working to advance road weather information systems technology

Performance Benchmarking of RWIS Pavement Temperature Forecasts

http://aurora-program.org

Aurora Project 2000-01

University of Waterloo

Final Report
March 2012

About Aurora

Aurora is an international program of collaborative research, development and deployment in the field of road and weather information systems (RWIS), serving the interests and needs of public agencies. The Aurora vision is to deploy RWIS to integrate state-of-the-art road and weather forecasting technologies with coordinated, multi-agency weather monitoring infrastructures. It is hoped this will facilitate advanced road condition and weather monitoring and forecasting capabilities for efficient highway maintenance, and the provision of real-time information to travelers.

The preparation of this report was financed in part through funds provided by the Iowa Department of Transportation through its "Second Revised Agreement for the Management of Research Conducted by Iowa State University for the Iowa Department of Transportation" and its amendments.

Technical Report Documentation Page

1. Report No.	**2. Government Accession No.**	**3. Recipient's Catalog No.**
Aurora Project 2000-01		

4. Title and Subtitle
Performance Benchmarking of RWIS Pavement Temperature Forecasts

5. Report Date
March 2012

6. Performing Organization Code

7. Author(s)
Tae J. Kwon and Liping Fu

8. Performing Organization Report No.
Aurora Project 2000-01

9. Performing Organization Name and Address
Transportation Systems Research Group
Department of Civil and Environmental Engineering
University of Waterloo
Waterloo, Ontario, Canada N2L 3G1

10. Work Unit No. (TRAIS)

11. Contract or Grant No.

12. Sponsoring Organization Name and Address

Aurora Program	Federal Highway Administration
Iowa Department of Transportation	U.S. Department of Transportation
800 Lincoln Way	1200 New Jersey Avenue SE
Ames, Iowa 50010	Washington, DC 20590

13. Type of Report and Period Covered
Final Report

14. Sponsoring Agency Code
Pooled Fund SPR 72-00-0003-042

15. Supplementary Notes
Visit www.aurora-program.org for color PDFs of this and other research reports.

16. Abstract

The researchers examined five categories of factors that may affect the accuracy of road weather information system (RWIS) pavement temperature forecasts: climatic trends, locational attributes, seasonal/monthly variations, diurnal trends, and forecast length. Five hypotheses were established accordingly and tested using one year of RWIS observations and forecasts obtained from several provinces in Canada. The RWIS networks were classified into three groups on the basis of the climatic nature of the region in which they are located: maritime, continental, and mixed.

- Pavement temperature forecasts from the maritime climate group had the highest quality and those from the mixed climate group had the lowest quality, both in terms of mean absolute errors (MAEs) and percent of acceptable forecasts (PAFs). The significant performance differences between the regions suggested that the RWIS forecasting performance may be affected by climatic trends, as in, the unique climatic patterns of the regions may have caused the differences in RWIS forecasting performance.
- The correlation between the forecasting accuracy of RWIS stations and their topographical features, such as altitude and amount of vegetation cover, and geographical features, such as distance to local lakes/waters, were investigated within each region. The researchers found that the forecasting performance for the RWIS stations in the maritime climate region near coastal areas had a negative correlation with the distance from a nearby large water body. On the other hand, no significant correlation was found in either the mixed or continental climate groups.
- Daytime forecasts were less accurate than the ones generated for nighttime. Furthermore, as expected, the accuracy of forecasts was found to deteriorate quickly as the forecasting horizon increases.
- Forecast errors were found to exhibit seasonal variations with forecasts for the shoulder months (October and April) tending to be poorer than other months.
- There was a clear quantitative relationship between forecast errors and forecasting time and length, suggesting that it is possible to quantify these errors based on the time a forecast is made and the time the condition is to be forecasted (forecasting horizon).

17. Key Words
Canadian road weather—climatic analysis—forecast accuracy—forecast errors—regional investigation—surface conditions

18. Distribution Statement
No restrictions.

19. Security Classification (of this report)	**20. Security Classification (of this page)**	**21. No. of Pages**	**22. Price**
Unclassified.	Unclassified.	37	NA

PERFORMANCE BENCHMARKING OF RWIS PAVEMENT TEMPERATURE FORECASTS

Final Report
March 2012

Authors
Tae J. Kwon, PhD and Liping Fu, PhD., PE, Professor
Transportation Systems Research Group
Department of Civil and Environmental Engineering, University of Waterloo

Sponsored by
the Aurora Program,
the Iowa Department of Transportation,
and the Federal Highway Administration (FHWA)
FHWA Pooled Fund Study SPR-3(042): AK, IA, IL, IN, KS, MI, MN, ND, NV, NY, OH,
Ontario MOT, PA, Quebec MOT, SD, Sweden NRA, TN, UT, VA, WI

Preparation of this report was financed in part
through funds provided by the Iowa Department of Transportation
through its Research Management Agreement with the
Institute for Transportation
(Aurora Project 2000-01)

A report from
Aurora Program
Institute for Transportation
Iowa State University
2711 South Loop Drive, Suite 4700
Ames, IA 50010-8664
Phone: 515-294-8103
Fax: 515-294-0467
www.intrans.iastate.edu

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

This research was conducted under the Federal Highway Administration (FHWA) Pooled Fund Study SPR-3(042). The authors would like to express their gratitude to the FHWA, the Aurora Program partners, and the Iowa Department of Transportation (lead state) for their financial support and technical assistance.

