



**A Case Study of the Persistence of
Weather Forecast Model Errors**

by Barbara Sauter

ARL-TR-3418

January 2005

NOTICES

Disclaimers

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

Citation of manufacturer's or trade names does not constitute an official endorsement or approval of the use thereof.

Destroy this report when it is no longer needed. Do not return it to the originator.

Army Research Laboratory

White Sands Missile Range, NM 88002-5501

ARL-TR-3418

January 2005

A Case Study of the Persistence of Weather Forecast Model Errors

Barbara Sauter
Computational and Information Sciences Directorate

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE (DD-MM-YYYY) January 2005		2. REPORT TYPE Final		3. DATES COVERED (From - To)	
4. TITLE AND SUBTITLE A Case Study of the Persistence of Weather Forecast Model Errors				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S) Barbara Sauter				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) U.S. Army Research Laboratory Computational and Information Sciences Directorate Battlefield Environment Division (ATTN: AMSRD-ARL-CI-EM) White Sands Missile Range, NM 88002-5501				8. PERFORMING ORGANIZATION REPORT NUMBER ARL-TR-3418	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Research Laboratory 2800 Powder Mill Road Adelphi, MD 20783-1145				10. SPONSOR/MONITOR'S ACRONYM(S) ARL	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S) ARL-TR-3418	
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution is unlimited.					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT Decision makers could frequently benefit from information about the amount of uncertainty associated with a specific weather forecast. Automated numerical weather prediction models provide deterministic weather forecast values with no estimate of the likely error. This case study examines the day-to-day persistence of forecast errors of basic surface weather parameters for four sites in northern Utah. Although exceptionally low or high forecast errors on one day are more likely to be associated with a similar quality forecast the following day, the relationship is not considered strong enough to provide beneficial guidance to users without meteorological expertise. Days resulting in average forecast errors showed no persistence in the quality of the subsequent day's forecast. More sophisticated methods are needed to generate and portray weather forecast uncertainty information.					
15. SUBJECT TERMS Weather Forecast Uncertainty; Forecast Errors; Error Persistence; WRF; MM5; Utah MesoWest					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT SAR	18. NUMBER OF PAGES 48	19a. NAME OF RESPONSIBLE PERSON Barbara Sauter
a. REPORT U	b. ABSTRACT U	c. THIS PAGE U			19b. TELEPHONE NUMBER (Include area code) (505) 678-2840

Content

List of Figures	iv
List of Tables	vi
Summary	1
1. Introduction	3
2. Methodology	3
2.1 Forecast Model Runs.....	3
2.2 Surface Validation Stations.....	4
3. Results	6
3.1 Case Study Considerations.....	6
3.2 Background Information on Station Observed Weather and Forecast Errors.....	6
3.2.1 Observed Weather.....	6
3.2.2 Forecast Errors.....	8
3.3 Forecast Error Persistence.....	10
3.3.1 Detailed Error Persistence Case Study Results.....	10
3.3.2 Summary Error Persistence Results by Station.....	23
3.3.3 Summary Error Persistence Results by Weather Parameter.....	25
3.3.4 Summary Error Persistence Results by Time of Day.....	26
3.3.5 Summary Error Persistence Results by Season.....	26
3.3.6 Summary Error Persistence Results by Model.....	27
4. Conclusions	28
4.1 Overall Case Study Results.....	28
4.2 Recommendations.....	29
References	31
Appendix. Average Forecast Error Ranges	32
Acronyms	39

List of Figures

Figure 1. Surface station locations.....	5
Figure 2. Observed temperatures most frequently occurred within the ranges shown for each station and time. Summer months depicted by a red line, winter months by a blue line.....	7
Figure 3. Observed dew-point temperatures most frequently occurred within the ranges shown for each station and time. Summer months depicted by a red line, winter months by a blue line.....	7
Figure 4. WRF dew-point temperature forecasts compared to the observations at 2 a.m. for the summer months at station DVE.....	10
Figure 5. Mean temperature errors at DVE for summer nighttime forecasts. Data markers are coded in green (better), yellow (average), or red (worse).....	11
Figure 6. Mean wind speed errors at DVE for summer daytime forecasts. Data markers are coded in green (better), yellow (average), or red (worse).....	13
Figure 7. The number of days with the dew-point temperature forecast error category of better (green bars), average (yellow bars), and worse (red bars), given the previous one-day forecast error category on the left and the previous two-day forecast error category on the right, for PRP at 2 p.m. in the summer.....	15
Figure 8. The daily dew-point temperature forecast error amounts for PRP at 2 p.m. in the summer.....	16
Figure 9. The daily temperature forecast error amounts at SNZ at 2 a.m. in the summer.....	18
Figure 10. The number of days with the dew-point temperature forecast error category of better (green bars), average (yellow bars), and worse (red bars), given the previous one-day forecast error category on the left and the previous two-day forecast error category on the right, for WBB at 2 a.m. in the summer.....	20
Figure 11. The daily dew-point temperature forecast error amounts for WBB at 2 a.m. in the summer.....	20
Figure 12. WRF wind speed forecasts compared to the observations at 2 a.m. for the summer months at WBB.....	21
Figure 13. Daily temperature forecast error categories for the daytime winter forecasts for WBB. The marker labels indicate the forecast day in order from 1 through 60, equating to the forecast period of January 1 through February 29, 2004.....	22

Figure 14. Each chart shows the combined results for all forecasts for stations (a) DVE, (b) PRP, (c) SNZ, and (d) WBB. The y-axis indicates the percentage of subsequent days with the forecast error category of better (green bars), average (yellow bars), and worse (red bars), given the previous day’s forecast error category, which is listed on the horizontal labels for the x-axis on the left half of each chart. The right half of each chart provides the same information, except the x-axis groupings are based on having two previous days in the same error category.24

Figure 15. Each chart shows the combined results for all forecasts for weather parameters (a) temperature, (b) dew-point temperature, (c) wind speed, and (d) wind direction. See figure 14 caption for chart details.25

Figure 16. Each chart shows the combined results for (a) the nighttime forecasts and (b) the daytime forecasts. See figure 14 caption for chart details.26

Figure 17. Each chart shows all the forecast results for (a) the summer forecasts and (b) the winter forecasts. See figure 14 caption for chart details.27

Figure 18. Each chart shows the all the forecast results for (a) the WRF and (b) the MM5. See figure 14 caption for chart details.27

Figure A-1. The range of absolute temperature errors used for the average error category is shown by the line for (a) MM5 summer forecasts, (b) WRF summer forecasts, (c) MM5 winter forecasts, and (d) WRF winter forecasts. One-third of the forecasts in each category had errors greater than the highest value and one-third had errors less than the lowest value depicted by the line. The asterisk associated with each line is the average absolute error for the given station and time.34

Figure A-2. The range of absolute dew-point temperature errors used for the average error category is shown by the line for (a) MM5 summer forecasts, (b) WRF summer forecasts, (c) MM5 winter forecasts, and (d) WRF winter forecasts. One-third of the forecasts in each category had errors greater than the highest value and one-third had errors less than the lowest value depicted by the line. The asterisk associated with each line is the average absolute error for the given station and time.35

Figure A-3. The range of absolute wind speed errors used for the average error category is shown by the line for (a) MM5 summer forecasts, (b) WRF summer forecasts, (c) MM5 winter forecasts, and (d) WRF winter forecasts. One-third of the forecasts in each category had errors greater than the highest value and one-third had errors less than the lowest value depicted by the line. The asterisk associated with each line is the average absolute error for the given station and time.36

Figure A-4. The range of absolute wind direction errors used for the average error category is shown by the line for (a) MM5 summer forecasts, (b) WRF summer forecasts, (c) MM5 winter forecasts, and (d) WRF winter forecasts. One-third of the forecasts in each category had errors greater than the highest value and one-third had errors less than the lowest value depicted by the line. The asterisk associated with each line is the average absolute error for the given station and time.37

List of Tables

Table 1. Surface station information.....	4
Table 2. Average differences between WRF forecast and observed values. Observed wind data at SNZ is considered unreliable.	8
Table 3. Percent of forecasts in each error category 24 hours later, given the previous day's error category.....	28
Table 4. Percent of forecasts in each error category on the third day, given a consistent error category over the previous two days.....	29

Summary

Military and civilian activities affected by the weather require decision makers to rely on uncertain weather forecasts. Specific information on the amount of uncertainty associated with a forecast has historically been non-existent, or based on a human expert indicating a level of confidence. Organizations and individuals involved in producing or using weather forecasts are becoming more aware of the value of including information about the amount of uncertainty associated with a forecast. At the same time, information processes are becoming more automated, relying less on human expertise.

Various organizations are investigating methods to automatically calculate and portray weather forecast uncertainty. One proposed method was to provide a “stoplight” chart indicating whether the validated model forecast, from either 24 or 48 hours earlier, had produced errors that were either better than, the same as, or worse than the average error for the particular model being used. Specifically, for each numerical weather prediction model available, the average error produced by that model over some historical period would be known. If the forecast validated the previous day produced a significantly lower than average error for a particular model, the Web link to that model output would be preceded by a green circle; whereas, if a model generated a typical error on the previous day relative to its own historical performance it would be preceded by a yellow circle, or if it produced a higher than average error it would be preceded by a red circle. Two color-coded circles would be presented if the previous two days’ results were considered. This case study looks at model forecast errors on consecutive days to determine how useful it might be for a user to be given a “better,” “average,” or “worse” marker based on the previous day’s forecast.

This case study is based on forecast errors for both the Pennsylvania State University/NCAR Mesoscale Model Version 5 (MM5) and the Weather Research and Forecasting (WRF) model for four specific sites in the vicinity of Salt Lake City, Utah. Ranges of forecast error values were defined as better, average, or worse for single parameter forecasts at the same location, time of day, season, and model. The divisions between the error categories were based on the thresholds that would produce an equal number of forecasts in each category. The summer months included 54 days and the winter months included 57 days, resulting in approximately 18-19 forecasts in each error category for each permutation of 4 stations, 4 weather parameters, 2 times of day, 2 seasons, and 2 forecast models.

A day-by-day analysis shows how frequently the error category persisted for consecutive days. The report provides many details of these analyses. Forecasts that were better or worse than the average range were more likely to be associated with forecasts in the same category the following day. These persistent better and worse error categories generally occurred on 40-50% of the following days, with about 25% of the following days falling in the opposite error

category. The days with average forecast errors were equally likely to be followed by any of the three forecast error categories. The original suggestion that precipitated this effort is not being pursued. The results seen in this study provide further justification that a simple stoplight marker based on a previous day's forecast error is not sufficient to allow weather forecast users to incorporate an appropriate level of uncertainty information into their decisions involving probability, risk, and benefits.

1. Introduction

Weather forecasters regularly use the term “persistence” to indicate non-changing weather from one day to the next. Experienced forecasters will also recall specific instances when a forecast model performed poorly, and suspect poor performance from that model when similar conditions occur. This case study examines the likelihood of a poor model forecast to persist from one day to the next, without consideration of the weather pattern influencing the area.

In a previous study, two mesoscale forecast models were run over northern Utah (1). These models were the Pennsylvania State University/National Center for Atmospheric Research (NCAR) Mesoscale Model Version 5 (MM5) and the first version of the Weather Research and Forecasting (WRF) model, under development by NCAR, the National Oceanic and Atmospheric Administration, and several universities (2, 3). The results of those hourly model forecasts compared to surface observations at 50-70 sites showed some very large errors. In order to examine the model performance in more detail, and to investigate any patterns in the quality of the forecasts, this case study focuses on four of the surface stations at one time during the night and one time during the day.

Although the goal of this study emphasizes the relative performance of the individual model at a specific site over a limited timeframe, many general error statistics are included. These individual error amounts do not accurately reflect the models’ capabilities to the degree indicated by the values given, since the surface sensors used for validation do not meet standard height or calibration requirements. The Utah MesoWest Cooperative provides error flags for questionable observations, but these flags were not available for this study (4). Only rudimentary error-checking was performed on the observations to ensure they were within possible limits.

2. Methodology

2.1 Forecast Model Runs

Both the MM5 and WRF models were initialized with Global Forecast System (GFS) model output data (5). The MM5 was run with a coarse domain and two nests on an Army High Performance Computing Resource Center’s Cray computer, while the non-nested version of the WRF was run on a U.S. Army Research Laboratory Major Shared Resource Center IBM system (6, 7). The model forecast data used for comparison with surface observations was based on 5-km grid spacing for both models. Model runs were initialized at 1800 Greenwich Mean Time (GMT) for 28 days in July 2003, 26 days in August 2003, 28 days in January 2004, and 29 days in February 2004. The initial 6 hours were used for spin-up purposes, so that the WRF validation times of 0900 and 2100 GMT used in this study would be considered a 9- and 21-hour

forecast. These times were 2 a.m. and 2 p.m. Mountain Standard Time (MST). Through an inconsistency in the handling of the WRF and MM5 validation files for this case study, the times available for the MM5 analysis were one hour later each day, valid at 3 a.m. and 3 p.m. MST. The forecast model parameters of temperature and dew-point temperature at 2 m and u - and v -wind components at 10 m were horizontally interpolated to the surface station locations. Wind speed and direction values have been calculated from the wind components and are used in the discussion of the results.

2.2 Surface Validation Stations

As mentioned in the introduction, this case study leverages an existing dataset. Four meteorological stations were selected for closer scrutiny based on their dew-point temperature errors for a single day, reflecting a wide range of forecast error amounts:

- Promontory Point, UT (PRP): much worse than average
- DV-Burns, UT (DVE): worse than average
- WBB/U Utah, UT (WBB): better than average
- Bountiful/F G House, UT (SNZ): much better than average

That criterion is not consistent for all the days used in this study, but the four stations previously selected were kept for this analysis. Site information for these stations is listed in table 1. PRP is situated on a hill on a peninsula extending into the Great Salt Lake, while DVE is in the Wasatch Mountains. WBB and SNZ are in urban areas, with WBB on the roof of a University of Utah building in Salt Lake City and SNZ on a 3-m tower in the backyard of a house in the vicinity of Bountiful, Utah. All four stations are part of the SNOwNET mesonet.

Table 1. Surface station information.

ID	Name	Latitude (°N)	Longitude (°W)	Elevation (ft)
PRP	Promontory Point	41.2575	112.4398	6926
DVE	DV - Burns	40.6228	111.4520	7333
WBB	WBB/U Utah	40.76623	111.84755	4910
SNZ	Bountiful/F G House	40.8754	111.8716	4760

Figure 1 shows the location of the four stations on a map of the Great Salt Lake/Wasatch Front areas, highlighting the complex terrain of northern Utah.

