THE VERIFICATION OF WEATHER FORECASTS COMPARED TO NUMERICAL MODEL GUIDANCE, CLIMATOLOGY, HUMAN FORECASTS AND SEASONAL VARIATIONS

A Thesis presented to the Faculty of the Graduate School
University of Missouri-Columbia

In Partial Fulfillment
Of the Requirements for the Degree
Master of Science

---

by
ERIC MICHAEL ALDRICH

Dr. Anthony Lupo, Thesis Supervisor

December, 2011

The undersigned, appointed by the Dean of the Graduate School,
have examined the thesis entitled:

THE VERIFICATION OF WEATHER FORECASTS COMPARED TO NUMERICAL MODEL GUIDANCE, CLIMATOLOGY, HUMAN FORECASTS AND SEASONAL VARIATIONS

Presented by: Eric Michael Aldrich
A candidate for the degree of Master of Science

And hereby certify that in their opinion it is worthy of acceptance.

_____
Professor Dr. Anthony R. Lupo

_____
Associate Professor Dr. Patrick S. Market

_____
Assistant Professor Dr. Patrick E. Guinan

_____
Associate Professor Dr. Stacey W. Woelfel

ACKNOWLEDGEMENTS

I would like to take this opportunity to thank Dr. Anthony Lupo for encouraging me to pursue an advanced degree in Atmospheric Science. Without his poking, prodding and encouragement, I wouldn't have even thought about going back to school to earn my Master's degree in Atmospheric Science! Dr. Lupo is and has been a great mentor and his constant guidance and advice means more to me than words could ever express. I would also like to thank Dr. Patrick Market for his advice, support and time. Dr. Market is a great synoptician and "forecasting guru" and I appreciate his feedback on this paper and research project. Also, I'm grateful for Dr. Pat Guinan's feedback and advice – from a state climatologist perspective. Dr. Guinan has provided useful nuggets of information and data throughout the years. Finally, I'd like to formally thank Dr. Stacey Woelfel, my supervisor at KOMU-TV. Stacey, thank you for understanding what it means to be a full-time employee and a student all at the same time. Thank you also for sitting on my thesis committee!

I would also like to thank the staff at the National Weather Service Forecast Office in St. Louis (NWS-STL) for their help in obtaining missing data, records and model data support. Mark Britt also showed great enthusiasm for my research and has invited me to present my findings to his office. Thanks for this opportunity, Mark!

Josh Kastman from the Soils, Environmental & Atmospheric Science Department (SEAS) provided data entry support and helped organize the data in Microsoft Excel. His ideas, encouragement and knowledge of macros and Excel was extremely important and instrumental when it came time for data analysis and entry. Thanks, Josh!

John Taylor, a great friend and statistical guru, helped with the statistical analysis and assisted in the creation of graphs, charts and statistical data for the research project. He was able to explain things in a very "down to earth" approach and also provided some good quality control looking over the data.

My research partner, Michelle Bogowith – thanks! That seems so inadequate when one considers all the hours we spent crunching data, analyzing data, creating graphs, charts, writing things down, getting lost in thousands and thousands of numbers, equations, charts, etc. Michelle and I got together from the get-go and decided that we were in this together and two years later, we've finally made it. Thanks for being an incredible help, encourager and co-worker! Your ideas and constant nagging was what it took to get this project underway and finally completed. It's been a pleasure working alongside you at KOMU-TV and in the classroom. You deserve this degree just as much as I do – I'm so glad I've been able to share the same path and achieve the same goals - Good job, friend!

My friends and family – thank you SO much for the incredible love and encouragement you have given me during the last two years. Without positive role models and words of encouragement, there is NO WAY I'd be able to finish this research project and degree program! Many hours spent studying, working, typing, looking at data – tons of hours where I could've been playing, talking, and relaxing. You've been a big help and I love each and every one of you!

Finally, my loving wife and best friend, Charlie. Wow! You've been by my side from day one and have never backed down from your encouragement, and understanding. I truly know what it means when someone says "...behind a good man there is a great woman!" With every day that passed, your questions, comments, advice – EVERYTHING – helped me get to this point. Through moments of fear, doubt and low spirits, you were there to build me up, give me a

hug and set me back on track. Without your support and constant love, this would not have been possible. Thanks! I love you more than you'll ever know!

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

THE VERIFICATION OF WEATHER FORECASTS COMPARED TO NUMERICAL MODEL GUIDANCE, CLIMATOLOGY, HUMAN FORECASTS AND SEASONAL VARIATIONS

Eric M. Aldrich

Dr. Anthony R. Lupo, Thesis Supervisor

ABSTRACT

The accuracy of weather forecasts has been a topic of discussion for many years, and a simple but useful method for showing forecast accuracy has been an issue for meteorologists over time. My study on the comparison of meteorological computer forecast models coupled with climatology data and a human produced forecast (from meteorologists at KOMU-TV, Columbia, MO) produced some interesting results. Not surprisingly, the computer models and human produced forecasts were, by far, more accurate than the climatology forecasts, but when I compared all of the data and results with the seasonality variance of temperature data, the outcome was more interesting. Essentially, all forecasts and outcomes had distinct seasonal variability. Some of the more surprising findings were that in seasons with large variability, the computer modeling did a decent job at producing an accurate forecast. It is my hope that the findings in this research will instill more confidence in a human produced forecast and forecasters can determine which computer model has a better handle on seasonal variance based on the model physics.

INTRODUCTION

1.1 – BACKGROUND ON WEATHER FORECASTING

The science and methodology of weather forecasting has seen substantial improvement over the last three centuries in part due to increased knowledge of the physics that govern atmospheric phenomena. One of our country's legendary pioneers, Benjamin Franklin added significantly to our understanding of atmospheric physics. He was said to have been the "Father of Weather" in America (Franklin, 1774; Miller, 1933). His experiments and investigations fueled by his desire to understand the atmosphere and its processes intrigued many. He was able to take simple measurements of common atmospheric variables (temperature, pressure, wind speed and direction, precipitation types and amounts), and, using these measurements, he was able to record the passage of specific weather events and even hypothesized that later in time the same weather events would impact areas to the east of his location. The knowledge gained from his research became a foundation stone for today's understanding of meteorology and the ability to forecast changes in the weather.

As time moved forward, so did scientific research into Earth's atmosphere - how and why it worked were the new questions of the day, and public as well as scientists needed answers. Even as recently as World War I, weather forecasting was unreliable and extremely elementary. Day to day weather observations were limited and, over the oceans, almost non-existent. Forecasting in the beginning of the 20th century was more guesswork and luck than science; basing many

forecasts on climatology and an overall "hunch." Even the concept of frontogenesis and cyclone development hadn't yet been explained, when a Norwegian scientist by the name of Vilhelm Bjerknes devised a methodology for forecasting atmospheric events. In his methodology, he spoke about the atmosphere in terms of a "diagnostic step" and a "prognostic step" (Bjerkness, 1914). The diagnostic procedure required correct and thorough observational data of the atmosphere. Once a clear picture of the atmosphere was obtained, a prognostication and prediction of how the current state of the atmosphere will change over time and space could be developed.

It wasn't until the advent of computers that the forecast process was expedited and computer modeling was implemented into the forecast process (before this time, atmospheric equations were solved by hand and the forecast was outdated before it could be issued). Today, scientists can model the atmosphere and make an accurate prediction (more or less) as to how the atmosphere will behave – based on the aforementioned variables, along with the laws that govern the atmosphere: Laws of Motion and the Laws of Conservation.

The science behind weather prediction has evolved greatly since the days of Benjamin Franklin. Researching how the weather changes with each season and how different forecasting techniques handle each of the seven meteorological variables has contributed to the overall accuracy of weather prediction. In the end, it's all about accuracy. Today's society practically demands it, and bases their activities and lifestyles around an accurate forecast. The results of this research will show how seasonality plays a role in the forecast process in central

Missouri, and how temperature forecasts at KOMU-TV compare to the computer model guidance at their disposal.

ANALYSIS & METHODS

2.1 – THE FORECAST

This research project and overall results were based on original forecasts developed by atmospheric science students and staff meteorologists at KOMU-TV in Columbia, Missouri. The data compiled and used in this research were delivered on air as actual weather forecasts presented to the public. The range covers three years worth of data, from November, 2007 through November, 2010. The forecast data collected consisted of both morning and afternoon/evening forecasts and included high and low temperature predictions for the following day and nighttime forecast periods (basically, a 24-hour time period). Sky cover, wind speed, wind direction, and precipitation data were not archived at the time of data acquisition, thus, are not included in this research project.

2.2 – MODEL FORECASTS

Numerical forecasting products (or Model Output Statistics, hereafter referred to as MOS) were obtained on the day used to produce the KOMU forecast and was archived over a period of three years (November, 2007 through November, 2010). Forecast highs and lows were obtained from two computer models and were analyzed for the 00Z and 12Z model runs. The numerical guidance used in this research project is consistent with meteorological modeling output used to produce weather forecasts worldwide, and is widely accepted by professional

meteorologists in all facets of the weather industry. All MOS products were gathered using the following website:

http://www.nws.noaa.gov/cgi-bin/mos/getall.pl?sta=KCOU

MOS provides a statistical relationship between observed weather variables and predicted weather elements by a dynamic model. MOS has been used in operational meteorology since its inception by Glahn and Lowry (1972) at the National Weather Service (NWS) Techniques Development Laboratory (now the Meteorological Development Laboratory), and has been used by NWS forecasters and other meteorologists for over 30 years. MOS has several advantages over forecasts produced directly by dynamic models. In particular MOS corrects for systematic model biases, partially accounts for some phase errors, predicts sensible weather elements not directly forecast by current dynamic models (e.g., visibility), produces reliable probability forecasts (e.g., probability of precipitation), and provides information on model predictability (Carter et al.1989). One disadvantage of MOS is that changes to the dynamic model, which are common with today's modeling systems, can alter model biases, reducing the accuracy of MOS equations that were developed with forecasts from the older model configuration (Dallavalle 1998). Humans are limited in their capacity to absorb atmospheric data in a certain amount of time. In turn, they are subject to a different sort of sampling error when compared to the computer models. However, humans can use their knowledge of how the atmosphere works and use pattern recognition to identify observations that computer models won't recognize or incorporate into the forecast output (Doswell, 2004).

The two computer models used in the study were the NAM (North American Mesoscale Model) and the GFS (Global Forecasting System). The NAM and GFS computer models are the ones most readily available and acceptable for use in creating a local weather forecast for public use. Both computer models used have limitations and certain biases that will result in greater/less forecast accuracy. The specifics and intricacies of the NAM and GFS models are listed in Table 1.

North American Mesoscale Model (NAM)

– Based on the "WRF-NMM"
– Short-range (out to 84 hours) forecasts for North America every 6 hours
– Grid spacing of ~12 km
–Available on lower-resolution grids
– Strengths relative to other numerical models
  -Terrain representation, mesoscale detail

– Weaknesses relative to other numerical models
  -Limited area, large-scale accuracy

Global Forecast System (GFS)

– Medium range (out to 384 hours) forecasts for global analyses
– Grid spacing of ~25 km to 192 hours and ~70km thereafter
– Available on lower-resolution grids

– Strengths relative to other numerical models
  -Accuracy of large-scale forecast

– Weaknesses relative to other numerical models
  -Terrain representation
  -Precip structure

Table 1.

