurately simulated with the better terrain representation found in the 4-km domain. The 12-km domain underpredicted precipitation along and immediately to the lee of major mountain crests and both domains overpredicted precipitation at most valley locations. The 12-km domain biases showed a strong correlation to model elevation error as significant underprediction was observed where model elevations were too low (e.g., mountain sites). Similarly, the 12-km domain produced too much precipitation where model elevations were too high (e.g., mountain valley sites). More realistic topography in the 4-km domain reduced correlations strictly between bias scores and model elevation errors that were more pronounced with the 12-km domain. However, the 4-km domain overpredicted precipitation in the lowlands immediately upstream of the Wasatch Mountains where the influence of the unrealistic,

broad windward slope in the model led to a westward shift of simulated precipitation and higher bias scores in portions of the Wasatch Front.

In contrast, overall differences in temperature accuracy between the 4- and 12-km domains were not significant. At lower-elevation mountain valley and Wasatch Front locations, where nocturnal or persistent cold pools were more prevalent, temperature MAEs and BEs for the 4-km domain were similar or higher than those of the 12-km domain. The inability of the MM5 to properly simulate boundary layer structure during cold pool events often led to better surface temperature forecasts by the 12-km domain at these locations despite the better topography in the 4-km domain. However, a closer inspection revealed that differences were more substantial at certain site types and under certain conditions. For example, the better representation of the Wasatch and other adjacent mountain ranges by the 4-km domain led to forecast improvements at high-elevation locations where conditions were more representative of the free atmosphere. These results highlight the requirement for improvements in model initialization and boundary layer parameterization before the advantages of high resolution can be fully realized over finescale basin-and-range topography.

This study has demonstrated that decreasing grid spacing from 12 to 4 km improves wind and precipitation forecasts over the finescale topography of the Intermountain West where the mountains are characterized by a half-width of ~5–10 km. Similar improvements from decreasing grid spacing below 12 km were not observed in studies over the Pacific Northwest (Colle et al. 1999, 2000; Mass et al. 2002) where the additional resolution of the broader Cascades (~50-km half-widths) offered by higher-resolution nests may not produce the same benefit. The results also suggest that

improvements in temperature are also possible if the ability of the model to simulate nocturnal or persistent cold pools were improved. While the effects of instrument siting and gauge type on undercatchment were broadly considered when analyzing the results from this study, future work should include better estimates of undercatchment based on the analysis of co-located temperature and wind observations over a longer period of record. However, surface observing instrument selection and siting is highly variable due to the large array of operational requirements and, as a result, the wide application of simple algorithms to estimate undercatchment presents significant challenges.

# CHAPTER 4

## SUMMARY AND CONCLUSIONS

This study has examined the effect of decreasing model grid spacing from 12 to 4 km and the use of mesoscale-model-based statistical techniques on forecast accuracy over finescale Intermountain topography. Numerical model and associated MOS forecasts from a multiply-nested version of the MM5 over the complex terrain of northern Utah were verified based on high-density observations collected by the MesoWest cooperative networks.

Unlike similar studies over broader terrain, wind and precipitation forecasts became more accurate as model grid spacing was decreased from 12 to 4 km. As a result, the direct application of high-resolution numerical model output could serve to improve fire weather forecasts and air quality prediction. The overall accuracy of temperature forecasts by the 4-km domain, however, was not significantly better than that of the 12-km domain. This lack of improvement was related to an inability to simulate the strength and duration of persistent and nocturnal cold pools leading to a warm/dry bias. Ironically, the improved terrain representation of the 4-km domain magnified this bias, resulting in larger temperature forecast errors. However, at upper-elevations, typically above the influence of cold pools, 4-km domain temperature forecasts were substantially more accurate. Improvements in model initialization and boundary layer parameterization could help mitigate errors during these cold pool episodes and, thereby, enable additional model skill

from decreases in model grid spacing over the finescale basin-and-range topography.