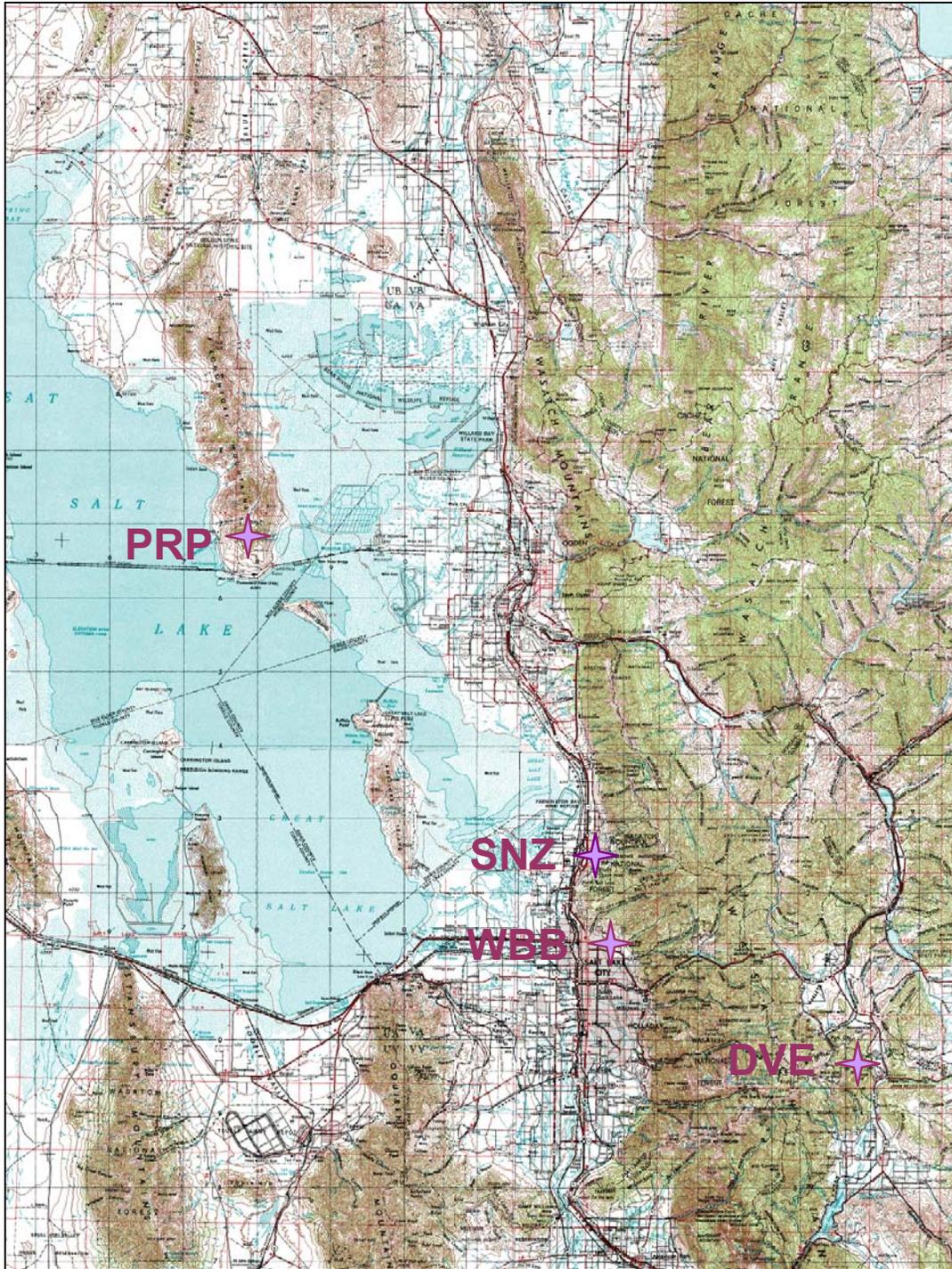


Figure 1. Surface station locations.

3. Results

3.1 Case Study Considerations

Although an attempt was made to narrow the scope of the thousands of results from the previous model output study, the resulting subset of data from two models, four stations, two times per day for two seasons still poses difficulties in analyzing and describing meaningful results. The primary goal of the analysis is to determine how often the quality level of a numerical model forecast persisted from one forecast to the following day's forecast. Some definition of a persistent error must be used. In order to accomplish this, the forecasts are classified as "better" (the third of the forecasts with the lowest errors), "average" (the third in the middle), or "worse" (the third with the highest errors). This grouping results in approximately equal numbers of forecasts in each classification. If no correlation exists between the quality of a forecast on a preceding day and the quality of the forecast on the subsequent day, each preceding day classification category would result in approximately 33% of each classification on the next day.

The sample size may be too small to obtain statistically significant results, and the classification groupings occasionally place forecasts with very small error differences in adjacent groups. With these limitations in mind, the following sections describe some of the results seen relating to the persistence of forecast errors from one forecast to the forecast 24 hours later.

3.2 Background Information on Station Observed Weather and Forecast Errors

3.2.1 Observed Weather

Within each season, the individual stations' temperatures most frequently fell within a 10-12 °C range over the two-month period, with occasional observations above or below that range. Figure 2 highlights these prevailing ranges, providing the summer temperatures with a red line and winter temperatures with a blue line. SNZ showed the highest daytime temperatures, but was cooler than the other urban station, WBB, at night during the summer.

The most frequently observed dew-point temperatures were typically within a 10-15 °C range. As expected, the dew-point temperatures were significantly lower than the temperatures in the summer months, but were usually quite close to the temperatures in the winter months.

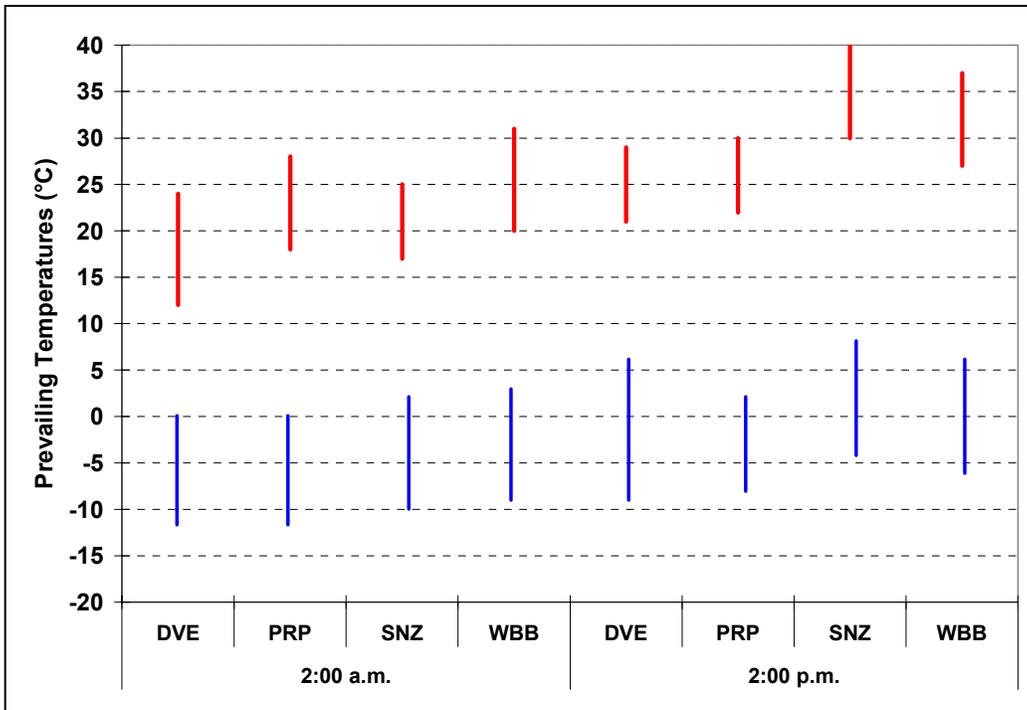


Figure 2. Observed temperatures most frequently occurred within the ranges shown for each station and time. Summer months are depicted by a red line, winter months by a blue line.

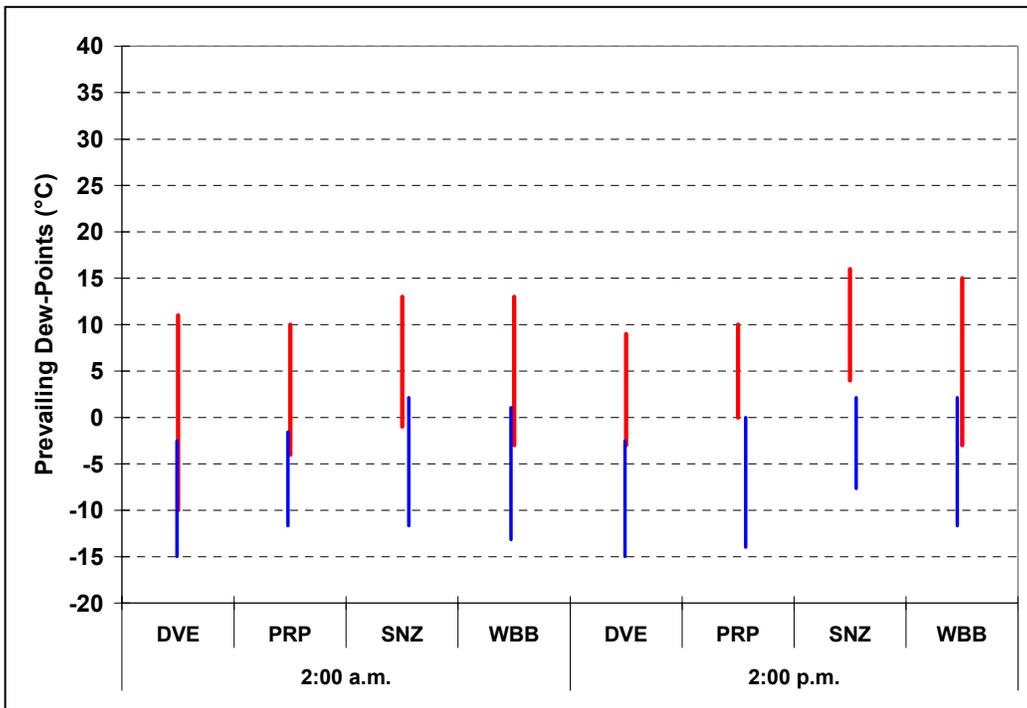


Figure 3. Observed dew-point temperatures most frequently occurred within the ranges shown for each station and time. Summer months are depicted by a red line, winter months by a blue line.

Wind speeds were predominantly light. Wind direction was only analyzed for this study for observations with corresponding wind speeds greater than 1 m/s. The nighttime winds at DVE were usually below 2 m/s, with primarily southerly winds in the summer and no determinant direction in the winter. DVE’s daytime wind reports were slightly stronger, concentrated between 0 and 4 m/s, and most frequently occurred from either a northerly or south to southwesterly direction. The other high elevation station, PRP, reported somewhat higher wind speeds, generally between 1 and 7 m/s. The summertime wind came from all directions during the night, while winds with a westerly component were most frequent during the day. Southerly winds were most common at PRP during the winter; however, all wind directions occurred a significant portion of the time. SNZ almost always reported wind speeds below 1 m/s. In addition to suspecting the observations were not reliable, SNZ did not report for ten days in mid-August. No analyses of the SNZ wind data will be provided. WBB showed both diurnal and seasonal variations in observed winds. During the summer months, wind speeds were generally between 1 and 6 m/s, but the direction varied from predominantly north to east at night and south to northwest during the day. The winter reports reflected wind speeds below 2.5 m/s from all directions during the night, with somewhat stronger winds, up to 5 m/s, from southwest to northwest during the day.

3.2.2 Forecast Errors

A summary of the average WRF model forecast errors is provided in table 2.

Table 2. Average differences between WRF forecast and observed values. Observed wind data at SNZ is considered unreliable.

		Summer				Winter			
		2 a.m.		2 p.m.		2 a.m.		2 p.m.	
		Mean Error	Absolute Error						
DVE	T (°C)	-2.1	2.6	0.3	1.3	1.1	1.9	2.6	2.9
	DP (°C)	8.6	8.6	7.1	7.1	3.2	3.3	2.6	3.0
	WS (m/s)	2.6	2.6	1.2	1.7	2.8	2.9	2.8	3.0
	WD (°)	32	40	-31	65	21	66	-2	69
PRP	T (°C)	-1.4	2.4	4.9	4.9	4.3	4.4	7.0	7.0
	DP (°C)	11.0	11.0	8.3	8.4	5.4	5.4	4.3	4.8
	WS (m/s)	-1.1	2.1	-1.3	1.6	-0.2	1.8	0.4	1.6
	WD (°)	-6	46	-54	65	-14	61	25	54
SNZ	T (°C)	-1.0	2.2	-5.0	5.3	1.5	2.9	1.4	3.4
	DP (°C)	5.1	5.2	1.9	2.9	0.6	2.4	-0.8	2.5
	WS (m/s)	4.3	4.3	2.6	2.6	2.7	2.7	3.5	3.5
	WD (°)	-5	6	11	49	71	71	26	27
WBB	T (°C)	-5.9	6.0	-2.6	3.0	0.5	2.4	3.4	4.2
	DP (°C)	8.2	8.2	5.3	5.4	1.4	2.2	1.4	2.8
	WS (m/s)	0.6	2.0	-1.1	1.6	0.5	1.1	1.4	1.9
	WD (°)	32	61	4	52	9	71	11	50

NOTE: Shading indicates unreliable data.

Positive values for temperature, dew-point temperature, and wind speed mean errors indicate the forecasts were higher than the observations. When the mean error listed for wind direction is positive, the forecast wind was from a direction clockwise from the observed direction.

The WRF exhibited a warm bias for all four stations during the winter, which was most pronounced at PRP, which is situated on a peninsula extending into the Great Salt Lake. The WRF forecasts for summertime temperatures were often too low, while the dew-point temperature forecasts at the same place and time were much too high, resulting in extremely overestimated relative humidity.

Average wind speed forecasts typically reflected a bias 2-3 m/s too high at DVE, with smaller (or even low) biases at the other stations, excluding SNZ. Average wind direction absolute errors consistently ranged between 40° and 70°.