(Obtained from Dr. Jim Steenburgh)
http://www.inscc.utah.edu/~steenburgh/classes/5010/lecture_notes/Forecast-Techniques-and-Tools.pdf

Because our understanding of the atmosphere is imperfect, it is natural to believe that even if we somehow had obtained perfect observations of the initial state, our predicted evolution of the initial state would fail to be perfect as a

consequence of an imperfect prognostic model (Doswell, 2004). Doswell brings up an excellent point here, in that, you can have all of the computer models in the world set to update and create a forecast for your specific area of interest, and while the computer will most definitely produce a forecast, the outcome will only be as accurate as the initial conditions that are put into the model.

2.3 – CLIMATOLOGY

We needed an "accepted base" for our research project and something to compare our results with (especially when comparing our accuracies to the various seasons). We used the 30-year daily averages from 1971-2000 (otherwise known as climatology) of observed highs and lows for Columbia, Missouri. The range of climatology data included the days from within November, 2007 through November, 2010. This data was provided in part by the National Weather Service Forecast Office in St. Louis, Missouri (NWSFO-STL), as well as from the National Climatic Data Center (NCDC). Additionally, this data can be obtained online at:
http://hurricane.ncdc.noaa.gov/climatenormals/clim20/mo/231791.pdf

2.4 – PROCEDURES

The data analysis and procedure for the skill score portion of this thesis project was carried out using Microsoft Excel. The data was broken into four categories: Created forecast, Observed Highs/Lows, MOS forecast, and Climatologically averaged forecasts. The MOS forecast was broken into NAM and GFS sub

categories. In order to classify the data based on accuracy, a temperature score formula was created in which we took the forecast temperature (F) and found the difference between the observed temperature (O). Further breakdowns show that a 0, 1, or 2 point score was given here. The forecast was considered "perfect" if the difference between the forecast temperature and the observed temperature was between zero and two (meaning a near/perfect temperature forecast had occurred). A temperature score of two was assigned at this point. If the temperature difference was between three and four degrees, a temperature score of one was assigned and was said to be a "good" forecast. Any temperature difference that was greater than four was a "bad" forecast and was therefore awarded a score of zero. The averages were created by removing the "special case" scores and taking the absolute values of the scores. This was done because negative numbers represent the same value as positive values on the success scale. Therefore the numbers were changed to all positives for analysis purposes to achieve a non zero (or less than zero) score. There were a few instances where midnight high temperatures occurred, or daytime low temperatures occurred, and those results were classified as "special cases." These cases were reviewed individually and were either included into the average score or omitted from the study entirely. An example of how these data was calculated can be seen in Table 2. In order to compute forecast skill, a skill score formula was created and was performed on all of the data sets. The skill score formula used for the analysis was adopted from Lupo and Market (2002).

$$S = \left(\frac{F - B}{P - B} \right) \times 100$$

Where *S* is the skill score number, *F* is the forecast temperature, *B* is the baseline (or the observed temperature) and *P* represents a perfect forecast (in this study, 2). This equation takes the forecasted temperature (*F*) and subtracts from it the observed baseline (*B*) dividing by the difference between a perfect forecast score (*P*) which is equal to 2.

Date	High KOMU	Low KOMU	High Observed	Low Observed	High Score	Low Score
12/21/2007	58	42	58	41	2.00	2.00
12/24/2007	38	25	40	22	2.00	1.00
12/31/2007	25	17	26	11	2.00	0.00

Table 2.

Table 2 is an example of what a portion of the MS Excel spreadsheet looked like that contained the data in this research. In the example, KOMU produced forecasts were compared to the observed values and are listed with the associated temperature scores. The difference between the KOMU high temperature forecast and the observed high for 12/21/2007 is 0. Therefore, the temperature score was rated a 2, meaning the forecast was perfect. The temperature score for low temperatures on 12/31/2007 is a zero because the difference between the low temperature forecast and the observed low temperature is a 6. The temperature score is a zero if the difference between forecast and observed is greater than or equal to 4.

Once temperature and skill score numbers were calculated, we organized all of the data into individual yearly seasons (i.e., Spring, 2009, Summer, 2009, Fall, 2009, etc.) and overall seasons (i.e., Spring 2007-2010, Summer 2007-2010, Fall 2007-2010, etc.). The results were ranked in terms of accuracy, and in which seasons those accuracies occurred.

2.5 – BREAKDOWN OF SEASONS

All seasons researched in this study were classified by meteorological seasons instead of the more common astronomical seasons that the general public is most familiar with. December, January and February, March, April and May, June, July and August, and September, October and November were classified as Winter, Spring, Summer and Fall, respectively (Table 3.)

Spring	**Summer**	**Fall**	**Winter**
March	June	September	December
April	July	October	January
May	August	November	February

Table 3.

Table 3 represents the meteorological seasons and the breakdown pattern that was used in this research study. These are different from the normally accepted astronomical seasons.

RESULTS

3.1 - THE RESULTS

There were a few surprising outcomes related to the overall results of this research project. Going into the research, we knew that climatology would fall behind in the accuracy column from the beginning. When we graphed our results and then compared climatology to the human forecast, model forecasts and the observed values, climatology had the lowest ranking. One of the other interesting results was to see how accurate the NAM model was at forecasting daily high and low temperatures. The overall results and breakdown of how everyone fared within the meteorological seasons will be discussed in this chapter.

3.2 – CLIMATOLOGY RESULTS

Climatology is the description of aggregate weather conditions - the sum of all statistical weather information that helps describe a place or region (Lutgens, Tarbuck, 1995). Or, more simply, a 30-year average of high and low temperature data for a certain location. This is used to quickly get an idea as to how the weather *typically* behaves over a certain location and time. Table 4 shows the 30-year seasonal average highs and lows for Columbia.

1971-2000	WINTER				SPRING			
YEAR	**DEC**	**JAN**	**FEB**	**AVG**	**MAR**	**APR**	**MAY**	**AVG**
AVG	32.0	27.8	33.7	31.2	44.0	54.4	63.7	54.0
MAX	41.5	37.4	43.9	40.9	55.1	65.9	74.6	65.2
MIN	22.5	18.2	23.4	21.3	33.0	42.9	52.8	42.9
	SUMMER				**FALL**			
	JUN	**JUL**	**AUG**	**AVG**	**SEP**	**OCT**	**NOV**	**AVG**
AVG	72.7	77.4	75.7	75.2	67.3	56.0	43.2	55.5
MAX	83.6	88.6	87.3	86.5	79.1	68.0	53.4	66.8
MIN	61.8	66.3	64.0	64.0	55.4	44.1	33.0	44.1

Table 4.

Table 4 shows the 30 year climatology data for Columbia, Missouri from 1971-2000. The data is broken down into the meteorological seasons and includes seasonal averages. This data was obtained from the National Climatic Data Center (NCDC).

The monthly and seasonal averages for the study period (2007-2010) are listed in Table 5. These data can be compared to the 30 year climatology data listed in Table 4 for a monthly/seasonal comparison, along with standard deviations.

2007/2008	WINTER				SPRING			
YEAR	**DEC**	**JAN**	**FEB**	**AVG**	**MAR**	**APR**	**MAY**	**AVG**
AVG	33.3	31.4	30.9	31.9	43.8	53.2	62.6	53.2
MAX	40.7	42.1	39.0	40.6	54.8	63.7	72.6	63.7
MIN	25.8	20.6	22.7	23.0	32.8	42.6	52.5	42.6
σ	0.6	2.4	2.0	0.3	0.1	1.1	0.8	0.7
	SUMMER				**FALL**			
	JUN	**JUL**	**AUG**	**AVG**	**SEP**	**OCT**	**NOV**	**AVG**
AVG	73.8	76.6	73.4	74.6	67.0	55.7	42.4	55.0
MAX	83.5	85.9	83.2	84.2	76.4	66.7	51.6	64.9
MIN	64.0	67.3	63.5	64.9	57.5	44.7	33.1	45.1
σ	0.7	0.8	1.9	0.7	0.5	0.5	1.0	0.6
2008/2009	**WINTER**				**SPRING**			
YEAR	**DEC**	**JAN**	**FEB**	**AVG**	**MAR**	**APR**	**MAY**	**AVG**
AVG	29.8	27.0	36.8	31.2	46.6	53.1	64.0	54.5
MAX	40.3	37.3	48.0	41.9	58.1	63.2	74.5	65.3
MIN	19.2	16.7	25.5	20.5	35.0	42.9	53.4	43.8
σ	1.9	0.7	2.1	0.2	1.8	1.1	0.1	0.3
	SUMMER				**FALL**			
	JUN	**JUL**	**AUG**	**AVG**	**SEP**	**OCT**	**NOV**	**AVG**
AVG	74.2	72.3	72.0	72.8	66.5	50.7	50.0	55.7
MAX	84.1	81.7	81.9	82.6	76.2	59.1	59.1	64.8
MIN	64.3	62.8	62.0	63.0	56.8	42.3	40.8	46.6
σ	1.1	3.9	2.9	1.9	0.8	4.0	4.4	0.1
2009/2010	**WINTER**				**SPRING**			
YEAR	**DEC**	**JAN**	**FEB**	**AVG**	**MAR**	**APR**	**MAY**	**AVG**
AVG	30.7	25.3	27.0	27.7	45.8	60.4	64.4	56.9
MAX	38.4	31.8	35.2	35.1	55.5	72.4	73.7	67.2
MIN	23.0	18.7	18.8	20.2	36.1	48.4	55.1	46.5
σ	1.1	3.9	2.9	1.9	0.8	4.0	4.4	0.1
	SUMMER				**FALL**			
	JUN	**JUL**	**AUG**	**AVG**	**SEP**	**OCT**	**NOV**	**AVG**
AVG	76.6	78.7	78.3	77.9	68.2	58.7	46.0	57.6
MAX	85.8	87.6	88.7	87.4	78.3	71.8	57.7	69.3
MIN	67.3	69.8	67.9	68.3	58.1	45.6	34.3	46.0
σ	2.7	0.7	1.6	1.7	0.4	1.7	1.6	1.2

Table 5.

Table 5 shows the climatic averages for highs and lows during the meteorological seasons and the respective years in this research study. Standard deviations are also listed.

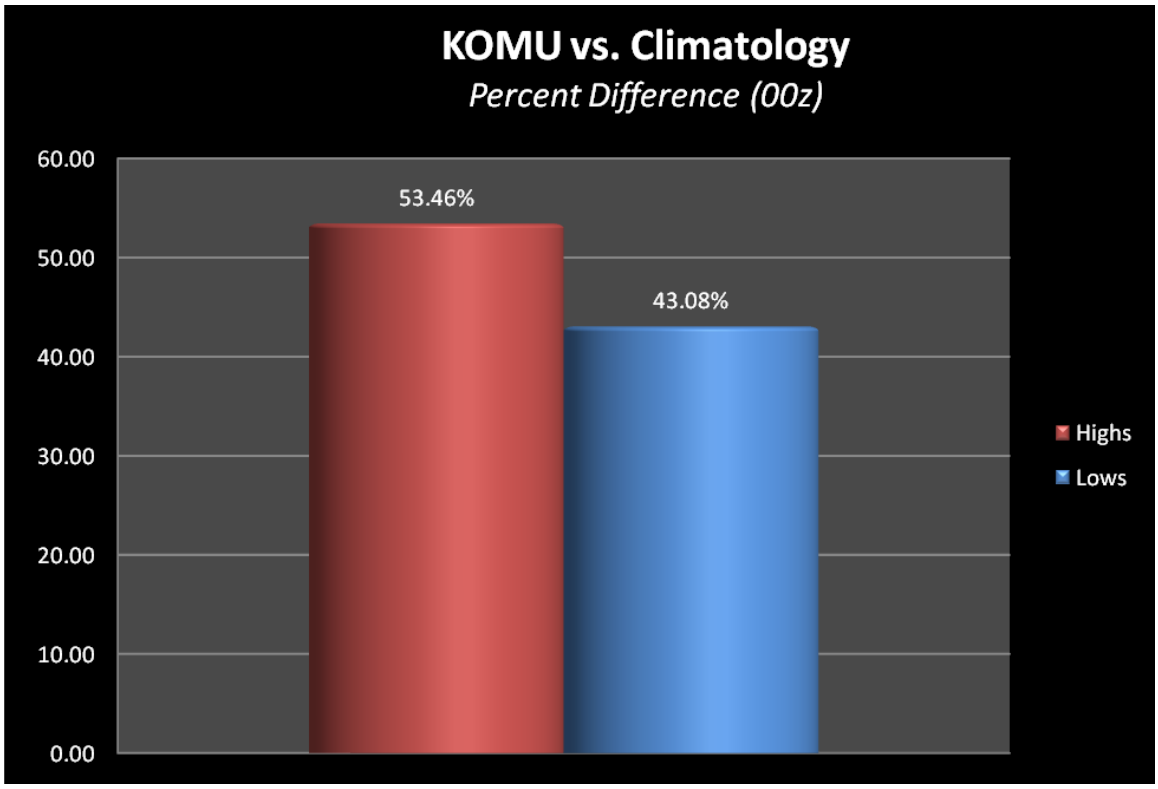

Figure 3.2.1

Figure 3.2.1 shows the percent difference between the KOMU produced forecast and the Climatology data at the 00z model timeframe. This figure shows that the KOMU highs and lows are 54% and 43% more skillful than the Climatology guidance.