MOS surface temperature, relative humidity, wind speed, and wind direction forecast errors were considerably smaller than those produced by the 12- and 4-km MM5 and comparable or smaller than those of the Olympic Forecast Team. These results highlight the need to improve model physics and suggest that traditional MOS, by accounting for and correcting systematic biases, continues to provide better forecasts than can be achieved by simply decreasing horizontal grid spacing (< 10 km). MOS or other statistical techniques based on high-density surface observations (e.g., MesoWest) could be used to improve gridded forecasts of surface sensible weather.

This study suggests that both statistical techniques and decreased grid spacing may be beneficial for environmental forecast applications over regions with finescale topography. For example, applications that require high-resolution, three-dimensional wind output with precipitation (e.g., smoke plume dispersion, fire weather forecasting, and air quality prediction) would benefit from the improved accuracy at 4-km grid spacing. On the other hand, applications requiring point-specific, surface sensible weather forecasts would benefit from statistically filtered model output, even if it is based on a relatively coarse resolution model or ensemble system. Specifically, gridded forecasts produced by the National Weather Service Interactive Forecast Preparation System would benefit from the use of MOS or other statistical techniques to improve model guidance.

## REFERENCES


Abdel-Aal, R. E., and M. A. Elhadidy, 1995: Modeling and forecasting the daily maximum temperature using abductive machine learning. *Wea. Forecasting*, **10**, 310–325.

Cheng, L., 2001: Validation of quantitative precipitation forecasts during the Intermountain Precipitation Experiment. M.S. thesis, Dept. of Meteorology, University of Utah, 137 pp. [Available from Dept. of Meteorology, University of Utah, 135 South 1460 East, Salt Lake City, UT 84112.]

Clements, C. B., C. D. Whiteman, and J. D. Horel, 2003: Cold-air-pool structure and evolution in a mountain basin: Peter Sinks, Utah. *J. Appl. Meteor.*, **42**, 752–768.

Colle, B. A., C. F. Mass, and K. J. Westrick, 2000: MM5 precipitation verification over the Pacific Northwest during the 1997-99 cool seasons. *Wea. Forecasting*, **15**, 730–744.

_____, _____, and D. Ovens, 2001: Evaluation of the timing and strength of MM5 and Eta surface trough passages over the eastern Pacific. *Wea. Forecasting*, **16**, 553–572.

_____, K. J. Westrick, and C. F. Mass, 1999: Evaluation of MM5 and Eta-10 precipitation forecasts over the Pacific Northwest during the cool season. *Wea. Forecasting*, **14**, 137–154.

Cooper, W. A., and C. P. R. Saunders, 1980: Winter storms over the San Juan Mountains. Part II: The microphysical processes. *J. Appl. Meteor.*, **19**, 925–941.

Davis, C., T. Warner, E. Astling, and J. Bowers, 1999: Development and application of an operational, relocatable, mesogamma-scale weather analysis and forecasting system. *Tellus 51A*, **5**, 710–727.

Doran, J. C., J. D. Fast, and J. Horel, 2002: The VTMX 2000 campaign. *Bull. Amer. Meteor. Soc.*, **83**, 537–551.

Dudhia, J., 1989: Numerical study of convection observed during the Winter Monsoon Experiment using a mesoscale two-dimensional model. *J. Atmos. Sci.*, **46**, 3077–3107.

Fritsch, J. M., and Coauthors, 1998: Quantitative precipitation forecasting: Report of the eighth prospectus development team, U. S. Weather Research Program, *Bull. Amer. Meteor. Soc.*, **79**, 285–299.

Glahn, H. R., 1985: Statistical weather forecasting. *Probability, Statistics, and Decision Making in the Atmospheric Sciences*, A. H. Murphy and R. W. Katz, Eds., Westview Press, 289–335.