The MM5 forecasts were valid one hour after the WRF forecasts. Most of these forecasts revealed average error trends similar to those seen in the WRF forecasts. The appendix provides the mean absolute errors for both the MM5 and WRF forecasts by season, station, and time of day. These charts in the appendix also highlight the range of errors categorized as “average” in the forecast error persistence analysis.

The purpose of this study is not to evaluate the forecast model performance in detail; however, the simple face validations of plots of forecast and observed data points, as shown in figure 4, strongly suggest the model forecasts contain some skill, even when the bias error is significant. The Pearson correlation coefficient for these summer dew-point temperature forecasts and observations is 0.74.

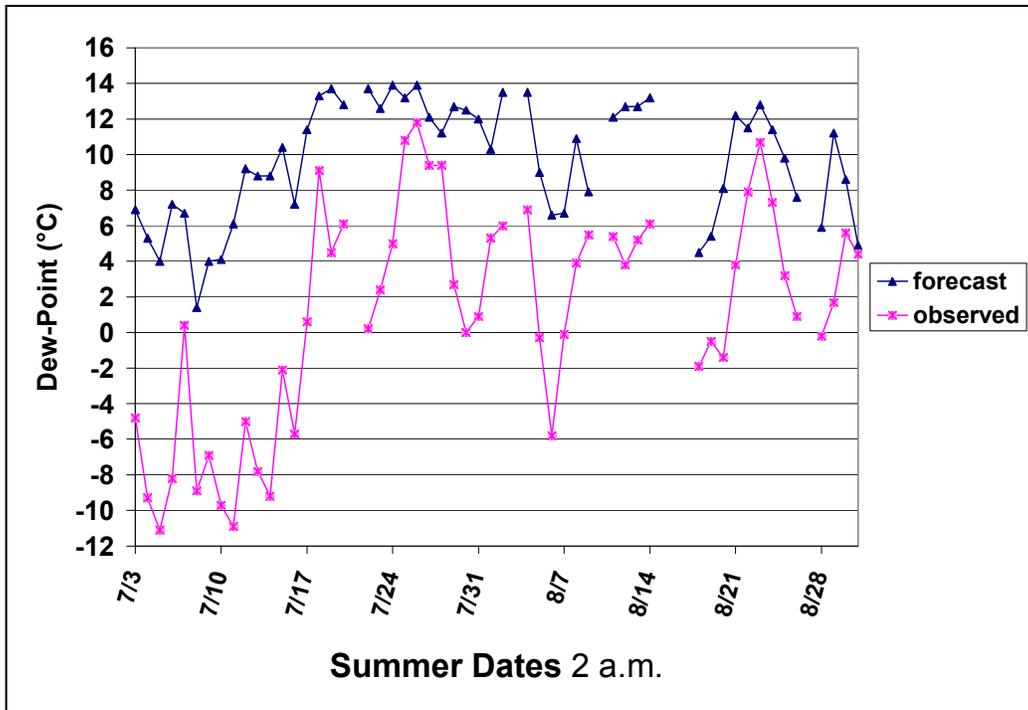


Figure 4. WRF dew-point temperature forecasts compared to the observations at 2 a.m. for the summer months at station DVE.

3.3 Forecast Error Persistence

The following sections provide the results of this case study on the day-to-day persistence of model forecast errors. Initial discussions provide some of the details seen at each of the four stations. Following those discussions, the findings are summarized by station, weather parameter, time, season, and forecast model. Whenever the forecast model is not explicitly named, the information is based on the WRF outcomes.

3.3.1 Detailed Error Persistence Case Study Results

3.3.1.1 DV-Burns (DVE)

DVE is situated several miles southeast of Park City, Utah. Although, at over 7000-ft elevation, it is the highest station used in this case study, higher mountain peaks are found both to the west and east of the site.

For the summer temperature forecasts valid at 2 a.m., a better result was only slightly more likely to be followed by a better result 24 hours later, as opposed to either an average or worse result. While more frequent than either of the other categories alone, the probability of the better forecast error category repeating was still less than 50%, which was almost always the case. An average forecast day was also only slightly more likely to be associated with an average forecast the subsequent day, rather than a better forecast, but was unlikely to be followed by a worse forecast. On the other hand, days when the 2 a.m. temperature forecast proved to be worse than usual were significantly followed by another worse temperature forecast the next day.

The categories of better, average, and worse are based strictly on absolute error amounts. Figure 5 highlights the mean temperature errors, with the absolute error category for each day displayed by a green, yellow, or red data marker. Wide day-to-day variations exist, but noticeable trends appear, including a majority of worse forecasts during the first half of July and primarily better forecasts during the second half of July. More variations are seen in the August results. The summer nighttime temperature forecast biases were predominantly too cold.

The 2 p.m. summer temperature forecasts were significantly more likely to be in the same or adjacent category than in the opposite category from the previous day. Out of the 14 days with better afternoon temperature forecasts, the following afternoon included a better forecast on 7 days, an average forecast on 5 days, and a worse forecast on 2 days. The days after an average afternoon temperature forecast were pretty evenly divided between better, average, and worse forecasts. The typical July temperature warm bias of about 2 °C in the afternoon was often found in the opposite direction as the nighttime bias, but was not as large. This study does not investigate correlations between the nighttime and daytime errors on the same date, but a cursory inspection of the DVE temperature forecasts seems to show little correlation between these errors.

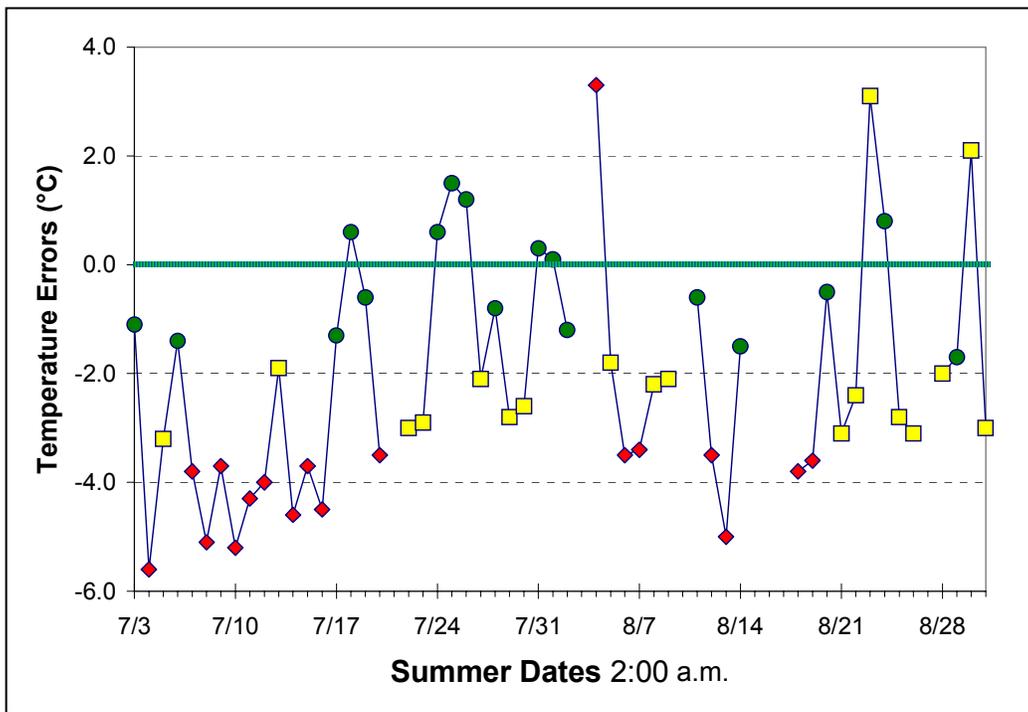


Figure 5. Mean temperature errors at DVE for summer nighttime forecasts. Data markers are coded in green (better), yellow (average), or red (worse).

The summer dew-point temperature forecasts for DVE included a very large warm bias, with average mean errors of 8.6 °C for the 2 a.m. forecasts and 7.1 °C for the 2 p.m. forecasts. Every nighttime forecast was too high, and all but one daytime forecast was too high. The actual amount of the forecast error frequently changed from one day to the next in the nighttime

forecasts, but worse forecasts were much more likely to be followed by worse forecasts again the following day. Of the 18 days in the worse category, 13 of the subsequent days also displayed worse forecast errors. This is one of the highest percentages of error persistence, and is due to dew-point temperature forecast errors greater than 10 °C on all but one day during the first two weeks in July. A single 3-day period from July 4-6 showed errors close to 15 °C each day, while a 4-day period from July 25-28 contained daily errors of approximately 2 °C each day. Otherwise, the errors differed by over 2 °C almost every day. Significant changes, by 5 °C or greater from one day to the next, occurred for 11 out of the 47 days with data available for consecutive days. The worse dew-point temperature forecasts valid at 2 p.m. were somewhat concentrated in the first half of July, like the 2 a.m. forecasts, but were more spread out over the rest of July and August with a weaker pattern of persistent errors.

Wind speed observations were frequently missing for station DVE at the 2 a.m. validation time. Almost all of the existing validations showed wind speed forecasts to be too high, with an average mean error of 2.6 m/s. It's also possible that the consistently very low wind speed observations at these times were not accurate, but other DVE reports from the same general timeframe do include higher wind speeds so the sensor was not stuck at low readings. Although few consecutive days are available, the nighttime wind speed error amounts seem to jump around fairly randomly. Most of the daytime reports for these summer dates were available including wind observations. The average mean error at 2 p.m. was 1.2 m/s with an average absolute error of 1.7 m/s. As shown in figure 6, no strong pattern exists in the persistence of absolute error category from one day to the next. The wind speed forecasts exhibit a consistent bias of too high in July, continuing but becoming less dominant in August.

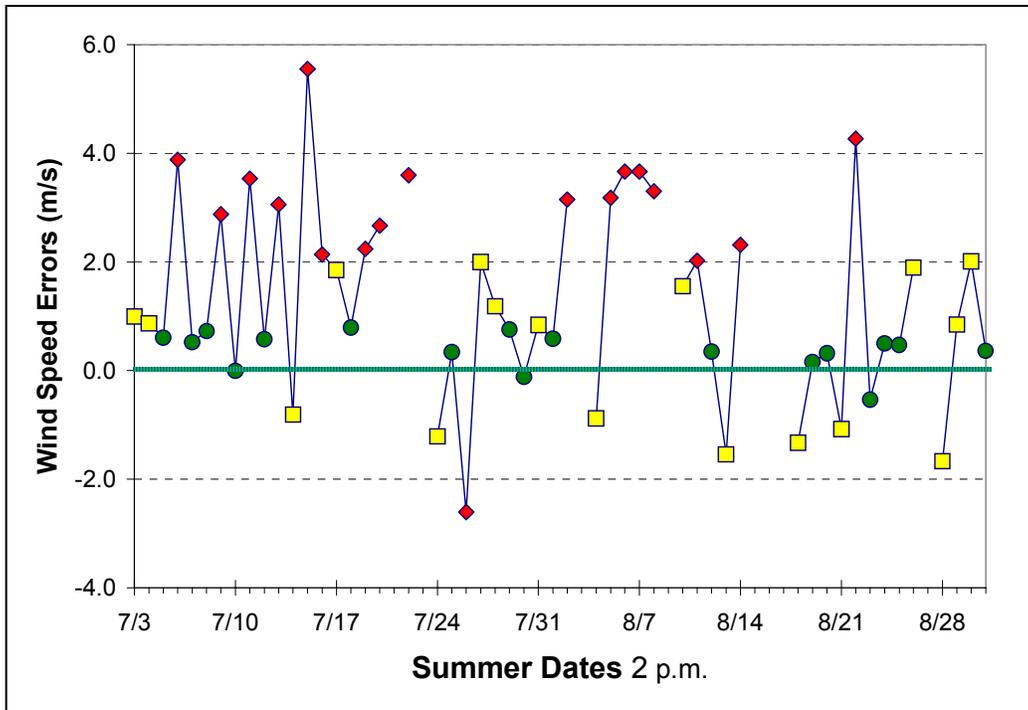


Figure 6. Mean wind speed errors at DVE for summer daytime forecasts. Data markers are coded in green (better), yellow (average), or red (worse).

Wind direction evaluations are only considered for observations with wind speed greater than 1 m/s, in order to lessen the impact of variable wind cases. Therefore, the summer dates at 2 a.m. provided too few wind direction observations to analyze. The 2 p.m. observations reflected no pattern in the amount of wind direction forecast errors for better or average forecast days, though the worse forecast days appeared less likely to be followed by a better forecast day.

The winter months included an even smaller error persistence relationship for nighttime temperature forecasts, which had only been evident in the worse category for the summer months. The persistence of the temperature forecast error category was stronger in the daytime for the summer months, and that pattern was repeated to the same or to a slightly greater degree in the winter months. The dew-point temperature forecast error averages were much smaller in the winter than in the summer months. The winter average mean and absolute errors were approximately 3 °C for both the 2 a.m. and 2 p.m. forecasts. The nighttime dew-point temperature forecasts in the better and average categories were no more likely to be followed by the same category the following day than a different category. Those forecasts falling in the worse category were slightly more likely to be worse again the next day. The daytime forecasts repeated the same category for the better and average forecasts 7 out of 18 days, while the worse forecasts preceded another worse forecast 8 out of 17 days. The summer occurrence of many missing 2 a.m. wind observations was repeated in the winter. The 2 p.m. wind observations displayed a noticeable tendency for better forecasts to be followed by better forecasts, and for worse forecasts to be followed by worse forecasts. This trend seems primarily due to the

changing pattern in wind speed forecast bias amounts, which are low or near zero in mid-January and late February. The remainder of the winter days reflected wind speed forecasts that were too high, but by varying amounts. Light winds or missing observations for the 2 p.m. forecasts provided very few consecutive days for wind direction in January. A pattern evidenced over several consecutive days in early February showed northerly observed winds associated with northwesterly forecast winds. Otherwise, the wind direction errors varied considerably from one day to the next in February.