While climatology is usually a good guide as to the daily behavior of weather conditions, it is merely that - a guide. There will be certain days (and even weeks) where special synoptic and mesoscale weather conditions will cause deviations from the acceptable climatology for a region (i.e., blocking high pressure systems leading to above average high temperatures, strong Canadian highs bringing unusually cold temperatures, etc.). It is because of the reasons listed above that climatology should be used as a guide only, and should never be taken seriously enough that 100% accuracy can be expected when using this method for high and low temperature predictions.

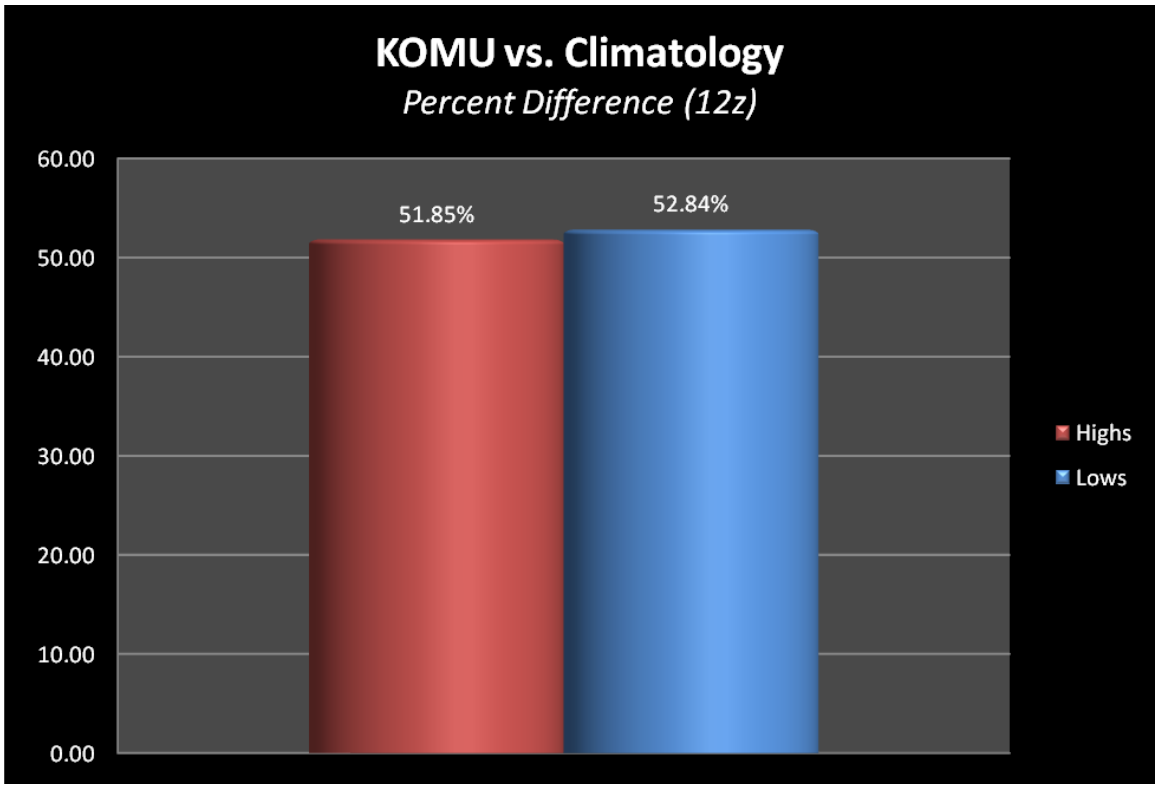

Figure 3.2.2

Figure 3.2.2 shows the percent difference between the KOMU produced forecast and the Climatology data at the 12z model timeframe. This figure shows that the KOMU highs and lows are ~52% and ~53% more skillful than the Climatology guidance.

When compiling and creating our 00z forecasts (here on out, 00z forecasts are the evening model runs used to create morning forecasts for the day and night time periods, respectively), we found that climatology underperformed when we compared those results to a human forecaster (KOMU). Figure 3.2.1 shows that both highs and lows were 54% and 43% more skillful, respectively, using the KOMU forecast when compared to the climatology outcome. However, both results (high and low climatology data) are far from a 100% accuracy level.

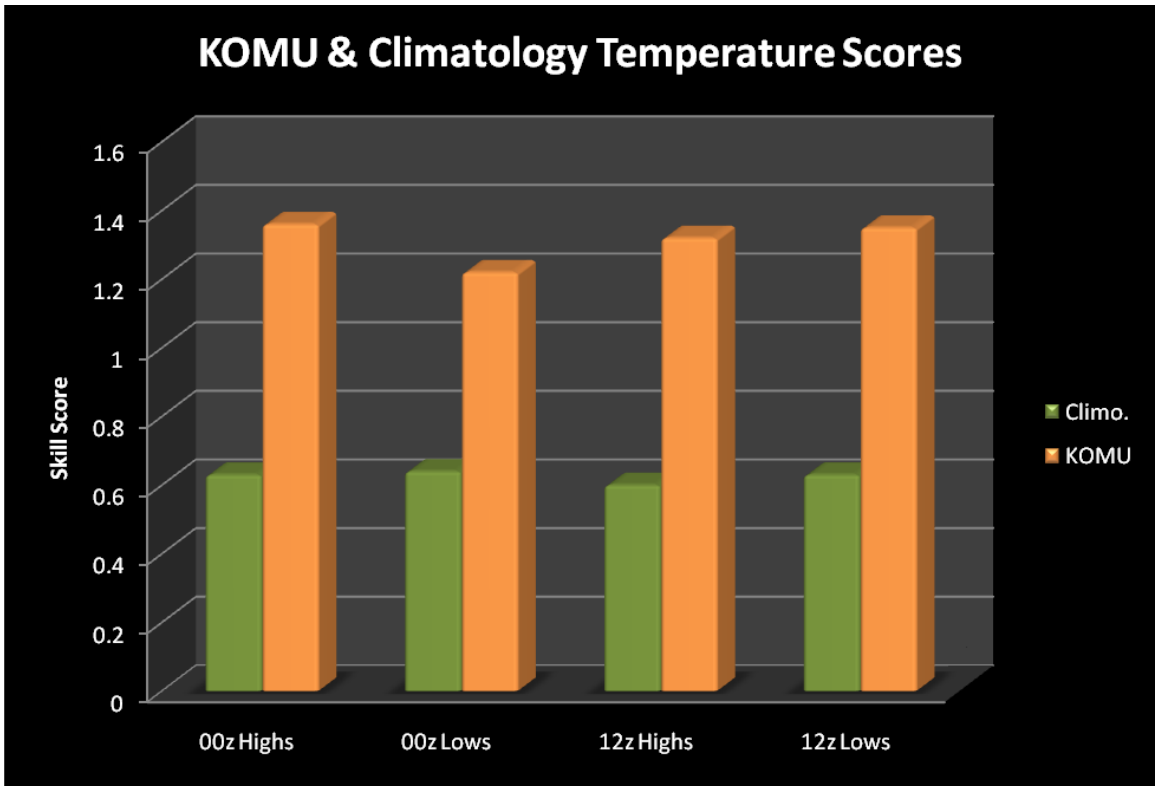

Figure 3.2.3

Figure 3.2.3 shows the Skill Scores for the KOMU and Climatology forecasts. The KOMU forecasts are superior to the climatology forecasts, as shown by the higher skill score values.

When looking at the results for the 12z forecasts (here on out, 12z forecasts are the morning model runs used to create evening forecasts for the night and day time periods, respectively), we found similar inaccuracies, but the high and low temperature results were reversed. Figure 3.2.2 indicates both highs and lows were ~52% and ~53% more skillful, respectively, using the KOMU forecast when compared to the climatology data. In terms of the skill score comparison between the KOMU forecast and climatology data, the differences are significant. Again, not surprising results, but it is certainly interesting to see the data first hand and see how things stack up. Figure 3.2.3 shows the temperature score comparisons for the KOMU forecast and the accepted values for climatology.

Based on our earlier temperature score equation, values that approach or equal two are accurate forecasts. Values that approach or equal zero are inaccurate. The human produced forecast excels when we compare it to our climatology findings.

3.3 – NAM MODEL

The NAM computer model was found to be the most accurate in predicting high and low temperatures, and in turn, most closely resembled the KOMU forecast. Based on the 00z forecasts, we found that the NAM computer model underperformed just slightly when we compared those results to a human forecaster. Figure 3.3.1 shows that both highs and lows were 10% and 8% more skillful, respectively, using the KOMU forecast when compared to the suggested values of the NAM model. This is a vast improvement over the climatology scheme (as would be expected), and would suggest that the equations and model physics are pretty well formulated to the region in terms of being able to produce an accurate forecast.