_____, and D. A. Lowry, 1972: The use of model output statistics (MOS) in objective weather forecasting. *J. Appl. Meteor.*, **11**, 1203–1211.

_____, and D. P. Ruth, 2003: The new digital forecast database of the National Weather Service. *Bull. Amer. Meteor. Soc.*, **84**, 195–201.

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. [Available from National Center for Atmospheric Research, P.O. Box 3000, Boulder, CO 80307.]

Groisman, P. Y., and D. R. Legates, 1994: The accuracy of United States precipitation data. *Bull. Amer. Meteor. Soc.*, **75**, 215–227.

Hallet, J., and S. C. Mossop, 1974: The production of secondary ice particles during the riming process. *Nature*, **249**, 26–28.

Hart, K. A., W. J. Steenburgh, D. J. Onton, and A. J. Siffert, 2004: An evaluation of mesoscale-model-based model output statistics (MOS) during the 2002 Olympic and Paralympic Winter Games. *Wea. Forecasting*, **19**, 200–218.

Homleid, M., 1995: Diurnal corrections of short-term surface temperature forecasts using the Kalman filter. *Wea. Forecasting*, **10**, 689–707.

Hong, S.–Y., and H.-L. Pan, 1996: Nonlocal boundary layer vertical diffusion in a medium-range forecast model. *Mon. Wea. Rev.*, **124**, 2322–2339.

Horel, J., and Coauthors, 2002a: MesoWest: Cooperative mesonets in the western United States. *Bull. Amer. Meteor. Soc.*, **83**, 211–226.

_____, T. Potter, L. Dunn, W. J. Steenburgh, M. Eubank, M. Splitt, and D. J. Onton, 2002b: Weather support for the 2002 Winter Olympic and Paralympic Games. *Bull. Amer. Meteor. Soc.*, **83**, 227–240.

Hsieh, W. W., and B. Tang, 1998: Applying neural network models to prediction and data analysis in meteorology and oceanography. *Bull. Amer. Meteor. Soc.*, **79**, 1855–1870.

Jacks, E., B. Bower, V. J. Dagostaro, J. P. Dallavalle, M. C. Erickson, and J. C. Su, 1990: New NGM-based MOS guidance for maximum/minimum temperature, probability of precipitation, cloud amount, and surface wind. *Wea. Forecasting*, **5**, 128–138.

Kain, J. S., and J. M. Fritsch, 1993: Convective parameterizations for mesoscale models: The Kain-Fritsch scheme. *The Representation of Cumulus Convection in Numerical Models, Meteor. Monogr.*, No. 46, Amer. Meteor. Soc., 165–170.

Lazarus, S. M., C. M. Ciliberti, J. D. Horel, and K. A. Brewster, 2002: Near-real-time applications of a mesoscale analysis system to complex terrain. *Wea. Forecasting*, **17**, 971–1000.

Majewski, D., 1997: Operational regional prediction. *Meteor. Atmos. Phys.*, **63**, 89–104.

Marwitz, J. D., 1986: *A comparison of winter orographic storms over the San Juan Mountains and the Sierra Nevada, Precipitation Enhancement--A Scientific Challenge*, R. R. Braham, Jr., American Meteorological Society, Boston, pp. 109–113.

_____, 1987: Deep orographic storms over the Sierra Nevada. Part II: The precipitation processes. *J. Atmos. Sci.*, **44**, 174–185.

Mass, C. F., 2003: IFPS and the future of the National Weather Service. *Wea. Forecasting*, **18**, 75–79.

_____, D. Ovens, K. Westrick, and B. A. Colle, 2002: Does increasing horizontal resolution produce more skillful forecasts? The results of two years of real-time numerical weather prediction over the Pacific Northwest. *Bull. Amer. Meteor. Soc.*, **83**, 407–430.

Onton, D. J., and W. J. Steenburgh, 2001: Diagnostic and sensitivity studies of the 7 December 1998 Great Salt Lake-effect snowstorm. *Mon. Wea. Rev.*, **129**, 1318–1328.