The MM5 error patterns are discussed relative to the WRF results described above for DVE:

- The summer MM5 temperature forecast errors seemed to be slightly less persistent for the nighttime forecasts, and about the same for the daytime forecasts.
- The summer MM5 dew-point temperature errors for both the 3 a.m. and 3 p.m. valid times showed a greater relationship between a better forecast one day being followed by a better forecast the next day than was seen in the WRF results. On the other hand, the persistency of the worse forecasts was evident but not to the degree found in the WRF results. Specifically, the MM5 results showed 9 out of 18 worse forecasts followed by another worse forecast, when the WRF results were 13 out of 18.
- The winter MM5 temperature forecast errors didn't exhibit strong persistence, but did show a slightly higher incidence of better forecasts after a previous better forecast at nighttime than seen in the winter WRF forecasts. The opposite occurred in the daytime forecasts, where the WRF error persistence was stronger, while the MM5 showed no relationship.
- As was the case for WRF, the winter MM5 dew-point temperature forecast errors were not particularly persistent, with the strongest but still small relationship found in the worse category in the afternoon forecasts.
- The summer and winter MM5 wind speed and wind direction results were similar to the WRF results.

3.3.1.2 Promontory Point (PRP)

PRP is the second highest station in this case study, at approximately a 6900-ft elevation. The site is complex not solely because of the hilly terrain, but also because of its location on a peninsula extending into the Great Salt Lake.

The summer 2 a.m. WRF temperature forecasts included a negative bias of 1.4 °C that was strongest in July. The same forecast error category as the day before did not repeat most frequently, except for the better category, which resulted in another better category only one day more than an average result. Having an opposite category occur was infrequent, however. For 16 nighttime temperature forecasts qualifying as a better result, only 3 of the following days were a worse result. Out of the 15 worse results, only 2 of the following days were a better result. The 2 p.m. temperature forecast errors were much larger than the nighttime errors, and

showed a stronger persistence from one day to the next. The mode for the second day was always the same category. In each category, about half of the subsequent days fell in the same category as the previous day. The two days with extremely high errors were each followed by days with smaller errors but still much higher than average.

The summer dew-point temperature forecasts included extremely large biases of 11 °C at 2 a.m. and over 8 °C at 2 p.m. The individual categories encompassed large error ranges, and the forecast error value generally changed by several degrees on adjacent days. The largest errors in the nighttime forecasts were always followed by another larger than average error. The daytime forecasts showed a strong modal response of the second day falling in the same category as the preceding day. Figure 7 depicts this on the left side of the chart labeled “Previous Day.” For each daily forecast in the category of better, average, and worse, the bars show how many subsequent days fell in each of those categories. The right side of the chart shows the same information, except that the x-axis categories of better, average, and worse consist of cases where the two previous days were in the same category rather than just one previous day. This setup seems to display a considerable amount of persistence in the quality of the forecast. As seen in figure 8, however, there are still frequently very large variations in the dew-point temperature forecast error from one day to the next.

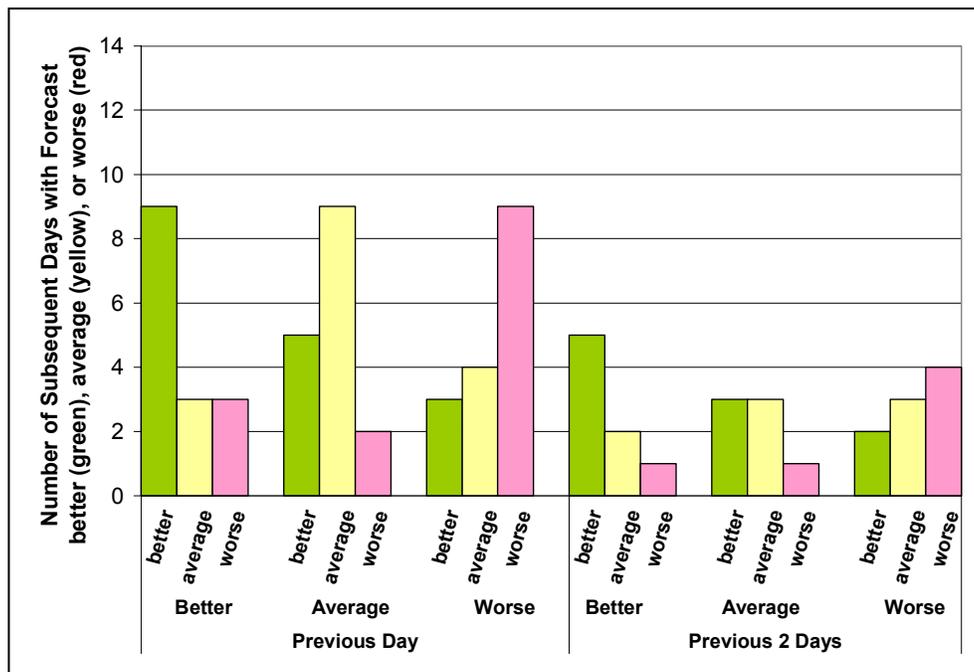


Figure 7. The number of days with the dew-point temperature forecast error category of better (green bars), average (yellow bars), and worse (red bars), given the previous one-day forecast error category on the left and the previous two-day forecast error category on the right, for PRP at 2 p.m. in the summer.

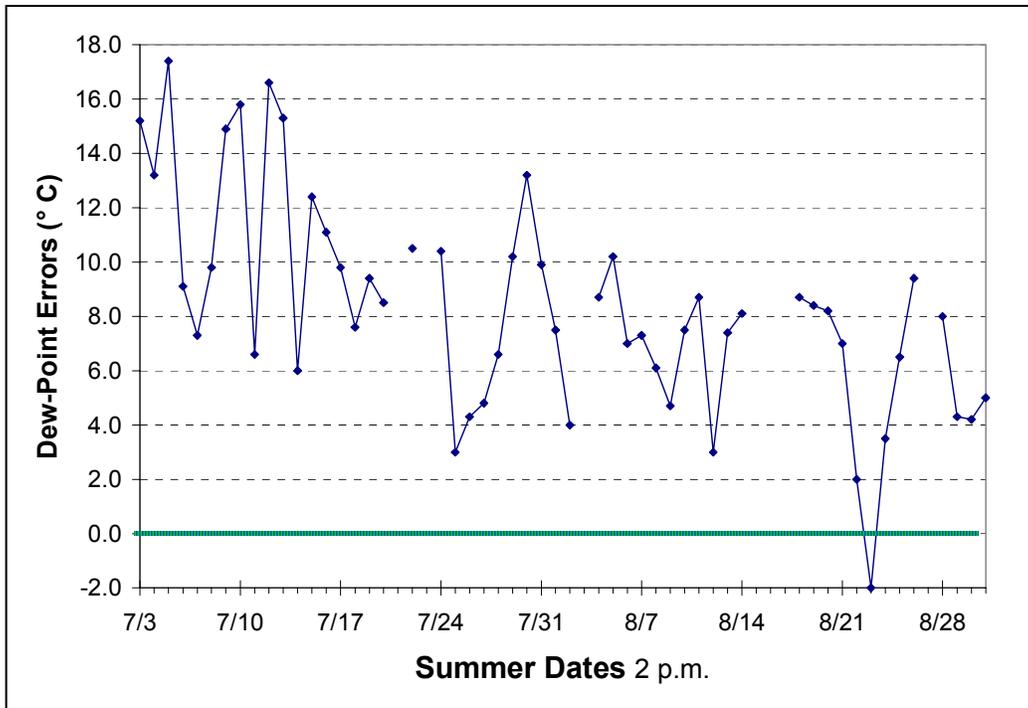


Figure 8. The daily dew-point temperature forecast error amounts for PRP at 2 p.m. in the summer.

The summer wind speed forecasts produced similar mean bias amounts of -1.1 m/s at 2 a.m. and -1.3 m/s at 2 p.m. However, the nighttime forecasts resulted in much more variation, including more frequent high errors in early July and late August and too low forecasts most of the rest of the time. The daytime wind speed forecasts had lower overall error amounts, with forecasts that were consistently too low. Neither the nighttime or daytime forecasts showed any persistence in the amount of absolute error from one day to the next. In fact, for the days with better or worse forecast results, the next day's forecast was more likely to be the opposite category than the same category.

The wind direction at PRP during the summer at 2 a.m. was predominantly out of the north. The WRF wind direction forecasts at these times did not show any bias and were often very accurate, with the wind direction forecasts categorized as average when the error was between 15 and 45°. Unlike the wind speed results, the nighttime wind direction errors reflected the common pattern of infrequent occurrence of the opposite category on days following a better or worse forecast. The observed wind directions were more variable at 2 p.m. The average forecast error category encompassed higher errors, ranging from 45° to 85°. As with the wind speeds, the daytime wind directions in the second day category were as likely or more likely to be the opposite than to be the same as the day before.

The winter temperature forecast errors for PRP were higher than the summer errors for both validation times. The nighttime forecasts exhibited a higher persistence in error category than was seen in the summer months, while the daytime forecasts showed less persistence than in the summer. On the other hand, the dew-point temperature forecast errors were only about half as

large in the winter, when compared to the 11 °C nighttime error and 8 °C daytime error experienced in the summer. The dew-point temperature errors showed some persistence from one day to the next. The nighttime pattern was similar to the summer results, but the daytime persistence was not as strong as seen in the summer. The wind speed errors were slightly more likely to show persistence between consecutive days in the winter than in the summer. The winter wind direction nighttime forecasts repeated the same category more frequently than either of the other categories the following day when the previous day fell in the worse category, but not when the previous day's forecast fell in the better or average categories. As seen in the summer forecasts, the daytime wind direction forecasts were more frequently in the opposite category than in the same category.

The MM5 error patterns are discussed relative to the WRF results described above for PRP:

- Any persistence in the summer MM5 temperature nighttime forecast errors was weak, with the opposite category occurring more frequently than seen in the WRF results. The pattern of persistence was stronger in the daytime forecast, and is similar to the WRF pattern.
- The summer MM5 dew-point temperature mean and absolute errors for both the 3 a.m. and 3 p.m. valid times were approximately 8 °C. Both times reflected patterns showing persistence of error amounts from one day to the next, as seen in the WRF results.
- The summer MM5 wind speed error categories mirrored the WRF findings of opposite category errors occurring more frequently than same category errors on subsequent days.
- The MM5 wind direction nighttime errors were slightly higher than the WRF errors in the summer months. The persistence of the error category was also somewhat less, with only the better category forecast days most likely to be followed by a forecast in the same category. The MM5 wind direction forecasts shared similar daytime error amounts as well as the lack of persistence produced by WRF.
- The bias in the winter MM5 temperature forecasts was generally in the same direction as the WRF forecasts, with similar or somewhat smaller absolute errors for the MM5. Both the MM5 and WRF nighttime results showed the repeat of average and worse forecasts to be more frequent than a different category, but this persistence still occurred in less than 50% of the cases. Better nighttime MM5 forecasts were followed by another better forecast in 7 out of 16 days, compared to 12 out of 17 for the WRF forecasts. The MM5 daytime temperature forecasts were even more likely to repeat an average or worse forecast; however, the better forecast days were followed by an equal distribution of each of the three categories.
- The winter MM5 dew-point temperature errors exhibited no particular persistence pattern compared to a small persistence in the larger WRF forecast errors.

- The winter MM5 wind speed and wind direction forecasts also showed no persistence in the forecast error from one day to the next.

3.3.1.3 Bountiful/F G House (SNZ)

SNZ is the lowest station used in this case study, at a 4760-ft elevation. It is located in the backyard of a house in Bountiful, which is a city of about 40,000 just north of Salt Lake City. As mentioned previously, no analyses of the wind forecast data is included for SNZ.

The summer temperature forecasts at SNZ at 2 a.m. encompassed a wide range of errors, as shown in figure 9. Several instances occurred with two or more consecutive days containing similar errors; however, there were just as many instances where the model produced wildly divergent errors on consecutive days. The temperature forecast bias was more consistently too low for the daytime forecasts, but no significant pattern appeared in the amount of the error from one day to the next.

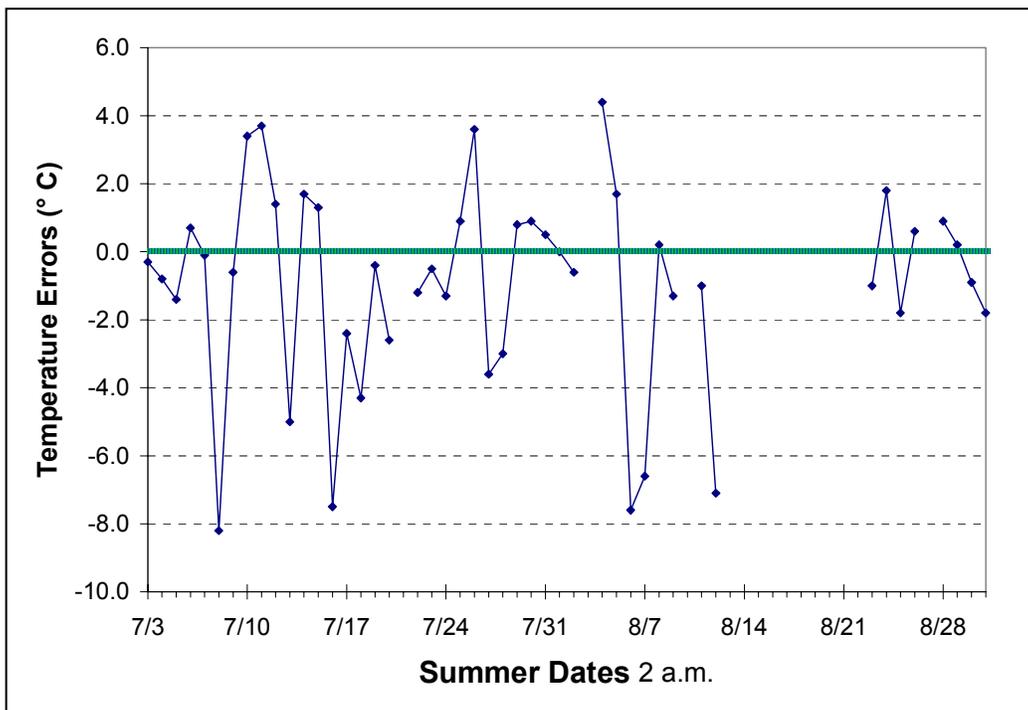


Figure 9. The daily temperature forecast error amounts at SNZ at 2 a.m. in the summer.

The summer dew-point temperature forecasts for SNZ also contained large errors, and frequent sizeable day-to-day changes. However, both the nighttime and daytime dew-point temperature error categories were more likely to repeat the following day than the temperature error categories.

The temperature and dew-point temperature forecasts for SNZ were much more accurate during the winter months. The only category exhibiting a significant persistence trend for temperature

forecasts was the worse forecasts. Each category was slightly more likely to be followed by the same category for the dew-point temperatures.