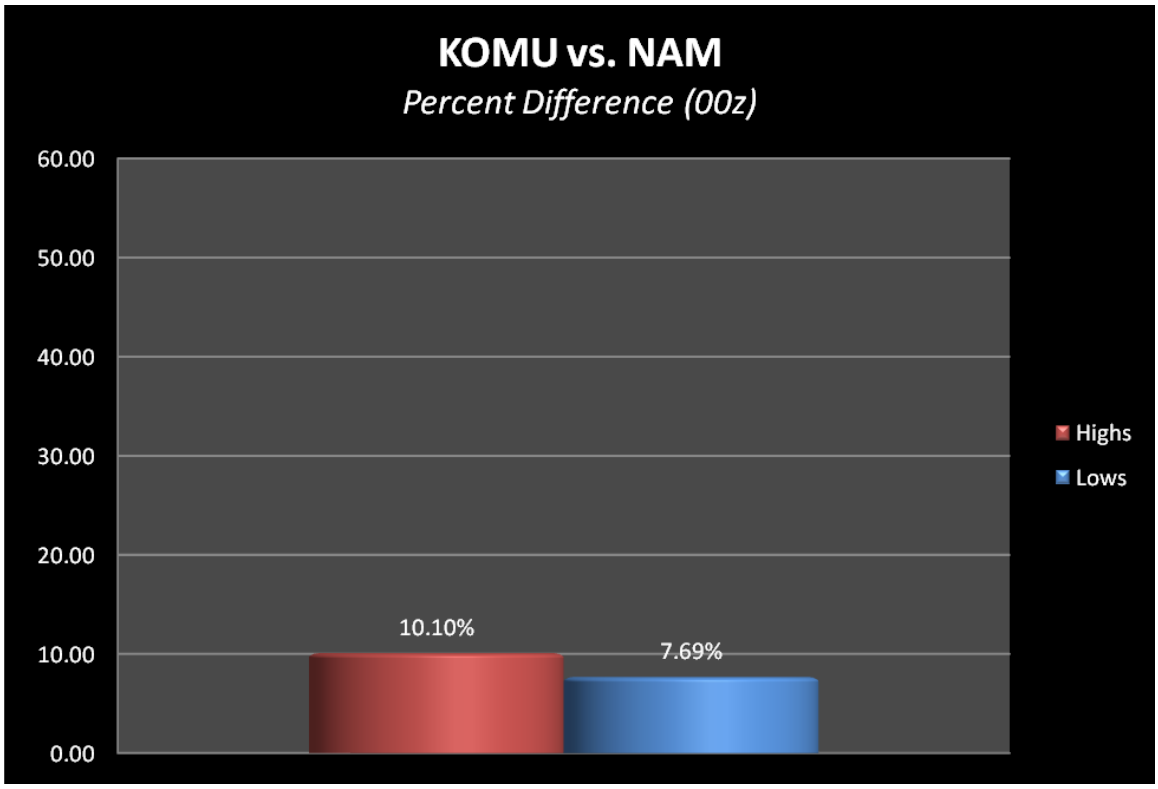

Figure 3.3.1

Figure 3.3.1 shows the percent difference between the KOMU produced forecast and the NAM computer model output at the 00z model timeframe. This figure shows that the KOMU highs and lows are ~10% and ~8% more skillful than the NAM computer model output.

When comparing the 12z forecast results, we found that the human produced forecast is 11% more skillful for high temperatures compared to the NAM model forecast output. In terms of low temperature accuracies, the human forecast is 16% more skillful than the model (Figure 3.3.2).

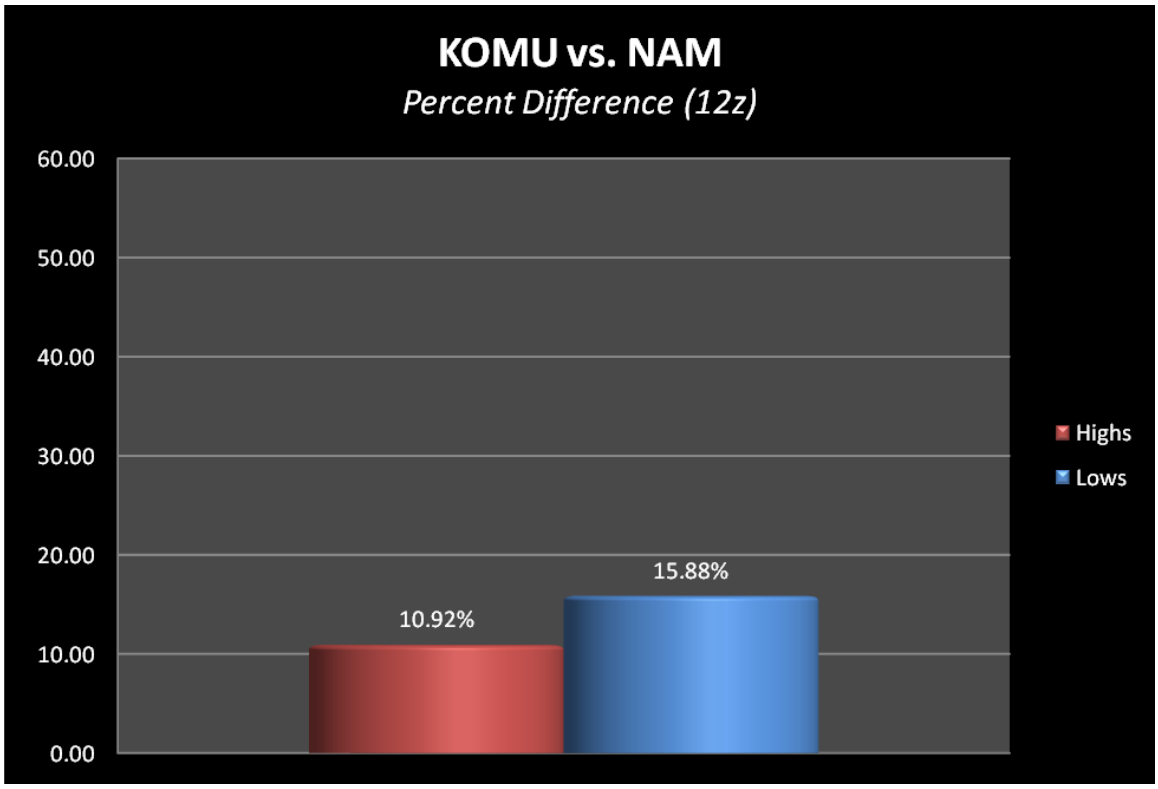

Figure 3.3.2

Figure 3.3.2 shows the percent difference between the KOMU produced forecast and the NAM computer model output at the 12z model timeframe. This figure shows that the KOMU highs and lows are ~11% and ~16% more skillful than the NAM computer model output.

As shown in the above figure, the forecast accuracy between the 12z NAM and KOMU is opposite from that of high/low accuracies from the 00z forecasts. This is likely due to model initialization and data fed into the models. The forecast periods are also closer to model initialization times, therefore the latest data would've been fed into the model equations, producing a more accurate depiction as to what would actually occur, temperature-wise. This is a recurring theme with the rest of the results this research produces.

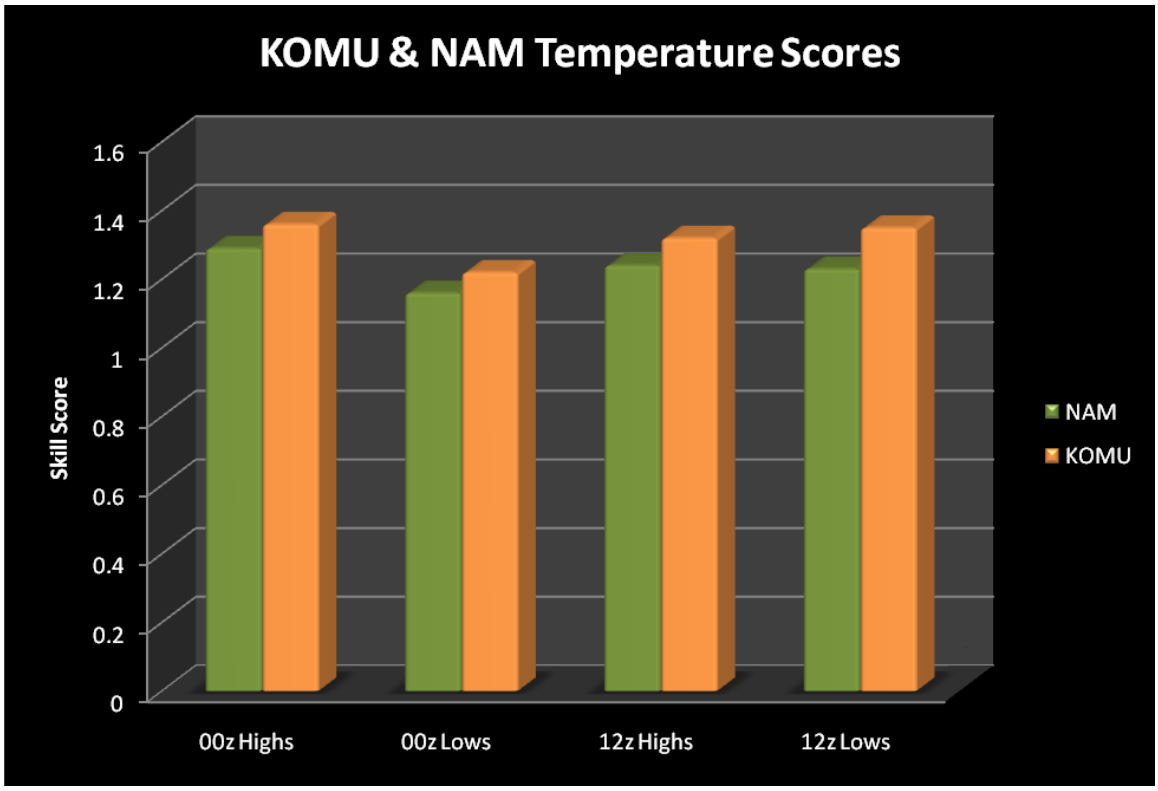

Figure 3.3.3

Figure 3.3.3 shows the temperature scores for the KOMU and the NAM computer model forecasts. The KOMU forecasts are superior to the NAM computer model forecasts, as shown by the higher temperature score values. However, there is a closer gap when compared to the climatology forecast guidance.

In terms of the temperature score comparison between the KOMU forecast and NAM forecast data, the differences aren't that significant when compared to those of climatology. Figure 3.3.3 shows the temperature score comparisons for the KOMU forecast and the NAM forecast data.

Based on our earlier temperature score equation, values that approach or equal two (2) are accurate forecasts. Values that approach or equal zero (0) are inaccurate. The human produced forecast still excels when we compare it to our NAM data, however, there seems to be better agreement on forecast highs and lows between the two samples.

3.4 – GFS MODEL

The GFS model is not as accurate as the NAM computer model, but we found that it is certainly more accurate than climatology. The discussion that follows are the results for the 00z and 12z runs of the GFS model compared to the human forecast. Comparing the human forecast to the GFS output doesn't show anything different from what we've already experienced, except for the idea that the errors begin to increase when we look at the observed highs/lows verses the model highs/lows. Model data and forecasts created based off of 00z GFS model data indicate that the model was some 23% behind the human produced forecast in terms of skill. For low temperatures, the GFS model was closer, but was still off by about 9% (Figure 3.4.1).