Paegle, J., R. A. Pielke, G. A. Dalu, W. Miller, J. R. Garratt, T. Vukicevic, G. Berri, and M. Nicolini, 1990: Predictability of flows over complex terrain. *Atmospheric Processes over Complex Terrain*, Meteor. Monogr., No. 45, Amer. Meteor. Soc., 285–299.

Panofsky, H. A., and G. W. Brier, 1958: *Some Applications of Statistics to Meteorology*. The Pennsylvania State University, 224 pp.

Rangno, A. L., 1986: How good are our conceptual models of orographic cloud seeding? *Precipitation Enhancement–A Scientific Challenge, Meteor. Monogr.*, No. 43, Amer. Meteor. Soc., 115–124.

Rauber, R. M., 1987: Characteristics of cloud ice and precipitation during wintertime storms over the mountains of northern Colorado. *J. Climate Appl. Meteor.*, **26**, 488–524.

Sall, J., A. Lehman, and L. Creightin, 2001: *JMP Start Statistics*, 2d ed. Duxbury Thomson

Learning, 491 pp.

Schultz, D. M., and Coauthors, 2002: Understanding Utah winter storms: The Intermountain Precipitation Experiment. *Bull. Amer. Meteor. Soc.*, **83**, 189–210.

Serreze, M.C., M. P. Clark, and A. Frei, 2001: Characteristics of large snowfall events in the montane western United States as examined using snowpack telemetry (SNOTEL) data. *Water Resources Res.*, **37**, 675–688.

Siffert, A.J., 2001: Point-specific MOS forecasts for the 2002 Winter Games. M.S. thesis, Dept. of Meteorology, University of Utah, 51 pp. [Available from Dept. of Meteorology, University of Utah, 135 South 1460 East, Salt Lake City, UT 84112.]

Smith, R., and Coauthors, 1997: Local and remote effects of mountains on weather: Research needs and opportunities. *Bull. Amer. Meteor. Soc.*, **78**, 877–892.

Stewart, J. Q., C. D. Whiteman, W. J. Steenburgh, and X. Bian, 2002: A climatological study of thermally driven wind systems of the U.S. Intermountain West. *Bull. Amer. Meteor. Soc.*, **83**, 699–708.

Vislocky, R. L., and J. M. Fritsch, 1995: Improved model output statistics forecasts through model consensus. *Bull. Amer. Meteor. Soc.*, **76**, 1157–1164.

Warner, T. T., R. A. Peterson, and R. E. Treadon, 1997: A tutorial on lateral boundary conditions as a basic and potentially serious limitation to regional numerical weather prediction. *Bull. Amer. Meteor. Soc.*, **78**, 2599–2617.

Whiteman, C. D., 1990: Observations of thermally developed wind systems in mountainous terrain. *Atmospheric Processes over Complex Terrain, Meteor. Monogr.*, No. 45, Amer. Meteor. Soc., 5–42.

_____, 2000: *Mountain Meteorology: Fundamentals and Applications*. Oxford University Press, 355 pp.

_____, S. Zhong, W. J. Shaw, J. M. Hubbe, X. Bian, and J. Mittelstadt, 2001: Cold pools in the Columbia Basin. *Wea. Forecasting*, **16**, 432–447.

Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sciences*. Academic Press, 467 pp.

Williams, P., Jr., and W. J. Heck, 1972: Areal coverage of precipitation in northwestern Utah. *J. Appl. Meteor.*, **11**, 509–516.

Wilson, L. J., and M. Vallée, 2002: The Canadian Updateable Model Output Statistics (UMOS) System: Design and development tests. *Wea. Forecasting*, **17**, 206–222.

Wolyn, P. G., and T. B. McKee, 1989: Deep stable layers in the Intermountain western United States. *Mon. Wea. Rev.*, **117**, 461–472.