The MM5 error patterns are discussed relative to the WRF results described above for SNZ:

- The summer MM5 temperature mean errors were similar to those seen in the WRF. The daytime forecasts displayed strong persistence in the error amounts for one five-day period and another four-day period.
- The summer MM5 nighttime dew-point temperature errors were smaller than the equivalent WRF errors, and showed no day-to-day pattern. The afternoon dew-point temperature errors were somewhat higher than the WRF errors, and also reflected the greater likelihood to repeat an error category the following day.
- The winter MM5 temperature forecasts were less accurate than the WRF forecasts at night, but as accurate during the day. Both times shared the WRF pattern of the worse forecasts being most likely to be followed by another worse forecast.
- The winter MM5 dew-point temperature forecast mean and absolute error amounts were very similar to those produced by the WRF. While the WRF forecasts showed persistence in the error categories for these dew-point temperature forecasts, the MM5 results showed no persistence.

3.3.1.4 WBB/U Utah (WBB)

WBB is located in the large urban area of Salt Lake City, slightly higher than SNZ at an elevation of 4900 ft. Whereas the SNZ sensors are on a mast in the backyard of a house, the WBB sensors are on the roof of a University of Utah building.

The summer temperature forecasts at WBB displayed a large bias of -6°C . With few forecasts accurate within 4°C , the error categories incorporated ranges of $1\text{-}5^{\circ}\text{C}$ for the better forecasts, $5\text{-}7^{\circ}\text{C}$ for the average forecasts, and $7\text{-}11^{\circ}\text{C}$ for the worse forecasts. Based on these relatively large ranges, each category was most likely to be associated with the repeat of the same error category the following day, but not by a high percentage. The afternoon temperature forecasts also contained a low bias, but only by about half as much as at night. These forecasts showed no persistence in the forecast error categories.

The WBB summer dew-point temperature forecast errors were even greater than the temperature errors, with the opposite bias of forecasts much higher than observed values. The nighttime results are portrayed in figure 10, showing some of the highest persistence occurrences in this study. This occurrence is primarily due to the change in bias errors from extremely large values in July to smaller values in August, as seen in figure 11. This pattern is similar to the one for the summer dew-point temperature errors at PRP. The summer daytime dew-point temperature forecast errors were smaller than during the night, but the bias and persistence trends were the same as the nighttime results.

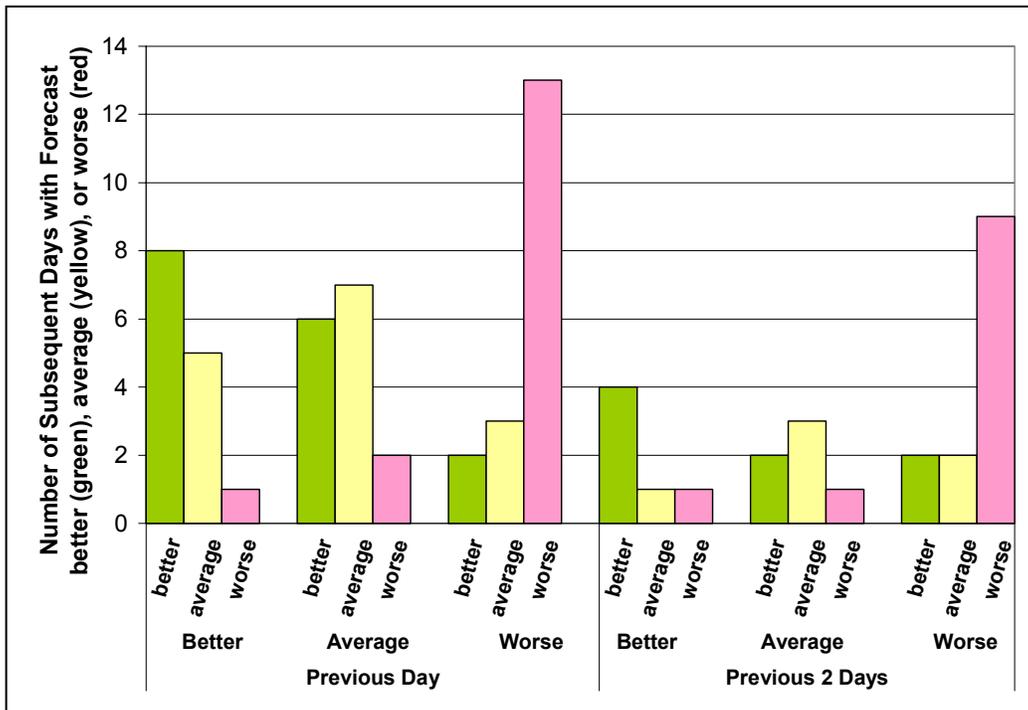


Figure 10. The number of days with the dew-point temperature forecast error category of better (green bars), average (yellow bars), and worse (red bars), given the previous one-day forecast error category on the left and the previous two-day forecast error category on the right, for WBB at 2 a.m. in the summer.

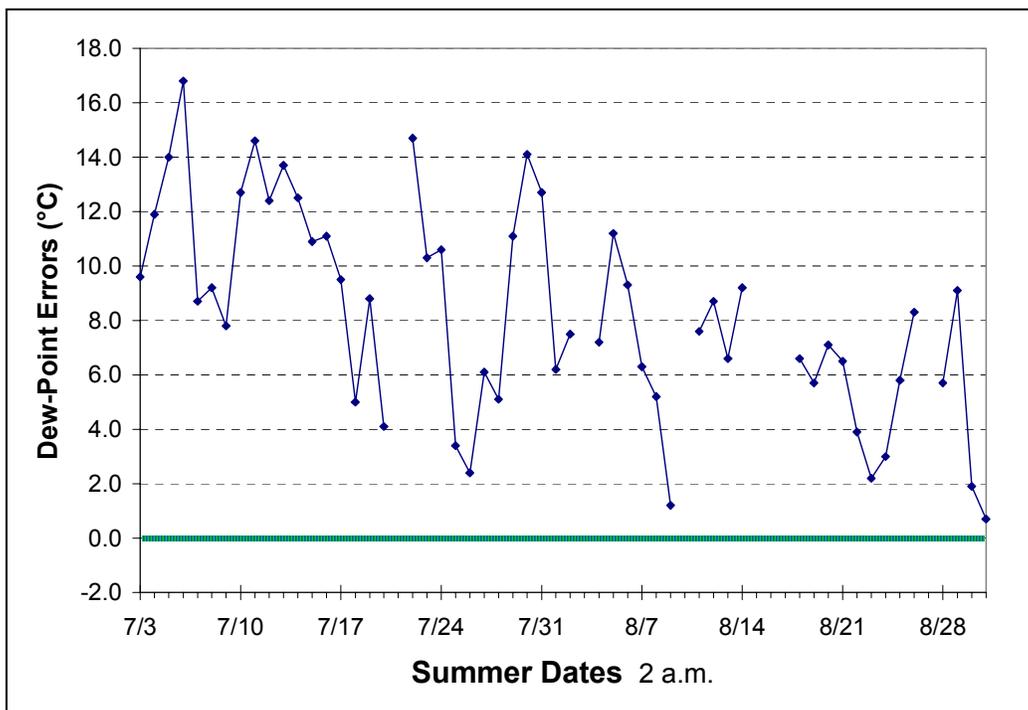


Figure 11. The daily dew-point temperature forecast error amounts for WBB at 2 a.m. in the summer.

The mean absolute error for the summer nighttime wind speed forecasts for WBB is just 2 m/s. However, figure 12 provides the forecast and observed values for each date, highlighting the lack of correlation between these data points. Their Pearson correlation coefficient is zero. With the lack of forecast skill, it is not surprising that the errors do not contain a pattern of persistence. The afternoon wind speed forecasts have a slightly positive correlation coefficient, but the days with better or worse forecast errors were more likely to be followed by the opposite category than the same category.

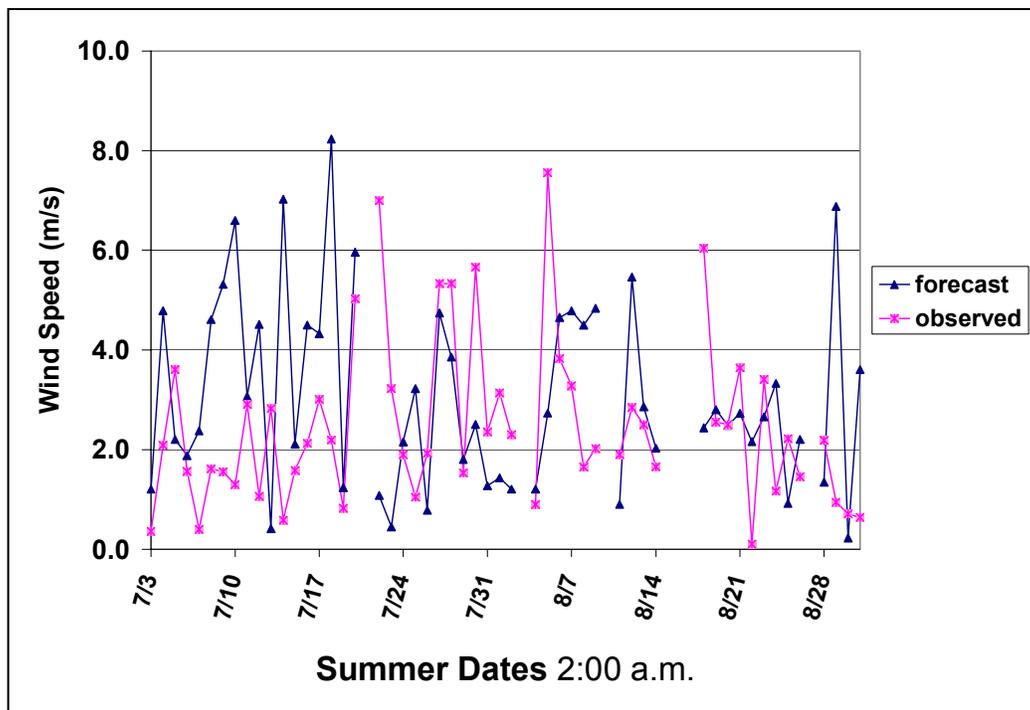


Figure 12. WRF wind speed forecasts compared to the observations at 2 a.m. for the summer months at WBB.

The summer wind direction forecasts had a mean absolute error of about 60° at 2 a.m. and slightly less at 2 p.m. Each time resulted in a slightly positive correlation coefficient. The better and worse categories were only slightly more likely to be followed by the same category the next day.

Winter temperature forecasts at the 2 a.m. valid time contained smaller errors than the ones in the summer. The persistence trend was similar, though, with each category slightly more likely to be followed by the same category the following day. The 4 °C mean absolute temperature error in the winter 2 p.m. forecasts was somewhat higher than in the summer forecasts. These daytime temperature errors showed the most persistence of any sample of the case study. Of the better forecast days, 11 subsequent days were also better, 6 were average, and 0 were worse. The average days were followed by 5 better, 10 average, and 3 worse forecast days. The worse forecast days were associated with 2 better, 2 average, and 14 worse forecasts on the next day. Figure 13 depicts the order of the occurrences of all the forecast categories, starting with day 1 as

January 1, 2004 and going through day 60 as February 29, 2004. The high rate of persistence is associated with a long period of 6-12 °C too high temperature forecasts from January 9 through January 24, followed by many days with forecast errors within 2 °C, and then with primarily 2-4 °C error amounts, including both overforecast and underforecast days, during the last half of February.

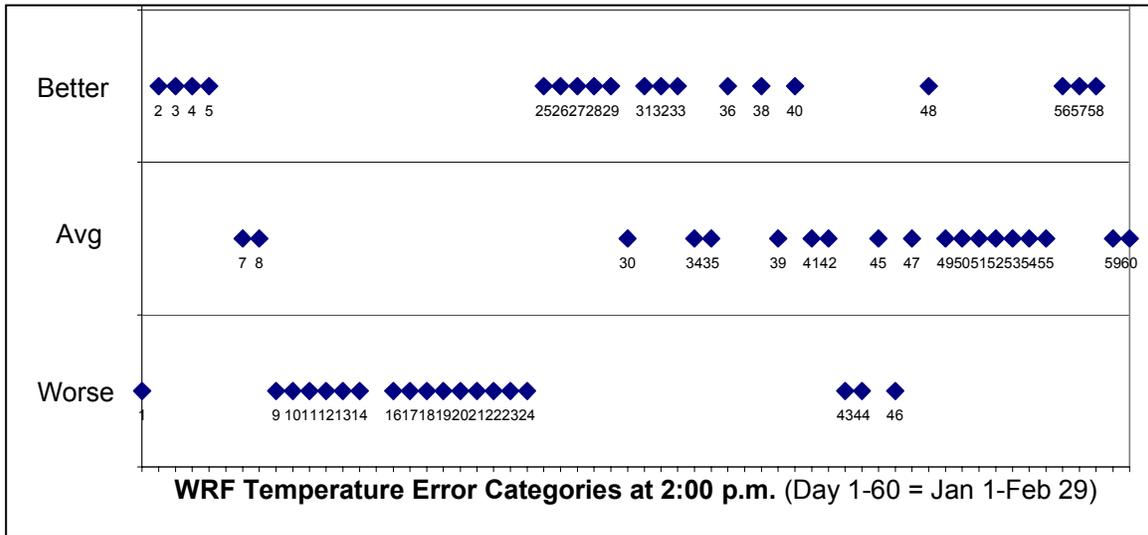


Figure 13. Daily temperature forecast error categories for the daytime winter forecasts for WBB. The marker labels indicate the forecast day in order from 1 through 60, equating to the forecast period of January 1 through February 29, 2004.

The winter dew-point temperature forecast errors were considerably smaller than the errors seen in the summer months. There was a general trend for the high bias to decrease through the two-month period, but many days deviated from the trend. The better and worse categories were slightly more likely to be followed by the same category for both the nighttime and daytime forecasts.

As seen in the summer results, the winter wind speed forecast errors were usually quite small. The first week in January and the second week in February produced several 2 a.m. forecast errors higher than any other period, but they did not occur on consecutive days. The daytime forecasts did not show any persistence pattern.

The winter wind direction forecasts produced mean absolute errors comparable to the summer errors. The wind direction errors do not display any persistence for either the nighttime or daytime.