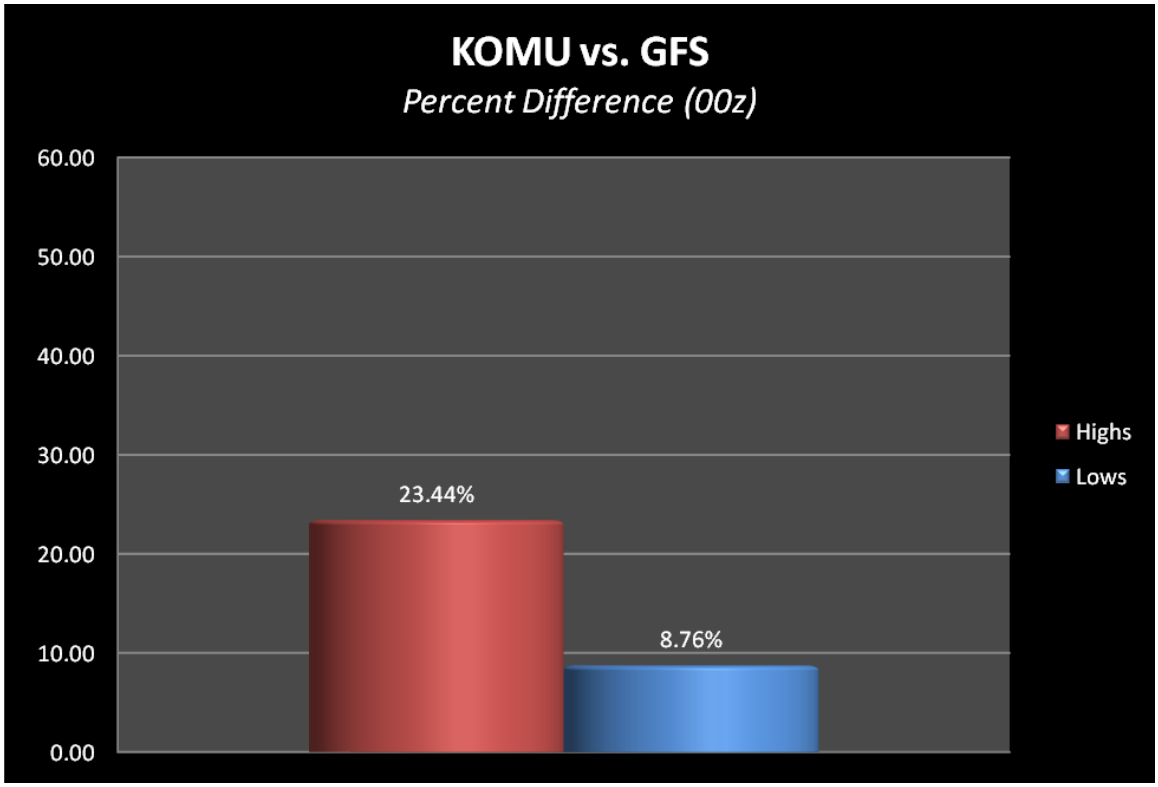

Figure 3.4.1

Figure 3.4.1 shows the percent difference between the KOMU produced forecast and the GFS computer model output at the 00z model timeframe. This figure shows that the KOMU highs and lows are ~23% and ~9% more skillful than the GFS computer model output.

For the same reasons listed above in section 3.3 regarding the accuracies associated with the different model initialization times and associated forecast accuracies, the same results were found with the 00z run of the GFS compared to the human forecast. Daytime high temperatures were demonstrated more skill (or were closer to the actual observed high temperature) compared to the skillfulness of the next morning's low temperature. Similar results were found with the 12z data (figure 3.4.2).

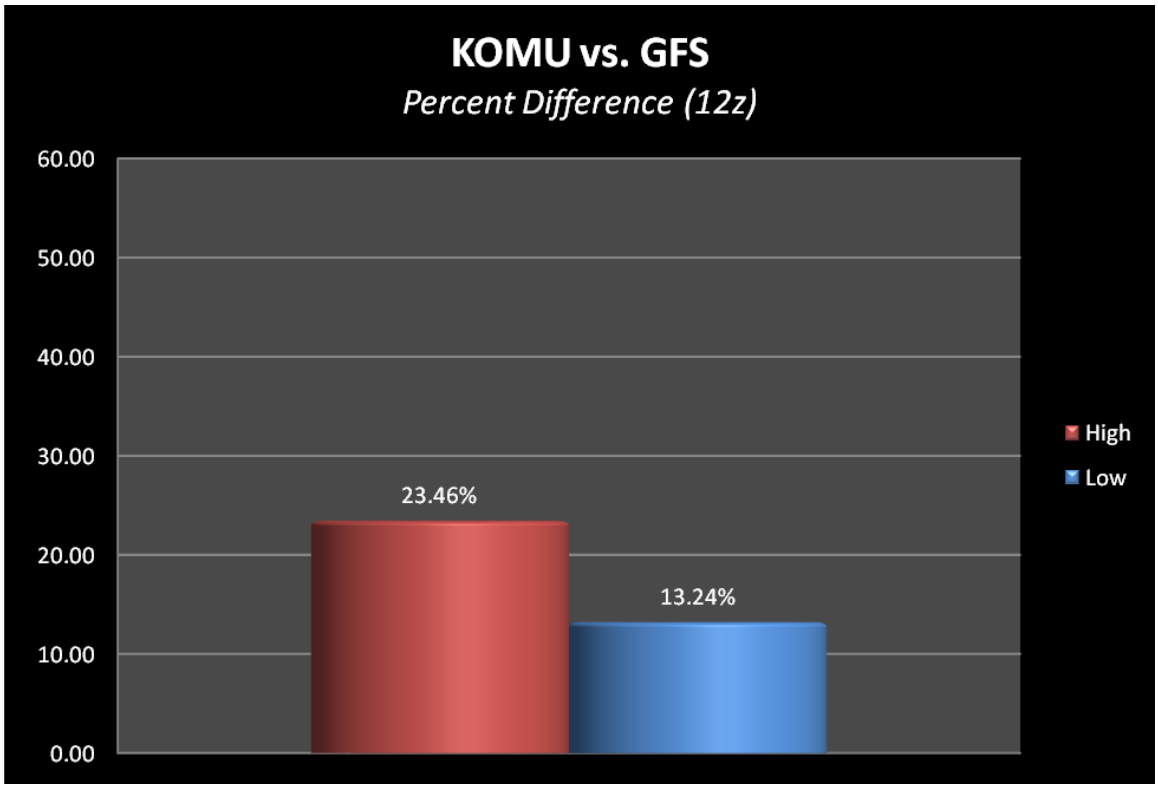

Figure 3.4.2

Figure 3.4.2 shows the percent difference between the KOMU produced forecast and the GFS computer model output at the 12z model timeframe. This figure shows that the KOMU highs and lows are ~23% and ~13% more skillful than the GFS computer model output.

Evaluating the differences between the GFS model forecast and the KOMU forecast for the 12z model run shows that in terms of high temperature predictions, the KOMU forecast outperformed the GFS MOS data by 23%. For low temperatures, the human produced forecast was better than the model by roughly 13%. In terms of the temperature score comparison between the KOMU forecast and GFS forecast data, the differences are going to be a bit more significant when compared to the NAM skill score values, but not as significant when compared to climatology.

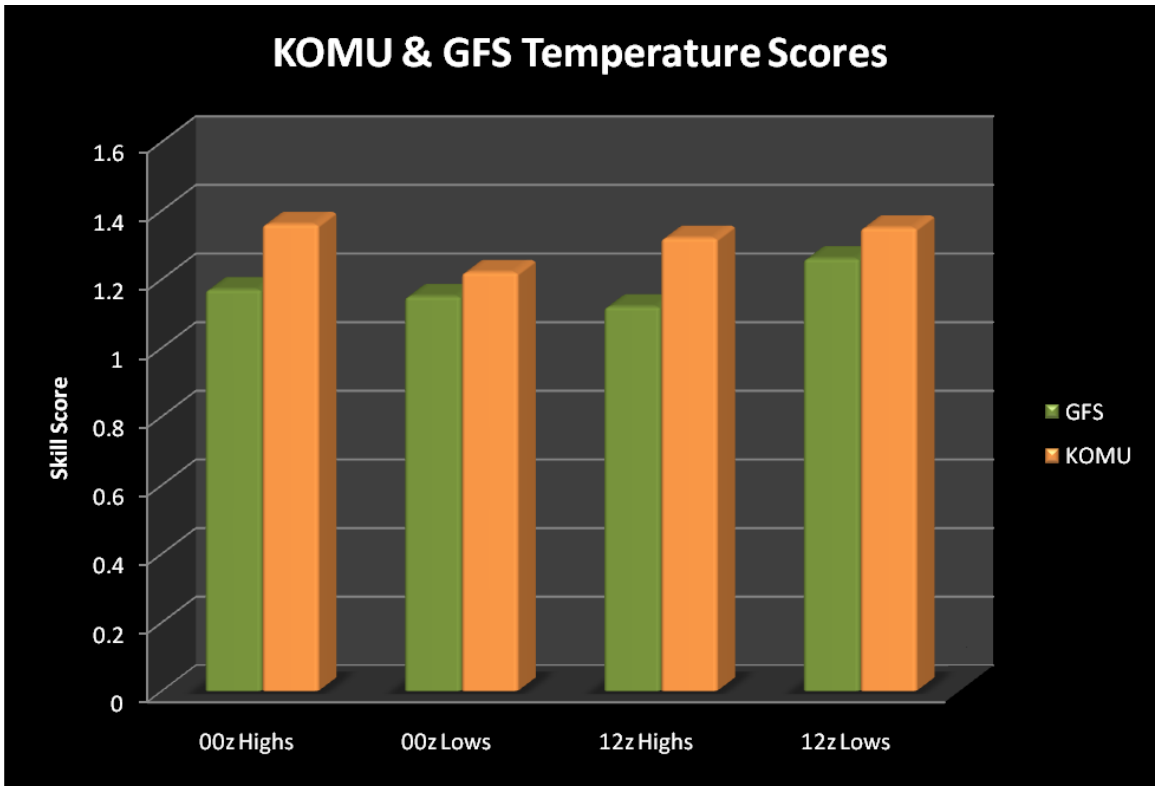

Figure 3.4.3

Figure 3.4.3 shows the temperature scores for the KOMU and the GFS computer model forecasts. The KOMU forecasts are superior to the GFS computer model forecasts, as shown by the higher temperature score values. However, there is a closer gap when compared to the climatology and NAM forecast guidance.

Figure 3.4.3 shows the temperature score comparisons for the KOMU forecast and the GFS forecast data. Based on our earlier temperature score equation, values that approach or equal two (2) are accurate forecasts. Values that approach or equal zero (0) are inaccurate. The human produced forecast still excels when we compare it to our GFS data, however, there seems to be better agreement on forecast highs and lows.

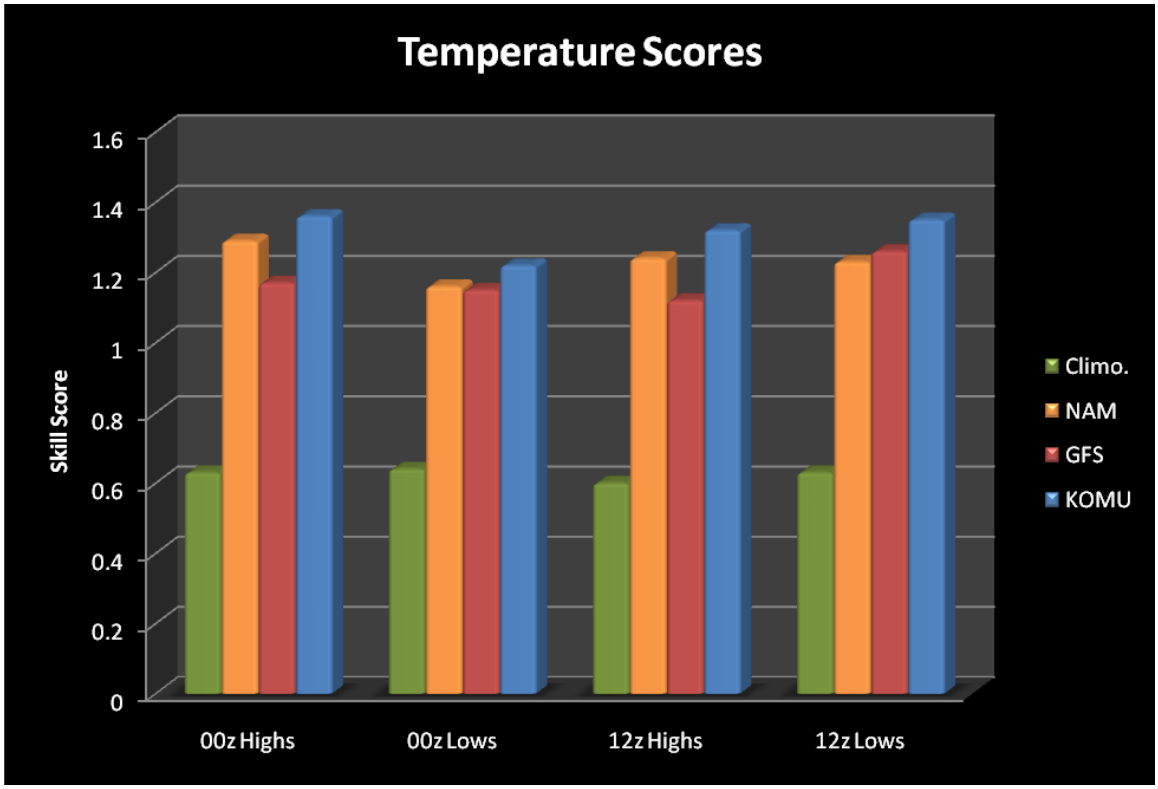

Figure 3.5.1

Figure 3.5.1 shows the temperature scores for all forecasts compiled (KOMU, GFS, NAM and Climatology forecasts). The KOMU forecasts are superior to the rest of the forecasting guidance for all model runs.