The MM5 error patterns are discussed relative to the WRF results described above for WBB:

- The MM5 summer temperature forecasts produced larger errors at the nighttime than at the daytime, as was the case for the WRF, but the errors were smaller than those generated by the WRF. Only the worse category was most likely to repeat on consecutive days for the 3 a.m. MM5 forecasts, and any persistence seen at 3 p.m. was very slight.

- The MM5 summer dew-point error day-to-day persistence was not quite as frequent as the WRF dew-point errors, but was still fairly noticeable.
- Summer MM5 wind speed errors were similar to the small WRF errors. Only the worse category in the nighttime forecasts showed any propensity to be followed by the same error category the next day.
- The MM5 summer wind direction forecasts did not include any persistence pattern.
- The MM5 winter temperature forecasts generated a more consistent warm bias than the WRF. The error categories showed some persistence but not to the extent evident in the equivalent WRF results. The MM5 nighttime worse category was repeated on 11 out of 17 days, while better forecasts were most likely to be followed by average forecasts and average forecasts by better forecasts. Each category was most likely to repeat the same category the next day for the MM5 daytime temperature forecasts, but not as frequently as seen in the WRF results.
- The MM5 winter dew-point temperature forecasts showed no persistence in the error categories.
- The MM5 winter wind speed forecast errors were comparable to the generally small errors produced by the WRF. The only discernable pattern in the errors was that there was a greater likelihood for the better and worse categories to be the same or the adjacent category the following day than in the opposite category, particularly for the nighttime forecast.
- The MM5 winter wind direction forecast errors were also similar in magnitude to the WRF errors, and did not display any persistence.

3.3.2 Summary Error Persistence Results by Station

Although the previous section provides many differences in the details of the forecast errors at the four station locations used in this case study, each station produced similar patterns for the error persistence, relative to that station. Nevertheless, the individual sites did produce different forecast error and bias amounts—the categories of better, average, and worse forecasts were determined separately for each seasonal grouping of forecasts for location, weather parameter, time, and forecast model.

Unlike the charts shown in figures 7 and 10, the remaining bar charts (figures 14-18) are based on the percentage of total cases in a given category rather than on the raw numbers of days. Therefore, the axes are equivalent and the charts can be reasonably compared.

Figure 14 shows the results for the DVE, PRP, SNZ, and WBB sites individually, based on the combined results for all weather parameters, both times, both seasons, and both forecast models. For each station, when the previous day had a forecast error in the better category, the following

day had the highest percentage of better forecasts, as opposed to average or worse forecasts. WBB had the lowest percentage of repeating better days (39%), but the difference wasn't substantial when compared to the highest at SNZ (44%). The differences were slightly accentuated when it was required that both of the previous two days were in the better category. In that circumstance, WBB showed a better forecast for the following day 37% of the time, while PRP fell in the better category 47% of the time.

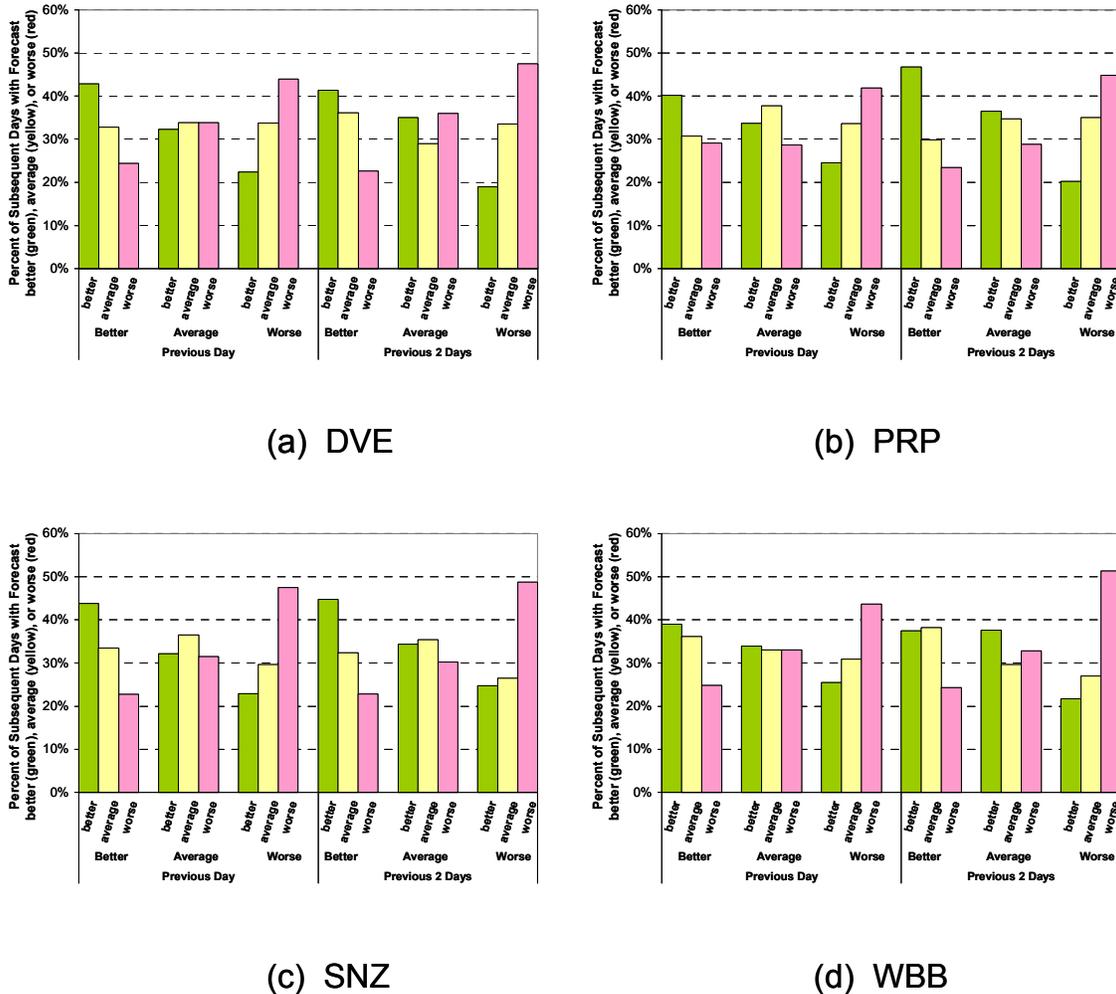


Figure 14. Each chart shows the combined results for all forecasts for stations (a) DVE, (b) PRP, (c) SNZ, and (d) WBB. The y-axis indicates the percentage of subsequent days with the forecast error category of better (green bars), average (yellow bars), and worse (red bars), given the previous day's forecast error category, which is listed on the horizontal labels for the x-axis on the left half of each chart. The right half of each chart provides the same information, except the x-axis groupings are based on having two previous days in the same error category.

The days with forecast errors falling in the worse category were also most likely to be in the same category the next day, and least likely to be in the opposite category. The percentage of repeating days was somewhat higher for the worse category than for the better category, ranging from a low of 42% at PRP to a high of 47% at SNZ. When the two previous days were both in

the worse category, the lowest and highest percentage of subsequent days in the worse category rose to 45% at PRP and 51% at WBB.

Days following an average forecast error were equally likely to produce a forecast error in any of the three categories at DVE and WBB. PRP generated a repeating average error the highest percentage of the time, at only 38%.

3.3.3 Summary Error Persistence Results by Weather Parameter

Figure 15 provides the same information as found in figure 14, except the individual charts total all the data by weather parameter rather than by station.

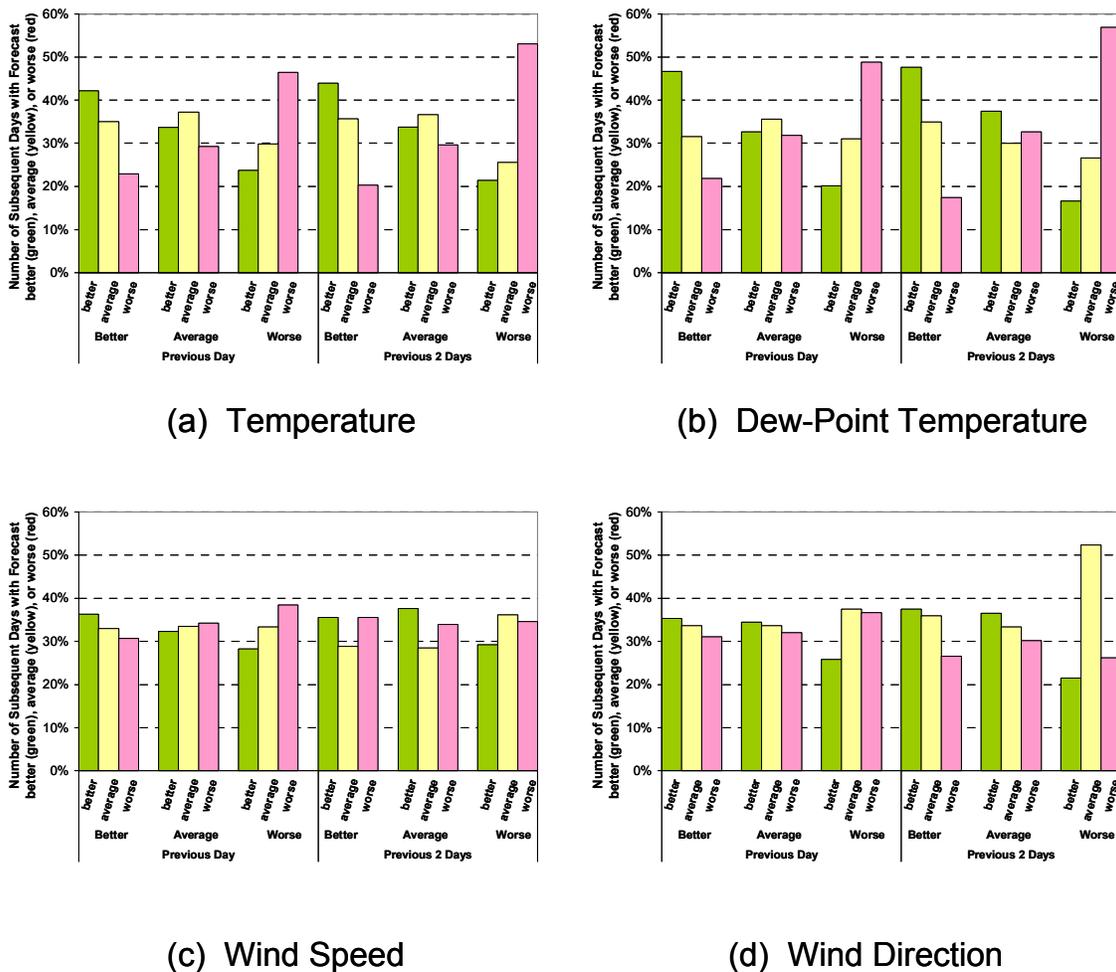


Figure 15. Each chart shows the combined results for all forecasts for weather parameters (a) temperature, (b) dew-point temperature, (c) wind speed, and (d) wind direction. See figure 14 caption for chart details.

The dew-point temperature errors were most likely to repeat the same error category on consecutive days. A better dew-point forecast was followed by a better forecast 47% of the time, an average forecast repeated 36% of the days, and a worse forecast was followed by another worse forecast 49% of the time. Temperature values were slightly lower, with 42% for the better

forecasts, 37% for the average forecasts, and 47% for the worse forecasts. The instances with two preceding days in the better or worse forecast error categories resulted in even higher recurrences of the same category, reaching 57% for the worse dew-point temperature forecasts. The wind speed and wind direction forecasts show less persistence, with the same category repeating the following day 34-38% of the time. With regard to the wind forecast, having two previous days in the same category did not lead to the same category on the third day occurring more frequently than a different category.

3.3.4 Summary Error Persistence Results by Time of Day

As seen in figure 16, results grouped by the nighttime and daytime forecasts produced similar trends, with the better 2-3 a.m. forecasts repeated 40% of the time and the better 2-3 p.m. forecasts repeating 42% of the time. Both the nighttime and daytime forecasts repeated the worse forecast category 44% of the time. For the days with two previous days in the same forecast category, the repetition was 5% higher for the 2-3 a.m. better category and the 2-3 p.m. worse category. No persistence pattern emerges on days with average forecast errors.

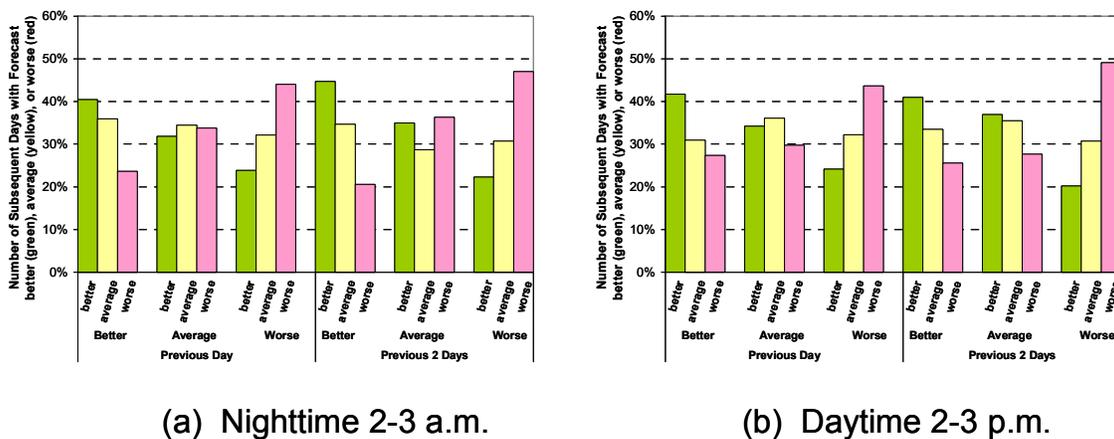


Figure 16. Each chart shows the combined results for (a) the nighttime forecasts and (b) the daytime forecasts. See figure 14 caption for chart details.

3.3.5 Summary Error Persistence Results by Season

The summer forecasts displayed greater error persistence than the winter forecasts, with the better and worse forecasts followed by the same category 44% and 45% of the time, as compared to 39% and 43% for the winter forecasts (figure 17). Days with average forecast errors were followed by another day with an average error 35% of the time for both the summer and winter months. The percentages of repeating better and worse category forecasts were similar for the summer and winter when the two preceding days were the same category, although the probability of an opposite category occurring was still higher in the winter than the summer.

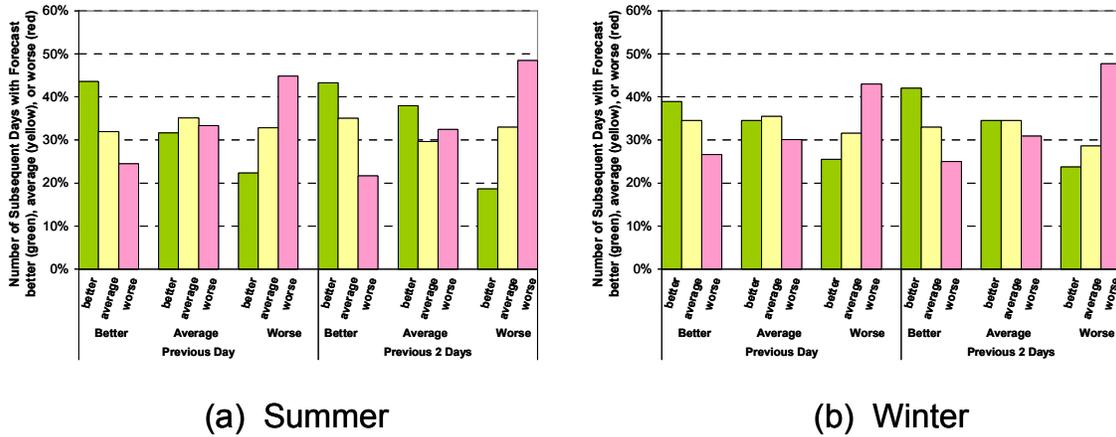


Figure 17. Each chart shows all the forecast results for (a) the summer forecasts and (b) the winter forecasts. See figure 14 caption for chart details.

3.3.6 Summary Error Persistence Results by Model

Although the WRF and MM5 forecasts were for slightly different times, and the forecast errors between the two models were frequently quite different, the weak propensity for the error amount to persist from one day to the next was comparable for both models (figure 18.) A better forecast was followed by a better forecast 41% of the time and by a worse forecast about 25% of the time for both the WRF and MM5. A worse forecast was followed by a worse forecast 45% of the time for the WRF and 43% of the time for the MM5, and by a better forecast 24% of the time for both models. An average forecast was only very slightly more likely to be followed by another average forecast than by either a better or worse forecast. When two consecutive days were the same category of better or worse, the WRF model forecasts were about 5% more likely to repeat that category a third day than the MM5 forecasts.

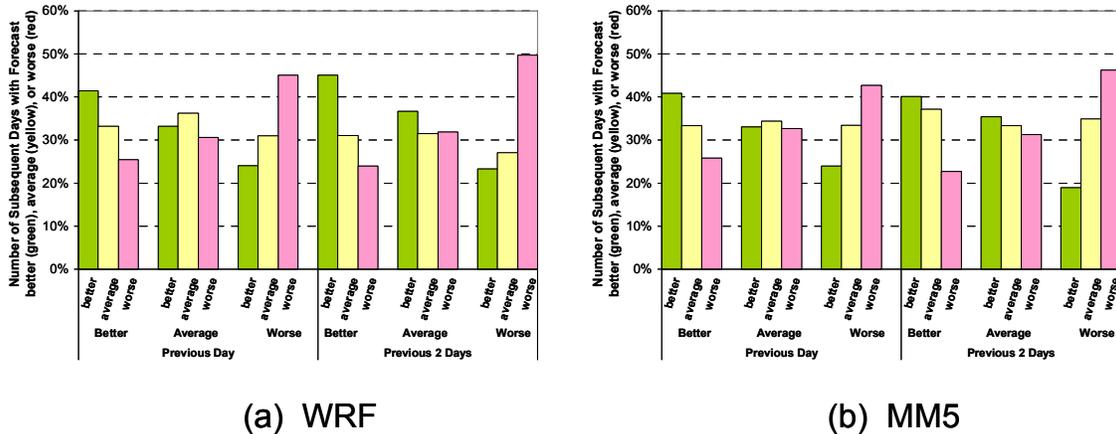


Figure 18. Each chart shows the all the forecast results for (a) the WRF and (b) the MM5. See figure 14 caption for chart details.

4. Conclusions

4.1 Overall Case Study Results

This case study examined the issue of using the quality of the previous one- or two-day numerical weather prediction model forecast to judge the quality of the following day's forecast, in the absence of knowledge of the actual weather situation. Hundreds of forecasts were divided into equal categories of better, average, and worse forecasts within individual combinations of the following:

- station location (four sites in northern Utah)
- weather parameter (temperature, dew-point temperature, wind speed, wind direction)
- time of day (2 a.m. and 2 p.m. for the WRF, 3 a.m. and 3 p.m. for the MM5)
- season (summer: July-August, 2003, and winter: January-February, 2004)
- forecast model (WRF and MM5)

Table 3 provides the summary of all the results combined. The rows indicate the previous day's forecast error category, with each row adding up to 100% of the forecasts for that given previous day's category. The columns labeled subsequent day give the percent of the time the equivalent forecast 24 hours later fell into each of the three categories. Since one-third of the forecasts were in each category, a random result would lead to each subsequent day's forecast category occurring approximately 33% of the time (rounding errors generate a slightly higher percentage in the subsequent day worse column than in the better column.) The cells where the forecast error persisted in the same category from one day to the next are shaded blue. The cells where the forecast error was in the opposite category are shaded peach.

Table 3. Percent of forecasts in each error category 24 hours later, given the previous day's error category.

		SUBSEQUENT DAY		
		better	average	worse
PREVIOUS DAY	better	41 %	33 %	26 %
	average	33 %	35 %	32 %
	worse	24 %	32 %	44 %

Table 4 is based on a subset of the data, considering only the days following occurrences when both the forecasts, 24 hours and 48 hours prior to the validation time, were in the same forecast error category. Since the validation days do not comprise the total dataset, these subsequent day results were not forced to be equally divided among the three categories, although it appears that the categories are evenly represented, with a few percent higher in the worse category than the average category. Again, the cells where the forecast error persisted in the same category from one day to the next are shaded blue. The cells where the forecast error was in the opposite category are shaded peach.

Table 4. Percent of forecasts in each error category on the third day, given a consistent error category over the previous two days.

		SUBSEQUENT DAY		
		better	average	worse
PREVIOUS 2 DAYS	better	43 %	34 %	23 %
	average	36 %	32 %	32 %
	worse	21 %	31 %	48 %

These results do support the theory that a forecast generating an exceptionally large or small absolute error amount will be more likely to be associated with a similar size error the following day. Although these repeating error categories are seen in more than 40% of the better and worse category forecasts, they occur less than half the time overall. Furthermore, they are followed by a forecast in the opposite category approximately one quarter of the time. An average forecast error is almost equally likely to be followed by an error in any of the categories the following day.

As described in section 3.3.1, the results for a specific time and location could deviate substantially from these average results; however, many instances show similar percentages. In addition, the overall averages for each individual station, time of day, season, and model reflect comparable results. The only significant difference is seen in the overall averages based on the forecast weather parameters. The dew-point temperature forecast errors proved to be the most persistent, while the wind forecast errors contained little or no persistence from one day to the next.

4.2 Recommendations

Several organizations are addressing the requirement to provide weather forecast uncertainty information to users. Suggestions have been raised to base this guidance on the previous day's accuracy of the numerical weather prediction model being used, or on a longer period of the

historical statistical results of the model. In general, these ideas have been discarded in favor of relying on ensembles of models to produce a range of forecasts based on the specific conditions for the forecast of interest. Although running ensembles of multiple models or a single model with multiple initial conditions requires more time and/or computational power, this case study supports the conclusion that simply providing the error amount from the previous day or two does not provide sufficient uncertainty information for a user. Several analyses in this case study showed model bias trends lasting for a week or two, with exceptions even within the short timeframe. Calculating and correcting recent model biases should lead to lower forecast errors overall, but a statistical approach based on a single model run does not appear adequate to provide forecast uncertainty information beneficial to decision makers.

References

1. Sauter, B.; Henmi, T. *Average Forecast Errors Using MM5 and WRF Over Complex Terrain: Utah, July/August 2003 and January/February 2004*; ARL-MR-597; U.S. Army Research Laboratory: White Sands Missile Range, NM, October 2004.
2. Grell, G.A.; Dudhia, J.; Stauffer, D.R. *A Description of the Fifth-Generation Penn State/NCAR Mesoscale Model (MM5)*; NCAR/TN-398+STR; NCAR Tech. Note; 1995, 122.
3. Michalakes, J.; Dudhia, J.; Gill, D.; Klemp, J.; Skamarock, W. *Design of a Next-Generation Regional Weather Research and Forecast Model: Towards Teracomputing*; World Scientific: River Edge, NJ, 1998, pp 117–124.
4. Horel, J.; Splitt, M.; Dunn, L.; Pechmann, J.; White, B.; Ciliberti, C.; Lazarus, S.; Slemmer, J.; Zaff, D.; Burks, J. MesoWest: Cooperative Mesonets in the Western United States. *Bulletin of the American Met. Soc.* **2002**, *83*, 211–225.
5. National Center for Environmental Prediction Global Forecast System: The GFS Atmospheric Model Home Page (status as of November 1, 2002). <http://www.emc.ncep.noaa.gov/modelinfo/index.html> (accessed November 2003).
6. Army High Performing Computing Research Center Home Page. <http://www.ahpcrc.org> (accessed December 2004).
7. U.S. Army Research Laboratory Major Shared Resource Center Home Page. <http://www.arl.hpc.mil/> (accessed December 2004).

INTENTIONALLY LEFT BLANK.

Appendix. Average Forecast Error Ranges

This appendix is part of ARL-TR-3418, U.S. Army Research Laboratory, White Sands Missile Range, NM 88002-5501.

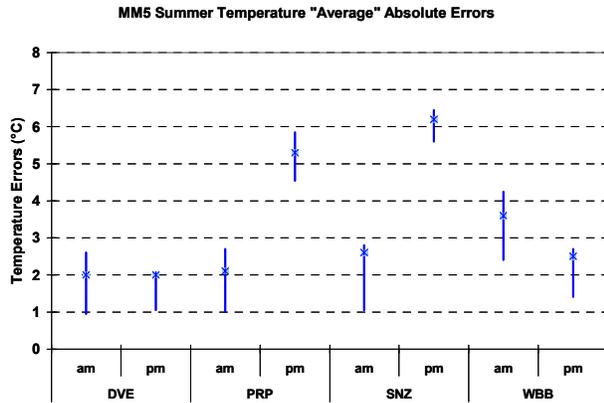
The following charts depict the specific forecast error ranges categorized as average errors in this study. The range is specific for each forecast model (MM5 or WRF), each season (summer or winter), each station (DVE, PRP, SNZ, or WBB), and each forecast time (2 a.m. or 2 p.m. MST for WRF, or 3 a.m. or 3 p.m. MST for MM5, denoted as “am” and “pm” on the charts.)

The chart data is as follows:

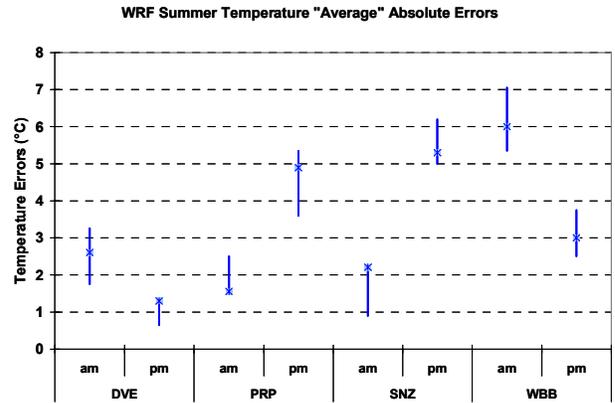
- On each chart, the “high-low” line encompasses the average forecast absolute error values encountered in the middle third of all the forecasts for the given model, season, station, and time.
- Forecast error values less than the lowest value of the “high-low” line comprise the third of the cases classified as better forecasts in this study.
- Similarly, the third of the cases classified as worse forecasts in this study have absolute errors greater than the value at the high end of the line.
- The simple average absolute error for all the forecasts in a category is shown by an asterisk.

Most categories contain approximately 55 forecasts, with about 18 forecasts in each classification of better, average, and worse. Wind direction results are frequently based on a smaller number of forecasts, since wind direction errors were not included for any forecast with an observed wind speed less than 1 m/s. Any category without at least 10 forecasts in each classification was not used and will not have a line or asterisk plotted.

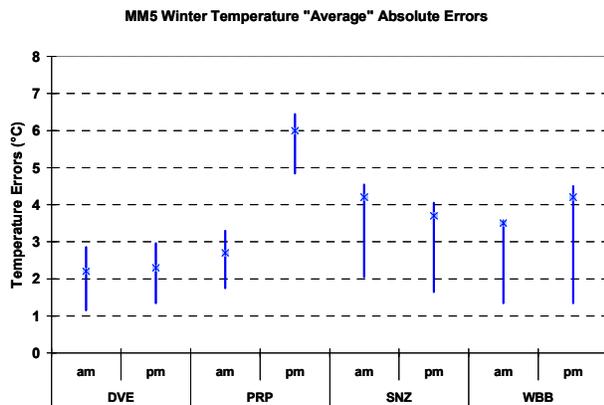
Temperature, dew-point temperature, wind speed, and wind direction error plots are shown in figures A-1 through A-4, respectively.