3.5 – KOMU VS. MODELS

Based on all of the data produced in this research, the most accurate forecast points toward the human forecast (or the forecast produced by staff meteorologists at KOMU-TV). There will be different meteorological events that the models will pick-up on and perhaps the human forecaster will miss. Occasionally, the computer models will actually be more accurate than the human produced forecasts, however, the human forecaster is able to take advantage of the ability to ask questions pertaining to the current state of the atmosphere and then based on past knowledge, can tweak his/her forecast

accordingly. Any human-produced forecast will add value to the forecast product whereas a model will not.

The computers can not take into consideration the different methods of forecasting (i.e., persistence, analogue, pattern-matching, etc.), and also can't analyze the accuracy of the initial conditions that were used to produce the forecast. As shown in figure 3.5.1, the human produced forecast is more accurate when compared to climatology data, and the computer generated forecasts. Although the models come close to a *perfect* forecast, they never achieve it. Of course, neither does the human produced forecast, but it does remain the most accurate when compared to the other data sets.

3.6 – SEASONALITY

The following section will compare the accuracy of the previous data sets with that of seasonality, and will determine which season, if any, is more accurate when using a computer model, climate data or a human produced forecast.

All seasons researched in this study were classified by meteorological seasons instead of the more common astronomical seasons that the general public is most familiar with. December, January and February, March, April and May, June, July and August, and September, October and November were classified as Winter, Spring, Summer and Fall, respectively.

3.6.1 – SPRING

Just as in the previous sections, the human forecast excelled in terms of accuracy. The NAM computer model came in behind the human produced forecast in terms of producing accurate forecasts for both the high temperatures and the low temperatures. However, the NAM 12z model fell behind the GFS 12z model for low temperature forecasts. In terms of using climatology as a reliable forecasting tool, it is a poor performer. With so much temperature variability in the spring months, this is not surprising. Figure 3.6.1 shows the results for the Spring Temperature Score.

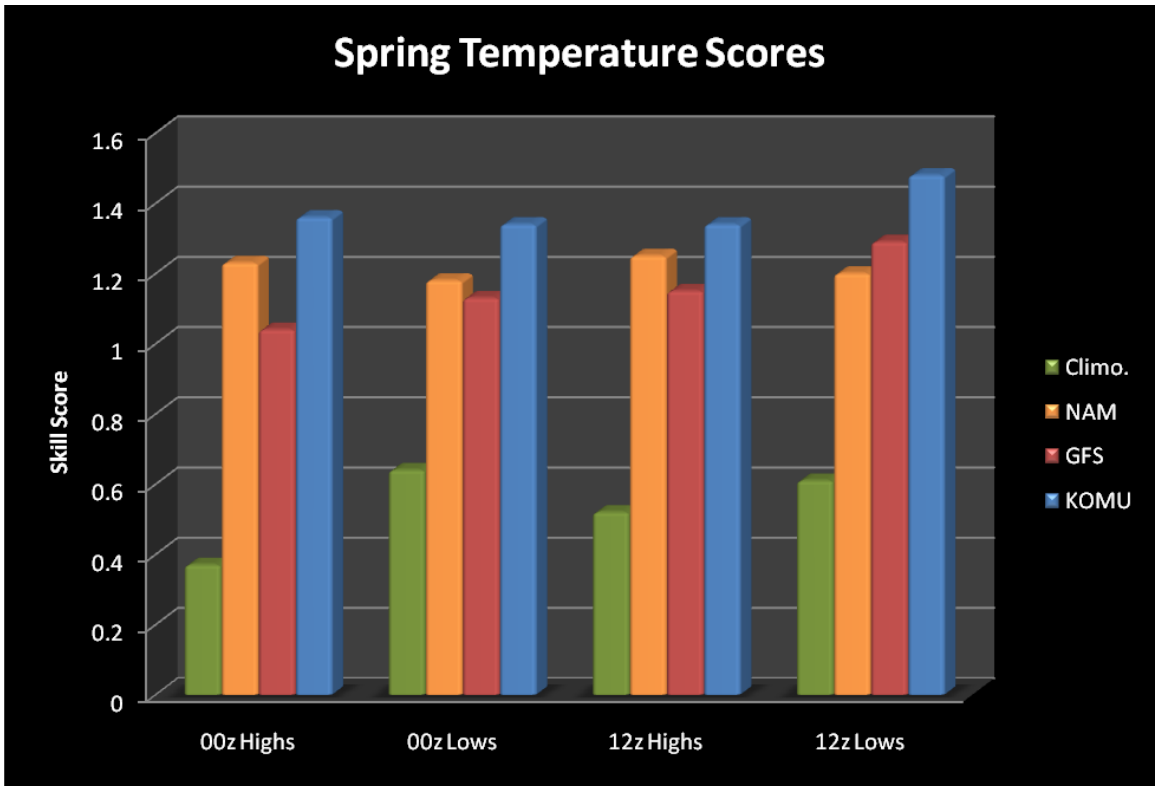

Figure 3.6.1

Figure 3.6.1 shows the temperature scores for all Spring forecasts compiled (KOMU, GFS, NAM and Climatology forecasts). The KOMU forecasts are superior to the rest of the forecasting guidance for all model runs.

3.6.2 – SUMMER

Very similar to the prior seasonal variations, the human forecast excelled in terms of accuracy. The NAM computer model was more skillful than the non-human test subjects in terms of producing accurate forecasts for both the high temperatures and the low temperatures. One interesting finding between the four samples is how far the degree of separation is with 12z and 00z high temperature forecasts. The KOMU and NAM forecasts are pretty close together in terms of skill. However, the skill score for the GFS 12z and 00z high temperature forecasts are far less than the KOMU and GFS forecasts.

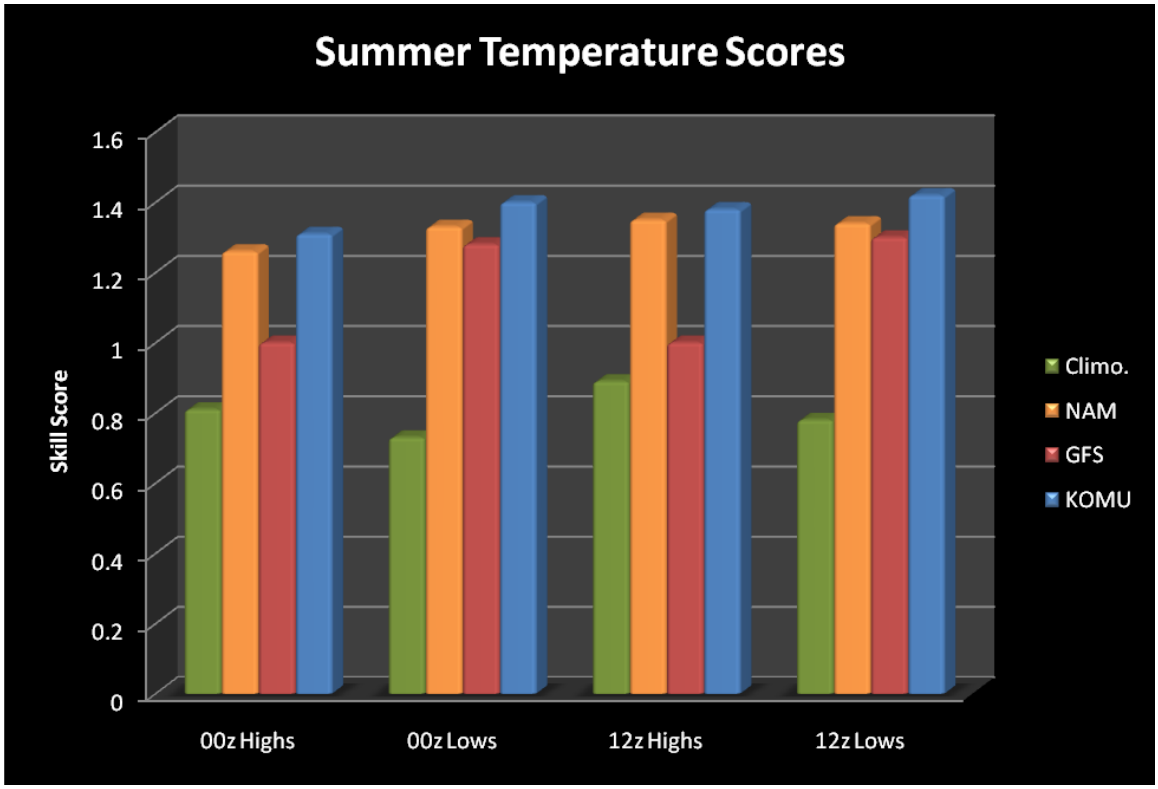

Figure 3.6.2

Figure 3.6.2 shows the temperature scores for all Summer forecasts compiled (KOMU, GFS, NAM and Climatology forecasts). The KOMU forecasts are superior to the rest of the forecasting guidance for all model runs.

The low temperature forecast scores are similar to what we found in the spring, and the scores are reasonably closer together. During our period of research, it would be conclusive to say the GFS is the least accurate model source for providing accurate high temperature forecasts. Climatology was the least reliable forecasting tool, however, the skill scores are increased from those found in the spring dataset. Temperature variability isn't as great in the summertime, so this is explains why the climatology skill scores increased. Figure 3.6.2 shows the results for the Summer Skill Score.