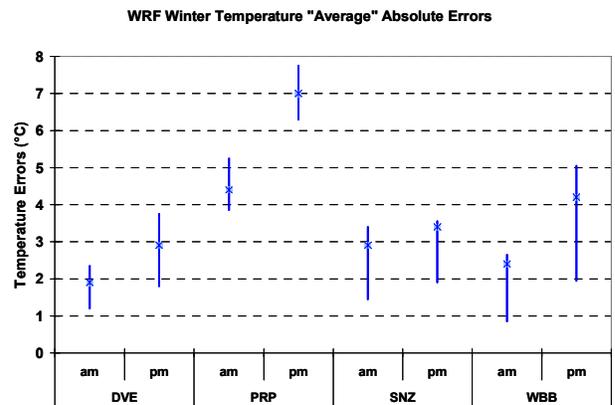
(a)



(b)

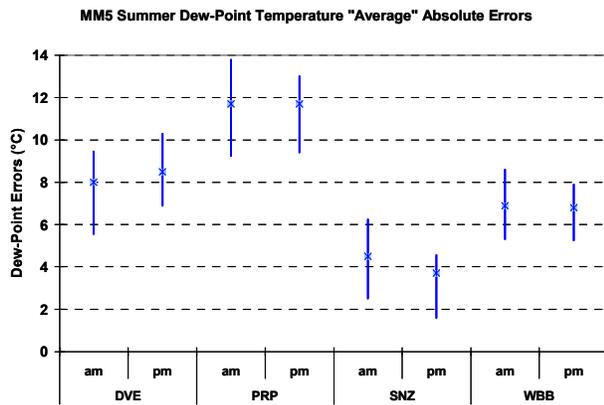


(c)

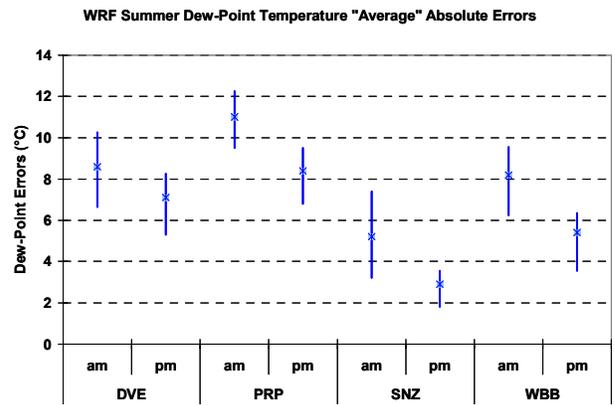


(d)

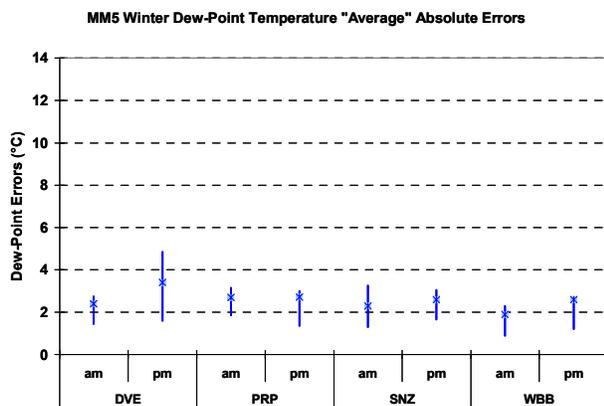
Figure A-1. The range of absolute temperature errors used for the average error category is shown by the line for (a) MM5 summer forecasts, (b) WRF summer forecasts, (c) MM5 winter forecasts, and (d) WRF winter forecasts. One-third of the forecasts in each category had errors greater than the highest value and one-third had errors less than the lowest value depicted by the line. The asterisk associated with each line is the average absolute error for the given station and time.



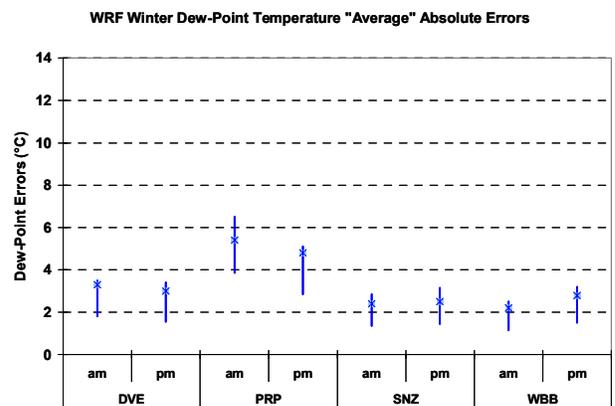
(a)



(b)

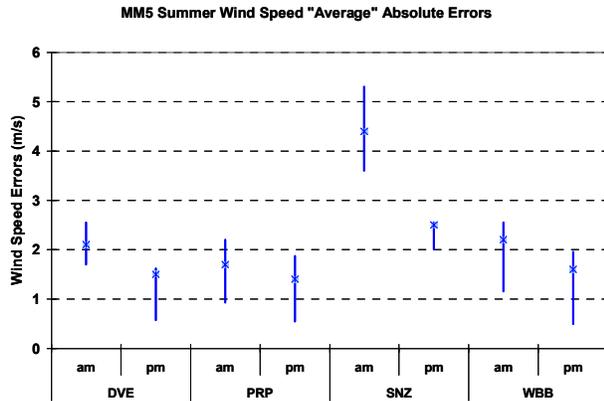


(c)

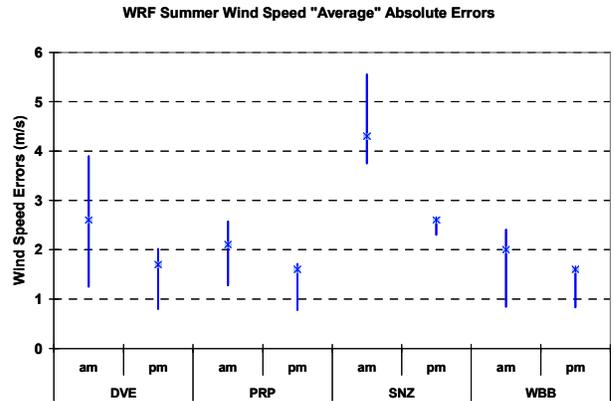


(d)

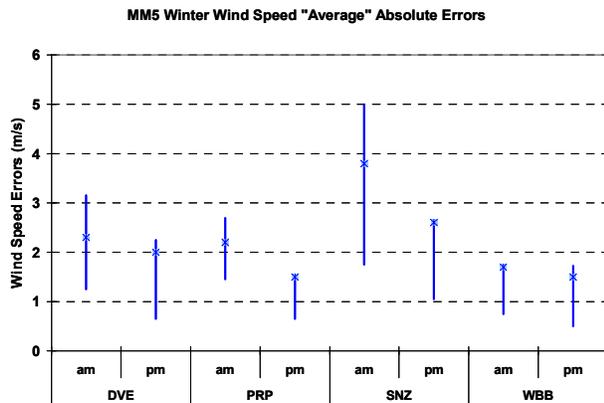
Figure A-2. The range of absolute dew-point temperature errors used for the average error category is shown by the line for (a) MM5 summer forecasts, (b) WRF summer forecasts, (c) MM5 winter forecasts, and (d) WRF winter forecasts. One-third of the forecasts in each category had errors greater than the highest value and one-third had errors less than the lowest value depicted by the line. The asterisk associated with each line is the average absolute error for the given station and time.



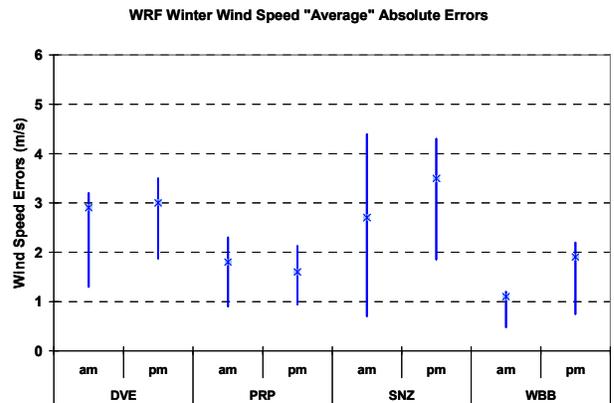
(a)



(b)

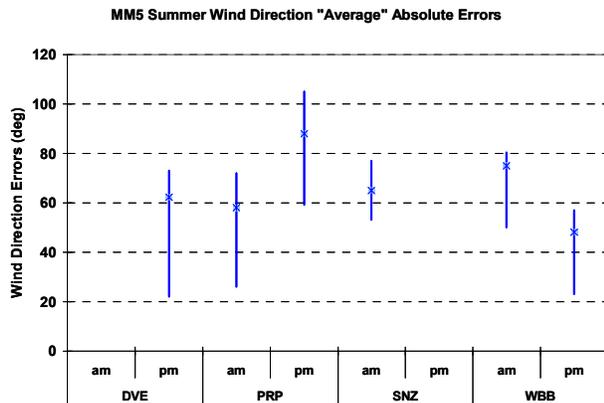


(c)

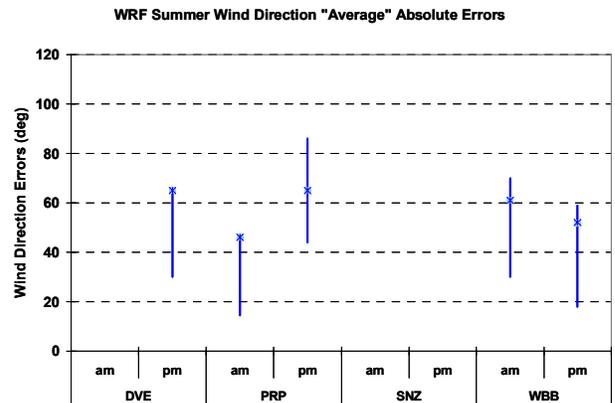


(d)

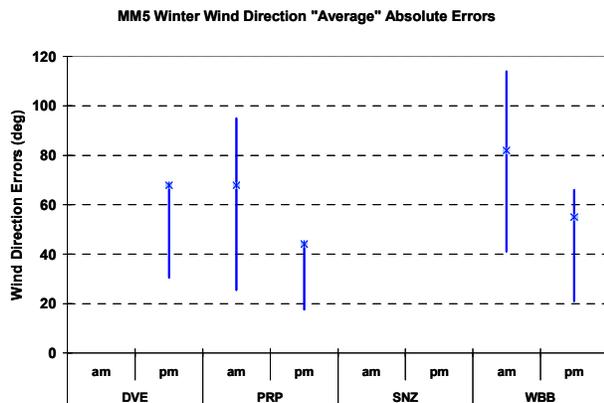
Figure A-3. The range of absolute wind speed errors used for the average error category is shown by the line for (a) MM5 summer forecasts, (b) WRF summer forecasts, (c) MM5 winter forecasts, and (d) WRF winter forecasts. One-third of the forecasts in each category had errors greater than the highest value and one-third had errors less than the lowest value depicted by the line. The asterisk associated with each line is the average absolute error for the given station and time.



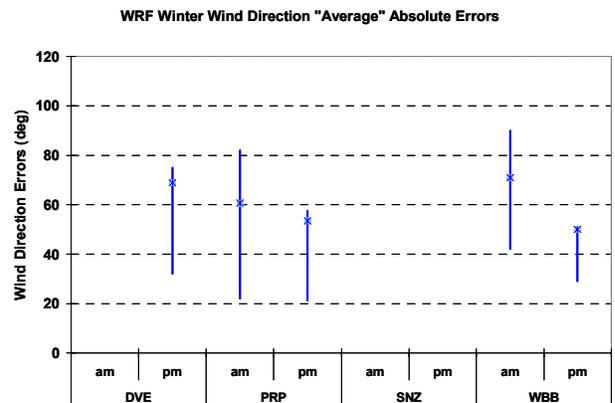
(a)



(b)



(c)



(d)

Figure A-4. The range of absolute wind direction errors used for the average error category is shown by the line for (a) MM5 summer forecasts, (b) WRF summer forecasts, (c) MM5 winter forecasts, and (d) WRF winter forecasts. One-third of the forecasts in each category had errors greater than the highest value and one-third had errors less than the lowest value depicted by the line. The asterisk associated with each line is the average absolute error for the given station and time.

INTENTIONALLY LEFT BLANK.

Acronyms

DVE	Meteorological station identification letters for DV-Burns, UT
GFS	Global Forecast System
GMT	Greenwich Mean Time
MM5	Pennsylvania State University/NCAR Mesoscale Model Version 5
MST	Mountain Standard Time
NCAR	National Center for Atmospheric Research
PRP	Meteorological station identification letters for Promontory Point, UT
SNZ	Meteorological station identification letters for Bountiful/F G House, UT
WBB	Meteorological station identification letters for WBB/U Utah, UT
WRF	Weather Research and Forecasting

Distribution List

	Copies		Copies
HEADQUARTERS DEPT OF ARMY DAMI POB WEATHER TEAM 1000 ARMY PENTAGON ROOM 2E383 WASHINGTON DC 20310-1067	1	HQ AFWA/DNX 106 PEACEKEEPER DR STE 2N3 OFFUTT AFB NE 68113-4039	1
US ARMY RESEARCH LAB AMSRD ARL D 2800 POWDER MILL ROAD ADELPHI MD 20783	1	AFRL TECH LIB 29 RANDOLPH RD HANSCOM AIR FORCE BASE, MA 01731-3010	1
US ARMY RESEARCH LAB AMSRD ARL CI E BATTLEFIELD ENVIR DIV ADELPHI MD 20783-1197	1	AFCCC/DOR TECH LIB 151 PATTON AVE., ROM 120 ASHEVILLE, NC 28801-5002	1
US ARMY RESEARCH LAB AMSRD ARL CI EM BATTLEFIELD ENVIR DIV WSMR NM 88002-5501	1	NAVAL RESEARCH LAB MARINE METEOROLOGY DIVISION 7 GRACE HOPPER AVENUE STOP 2 MONTEREY, CA 93943-5502	1
US ARMY RSRC OFC ATTN AMSRD ARL RO E DR BACH PO BOX 12211 RTP NC 27009	1	NCAR LIBRARY SERIALS NATL CTR FOR ATMOS RSCH PO BOX 3000 BOULDER CO 80307-3000	1
US ARMY DUGWAY PROVING GROUND STEDP MT DUGWAY UT 84022	1	US ARMY RESEARCH LAB AMSRD ARL CI EM B SAUTER WSMR NM 88002	3 hard copies
US ARMY FIELD ARTILLERY SCHOOL ATSF TSM TA FT SILL OK 73503-5600	1	US ARMY RESEARCH LAB ATTN IMNE AD IM DR MAIL & RECORDS MGMT ADELPHI MD 20783-1197	1
US ARMY TOPO ENGR CTR CETEC ZC 1 FT BELVOIR VA 22060-5546	1	ADMNSTR DEFNS TECHL INFO CTR ATTN DTIC OCP ELECT CPY V MADDOX 8725 JOHN J KINGMAN RD STE 0944 FT BELVOIR VA 22060-6218	1
WSMR TECH LIBRARY BR STEWIS IM IT WSMR NM 88002	1	US ARMY RESEARCH LAB AMSRD ARL CI OK TL 2800 POWDER MILL ROAD ADELPHI MD 20783-1197	2
		TOTAL	21