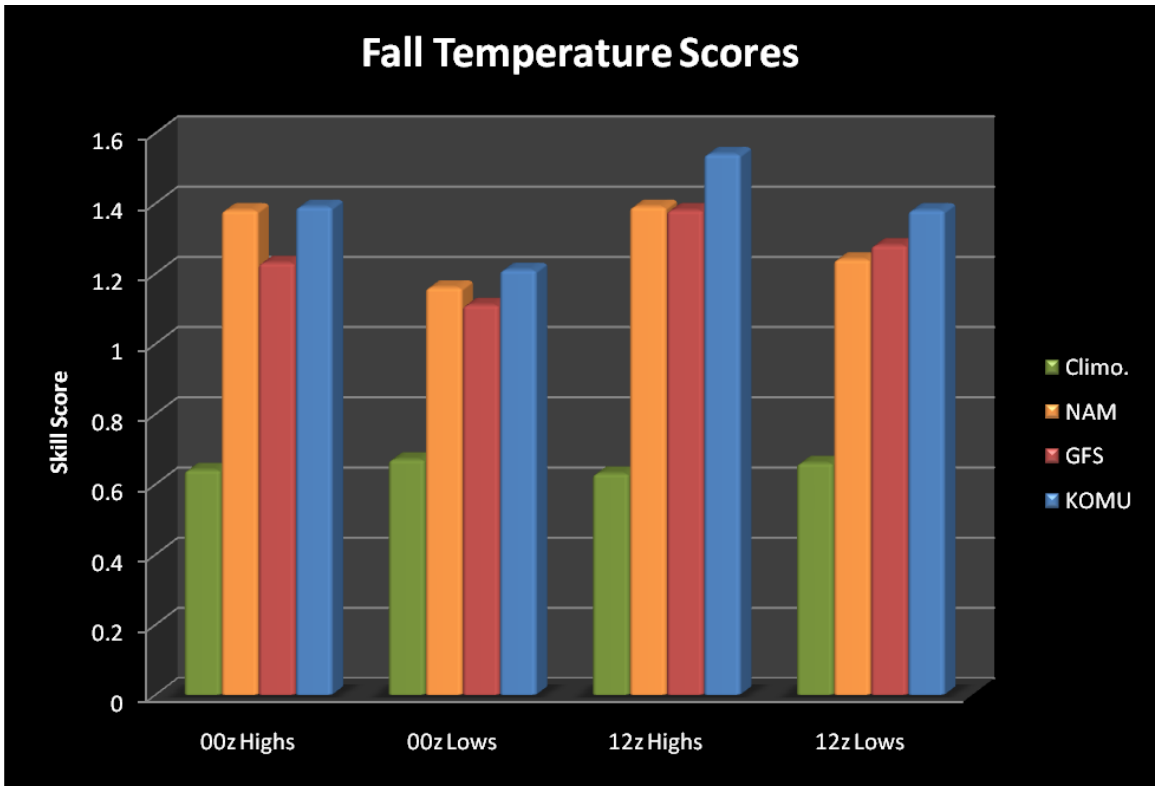

Figure 3.6.3

Figure 3.6.3 shows the Skill Scores for all Fall forecasts compiled (KOMU, GFS, NAM and Climatology forecasts). The KOMU forecasts are superior to the rest of the forecasting guidance for all model runs.

3.6.3 – FALL

Not surprising, the human forecast excelled in terms of accuracy during the fall season. One of the most interesting results in the Fall comparison is the almost identical results of the Spring accuracies. Figure 3.6.3 shows the results for the Fall Temperature Score, and in terms of similarities, the results closely resemble the Spring Temperature Scores (Figure 3.6.1). The temperature score values for Fall are increased slightly when compared to the Spring temperature scores, but the model rankings are exactly the same. The higher temperature score values are likely contributed to the similar temperature variability that was experienced during the summer, as the variability in temperature during the Fall season

doesn't normally change until late in the season. Climatology temperature score values are also increased when compared to the temperature score values of the remaining samples, but are still on the lower end of temperature score accuracy. These results are slightly elevated due to the fact that temperature variability isn't as great as it would be in the winter or spring.

3.6.4 – WINTER

Just as in the previous sections, the human forecast excelled in terms of accuracy, with one exception. The 12z and 00z high and low temperature forecast score values were significantly lower than in previous seasons, and during the 12z low temperature forecast period, both models (NAM and GFS) outperformed the human forecaster. There are multiple reasons why this occurred, and one can speculate that those reasons would include model/human biases, as well as model initialization issues. Another observation that wasn't as surprising, but does support earlier hypothesis listed above regarding climatology, is that the climatology temperature scores drop off dramatically from Fall into Winter. Temperature variability increases during the Winter and Spring seasons, so the overall accuracy of a localized climate model would drop off significantly, and that is dually noted in Figure 3.6.4.

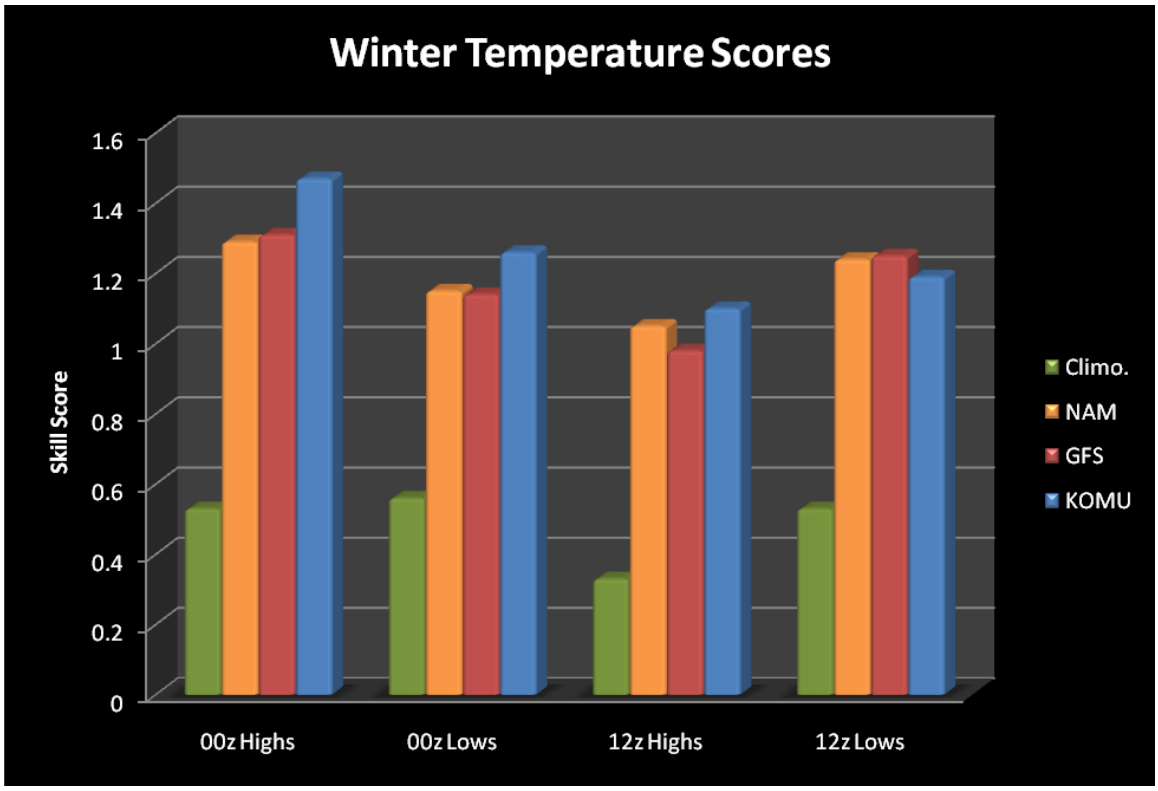

Figure 3.6.4

Figure 3.6.4 shows the temperature scores for all Winter forecasts compiled (KOMU, GFS, NAM and Climatology forecasts). The KOMU forecasts are superior to the rest of the forecasting guidance for all model runs, with the exception being the 12z low temperature forecasts.

3.7 – SEASONALITY COMPARISONS

Based on the results from this study, winter is one of the hardest seasons for which to produce an accurate temperature forecast. Various meteorological features contribute to the variability of not only temperature, but to precipitation as well (which are both directly and indirectly related). The location of the polar branch of the jet stream in the winter and various Arctic high pressure systems will influence temperature patterns in the winter, thus, creating a difficult forecasting regime. More persistent weather patterns exist during the summer months, and will create a more predictable (and accurate) type of weather pattern, which is why forecast accuracy is slightly higher during the summer

months. Distinct blocking weather patterns are more common during the summer months, and that is a chief contributor to this reasoning (Barriopedro et. al, 2006).

When comparing the two model runs (and forecasting shifts at KOMU-TV), we can break down the accuracies further, although the end results are typically the same. As shown in Figure 3.7.1, there seems to be a similarity with the KOMU forecast and the GFS forecast temperature score trend from season to season. The KOMU forecast still has the higher temperature score value, but the "up and down" trends both of these test subjects portray is very similar. Comparing the temperature score trend lines for the NAM and Climatology output, one finds a similar path as well. In Figure 3.7.2, the same trend patterns show up and the KOMU and GFS forecast seem to take on a gradual rise on the temperature score track, especially when you look at the transition from summer into fall. NAM and Climatology forecasts tend to follow the same path as shown in Figure 3.7.1, and will gradually develop less skill as you go from summer into winter. Overall, there appears to be greater variability between the human forecasts, and the computer models for the 12z data than there is with the 00z products. One interesting aspect of this seasonality comparison is the difference in the temperature scores when compared to the two different model runs and forecast shifts at KOMU-TV. There is a much wider discrepancy in the temperature score values for 12z data, while the 00z data is pretty much uniform (or leveled out).

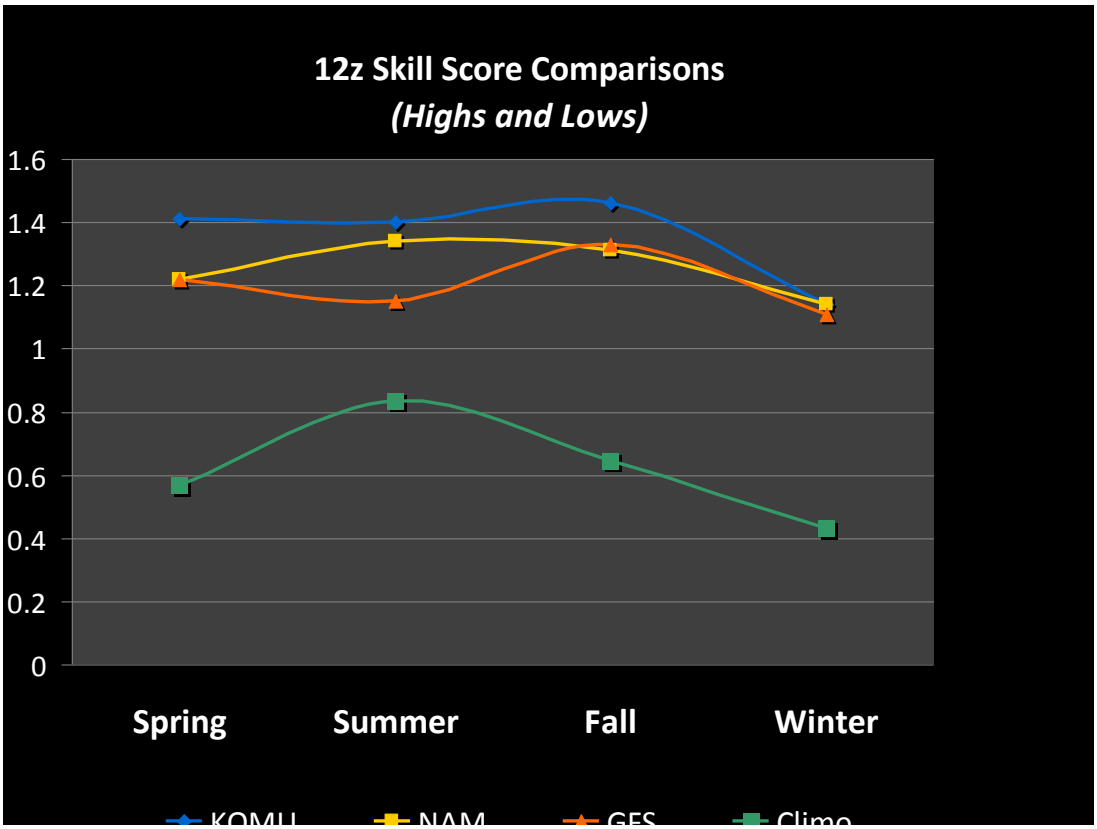

Figure 3.7.1

Figure 3.7.1 shows the temperature scores for all seasons along with the 12z human and model forecasts compiled (KOMU, GFS, NAM and Climatology forecasts).

Overall, the most accurate forecast (model or otherwise) period is found to be when temperature variability is relatively low. Obviously, meteorological effects will bring about changes in the overall accuracy of the models and the humans, and how each handles the meteorological characteristics differently. Biases and forecast experience will play a role as well, but this is something that is constantly being improved as atmospheric research continues to grow and one's knowledge about the atmosphere and the equations that govern its motions are better understood.

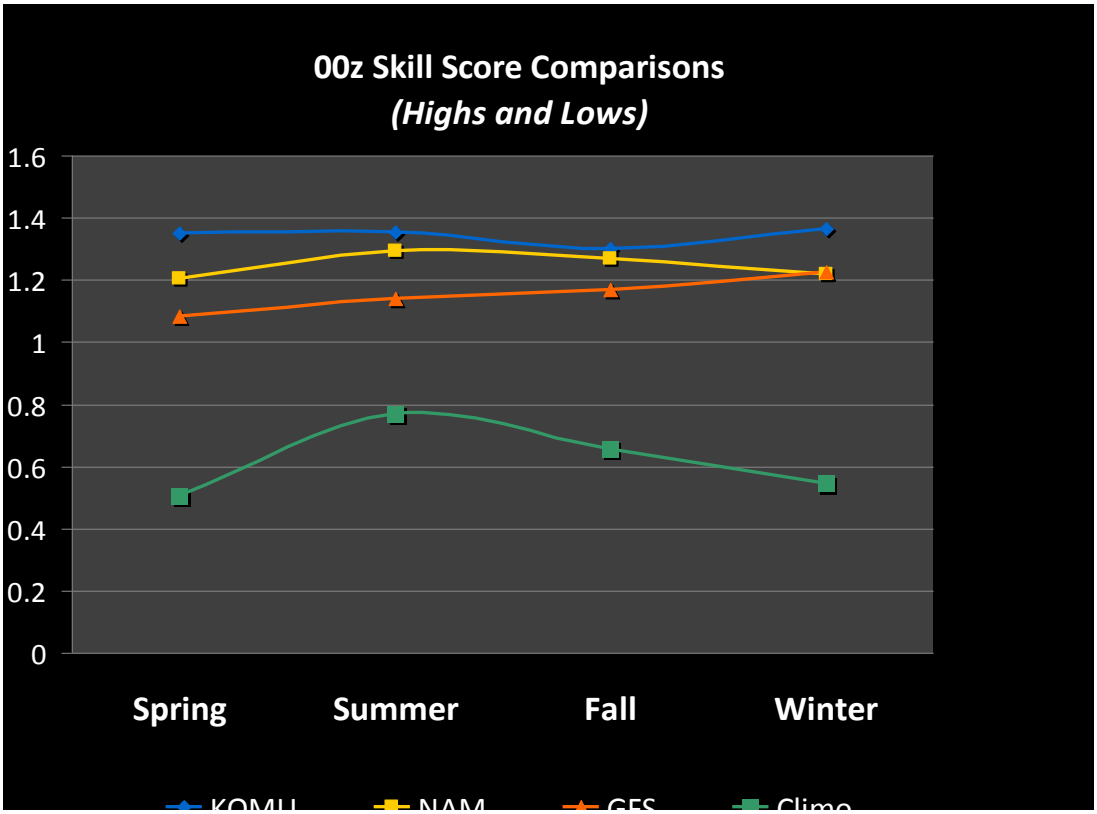

Figure 3.7.2

Figure 3.7.2 shows the temperature scores for all seasons along with the 00z human and model forecasts compiled (KOMU, GFS, NAM and Climatology forecasts).

3.8 – STATISTICAL DATA AND RESULTS

In this study, there was also a breakdown of the temperature scores and standard deviations associated with the human-produced forecast, the models and the four meteorological seasons. All of the data were tested for significance at the 95% confidence interval and none of the results were significant (when compared to climatology). Only two of the data sets were significant at the 90% confidence level as noted by an asterisk in Table 6a and 6b.

12Z HIGHS	*SPRING*	*SUMMER*	*FALL*	*WINTER*
KOMU	1.34/0.817	1.38/0.753	***1.54/0.713**	1.10/0.837
GFS	1.15/0.876	1.00/0.868	1.38/0.825	0.98/0.884
NAM	1.25/0.804	1.35/0.801	1.39/0.790	1.05/0.884
12Z LOWS	*SPRING*	*SUMMER*	*FALL*	*WINTER*
KOMU	1.48/0.768	1.42/0.769	1.38/0.768	1.19/0.783
GFS	1.29/0.877	1.30/0.836	1.28/0.801	1.25/0.822
NAM	1.20/0.886	1.34/0.830	1.24/0.843	1.24/0.822

Table 6a.

(Temperature Score/Standard Deviation)
**Forecast is better than climatology at the 90% confidence interval.*

00Z HIGHS	*SPRING*	*SUMMER*	*FALL*	*WINTER*
KOMU	***1.36/0.772**	1.31/0.797	1.39/0.780	1.47/0.829
GFS	1.04/0.869	1.00/0.814	1.23/0.818	1.31/0.875
NAM	1.23/0.826	1.26/0.830	1.38/0.828	1.29/0.807
00z Lows	*SPRING*	*SUMMER*	*FALL*	*WINTER*
KOMU	1.34/0.852	1.40/0.746	1.21/0.848	1.26/0.889
GFS	1.13/0.841	1.28/0.789	1.11/0.852	1.14/0.821
NAM	1.18/0.868	1.33/0.843	1.16/0.860	1.15/0.895

Table 6b.

(Temperature Score/Standard Deviation)
**Forecast is better than climatology at the 90% confidence interval.*

However, the data sets that did meet the 90% confidence level represent a stronger result when compared to the others. Lupo and Market (2002) showed similar results when applying the same statistical tests, and in fact, showed that

very few temperature results met the 90% confidence level. Conclusively, more datasets did meet the 90-95% confidence intervals in their study than here, solely analyzing temperatures. Lupo and Market (2002) also included more meteorological variables in their study, including those such as precipitation and cloud cover which increased the overall forecast performance as demonstrated in their study. The fact that the majority of the results did not meet the 90% confidence level should not be discouraging, however, as the temperature score and regression data still suggest greater forecast accuracies and improvement in a human produced forecast.

CONCLUSIONS

4.1 – THE CONCLUSIONS

Weather is something that impacts everyone on a daily basis. An accurate forecast is one which a person may or may not notice, but an inaccurate forecast almost always brings about awareness. The latter has been the subject of constant research, trying to improve upon past data and occurrences to bring about improvement in the future. Regarding the future, a subject that has been considered as of late: what is the future of human forecasting? Roebber and Bosart (1996) argue that the relative advantage of human forecasters reflects the ability to recognize instances of model forecast bias to certain synoptic situations. Identifying the events that the models predict accurately (and inaccurately) is key to moving forward in this endeavor. By knowing the biases in the computer models and how they handle the changes in meteorological seasons is important in determining which model output should be used as guidance in predicting meteorological events. The intent here is to use the models as *guidance*, along with one's education to produce an accurate forecast. Experience in forecasting is a very important aspect of forecasting skill, and time should be devoted to discern how the models will handle different meteorological events – accurately or inaccurately. From this, the forecaster can proceed to put together an accurate forecast using the guidance data and past experience to excel in an accurate outcome.

In this research, forecast and observed data was analyzed over a three year period in Columbia, Missouri. Data used in this study was MOS data from the NAM and GFS computer models, climatology data (based on the thirty year climate data obtained from the National Climatic Data Center (NCDC)), as well as daily high and low temperature forecasts produced by meteorology students from the University of Missouri and on-air meteorologists at KOMU-TV. Based on this research, it is evident that climatology is not a preferred (or accurate) method for producing a forecast. Rather, it is a much better tool for analyzing temperature trends over time and producing a seasonal or climatological outlook. In this work, it was shown that there was a distinct difference in the skill score values of the climatology data compared to the data compiled with respect to a human produced forecast (KOMU) and the computer models (NAM and GFS). It was also shown that a human produced forecast (KOMU) is more accurate than the other data sources used in this study. When comparing the two computer models, the NAM is superior over the GFS and even shows a distinct accuracy in the high temperature forecast data (12z and 00z high temperatures).

When dealing with the overall forecast accuracy of the model runs and forecasting shifts (KOMU forecasts), the forecast time period closest to the time of the model run, or forecast issuance, was always more accurate. For example, a daytime high temperature forecast created on the same day the high was to occur resulted in better accuracy than if that forecast would have been issued the day before. The same pattern can be applied to low temperature forecasts as well.

However, there does seem to be more variability in the forecast accuracy (KOMU and MOS forecasts) on the 12z forecasts and model runs than with the 00z data. The results show this as the skill score values increase and decrease as one moves from season to season (although, there is a significant shift in the skill score values as we move from fall into winter). This same pattern is not realized on the 00z level, however.

Finally, seasonality was examined to determine who was more accurate during the four meteorological seasons. Based on the results from this study, winter is one of the hardest seasons to produce an accurate temperature forecast. Various meteorological features contribute to the variability of not only temperature, but to precipitation as well (which are both directly and indirectly related).

One consistency, however, is that climatology accuracies peak during the summer and drop off quickly as temperature variability increases toward the fall and winter seasons. This result is also demonstrated, although on a smaller scale, when looking at the overall tendency for skill scores to drop in the late summer, fall and winter months for the NAM, GFS and KOMU-TV forecast. Overall, the most accurate forecast (model or otherwise) period is found to be when temperature variability is relatively low. In addition, very similar results in variability were observed with this study as compared to Lupo and Market (2002).

Overall, the results in this study show that humans are needed to produce accurate forecasts, and further more, humans add value to any forecast produced. Models can get close to a consistent accuracy, but without human interaction or thought, putting complete trust in a computer model could be dangerous.

Nonetheless, using the results of this research, it can be shown that human forecast accuracy is great and this should dispel the belief that meteorologists are consistently incorrect.

REFERENCES

Barriopedro, D., García-Herrera, R., Hernández, E., Lupo, A.R., 2006: A Climatology of Northern Hemispheric Blocking. *Journ. of Climate.*, **19**, 1042-1063.

Bjerkness, V., 1914: Meteorology as an Exact Science. *Mon. Wea. Rev.,* **43**, 14.

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

Carter, G. M., J. P. Dallavalle, and H. R. Glahn, 1989: Statistical forecasts based on the National Meteorological Center's numerical weather prediction system. *Wea. Forecasting,* **4**, 401-412.

Dallavalle, J.P., 1998: A perspective on the use of model output statistics in objective weather forecasting. Preprints, *Sixteenth Conf. On Weather Analysis and Forecasting,* Norfok, VA, Amer. Meteor. Soc., 479-482.

Doswell, C.A. III, 2004: Weather Forecasting by Humans: Heuristics and Decision Making. *Wea. Forecasting,* **19**, 1115-1126.

Franklin, B., 1774: *Experiments and Observations on Electricity*. Reprinted 1941. *Benjamin Franklin's Experiments,* I. B. Cohen, Ed., Harvard University Press, 453 pp.

Lupo, A.R., Market, P.S., 2002: The Application of a Simple Method for the Verification of Weather Forecasts and Seasonal Variations in Forecast Accuracy. *Wea. Forecasting,* **17**, 891-897.

Lutgens, F.K., Tarbuck, E.J., 1995: The Atmosphere. 6th Edition

Miller, E.R., 1933: American Pioneers in Meteorology. *Mon. Wea. Rev.,* July, 1933. 189-193.

Roebber, P.J., Bosart, L.F., 1996: The Complex Relationship between Forecast Skill and Forecast Value: A Real-World Analysis. *Mon. Wea. Rev.,* December, 1996. 544.