There are many people who provided a tremendous wealth of information and shared their knowledge throughout the course of completing this project.

We are deeply grateful to Max Perchanok at the Ontario Ministry of Transportation (MTO) for his continuous support and helpful comments.

Special thanks to meteorologists Paul Delannoy, Elizabeth Loder, Jeremy Duensing, and Gregg Benedict for their valuable consultation sessions and fruitful discussions.

Our heartfelt appreciation goes to Sheldon Drobot and Amanda Anderson at the National Center for Atmospheric Research (NCAR) for providing thoughtful suggestions and the data that were essential to complete this project.

Many thanks to Jeff Tilley at the University of North Dakota for developing the project concept and the performance metrics.

Thanks to Brian Mills and Daniel Huang at Environment Canada for sharing their expert knowledge.

Finally, we would like to thank Beata Bielkiewicz and Gervais Arel with the Alberta Ministry of Transportation (MOT) and Olga Kidson with the Nova Scotia Department of Transportation (DOT) who delightfully provided the location information of their road weather information system (RWIS) stations.

EXECUTIVE SUMMARY

The researchers examined five categories of factors that may affect the accuracy of road weather information system (RWIS) pavement temperature forecasts: climatic trends, locational attributes, seasonal/monthly variations, diurnal trends, and forecast length. Five hypotheses were established accordingly and tested using one year of RWIS observations and forecasts obtained from several provinces in Canada.

The RWIS networks were classified into three groups on the basis of the climatic nature of the region in which they are located: maritime, continental, and mixed. The main findings are summarized as follows:

- Pavement temperature forecasts from the maritime climate group had the highest quality and those from the mixed climate group had the lowest quality, both in terms of mean absolute errors (MAEs) and percent of acceptable forecasts (PAFs). The significant performance differences between the regions suggested that the RWIS forecasting performance may be affected by climatic trends, as in, the unique climatic patterns of the regions may have caused the differences in RWIS forecasting performance.

- The correlation between the forecasting accuracy of RWIS stations and their topographical features, such as altitude and amount of vegetation cover, and geographical features, such as distance to local lakes/waters, were investigated within each region. The researchers found that the forecasting performance for the RWIS stations in the maritime climate region near coastal areas had a negative correlation with the distance from a nearby large water body. On the other hand, no significant correlation was found in either the mixed or continental climate groups. It should be noted that a more detailed statistical analysis with additional data is required to determine the exact rationales as to why such a correlation was/was not found and to arrive at a definitive conclusion.

- Daytime forecasts were less accurate than the ones generated for nighttime. Furthermore, as expected, the accuracy of forecasts was found to deteriorate quickly as the forecasting horizon increases.

- Forecast errors were found to exhibit seasonal variations with forecasts for the shoulder months (October and April) tending to be poorer than other months. This could be due to the presence of a mixture of two different weather extremes within those shoulder months.

- There was a clear quantitative relationship between forecast errors and forecasting time and length, suggesting that it is possible to quantify these errors based on the time a forecast is made and the time the condition is to be forecasted (forecasting horizon).

1. INTRODUCTION

The cost of winter road maintenance is substantial in many northern countries such as Canada. Canadian transportation agencies expend more than a billion dollars a year on various winter road maintenance activities (Ye et al. 2009). These activities include the use of large amounts of road salts for deicing and anti-icing, which has become an increasing public concern due to the detrimental effects on the environment and corrosive effects to the infrastructure and vehicles. To reduce the costs of winter road maintenance and the use of salts, many state and municipal governments are seeking ways to improve the efficiency and effectiveness of winter maintenance operations.

One approach to achieving this goal is to improve the decision-making of maintenance operations by making use of real-time information on road weather and surface conditions from road weather information systems (RWISs). RWISs are automated road weather reporting stations that measure various meteorological parameters such as air and pavement temperature, precipitation, and wind speed using a variety of environmental sensors situated in the road and/or on towers adjacent to the road.

RWIS information can help road maintenance personnel develop cost-effective deicing and anti-icing programs to help maintain safe road conditions. This is especially true with information on pavement temperature, which is a major factor that determines if treatment is necessary, when to start maintenance activities, which chemicals to use, and the amount of chemicals to apply. In the context of anti-icing and resource planning, RWIS pavement temperature/condition forecasts are particularly important given the effectiveness of these programs depends on the accuracy and reliability of these forecasts.

Despite the critical importance of RWIS pavement temperature forecasts, there is little systematic information on the performance of various RWIS forecasts from different regions. In particular, what is the overall accuracy of the RWIS pavement temperature forecasts? What is the comparative performance of the RWIS stations from different regions, or models? And, what are the main factors influencing the magnitude of forecast errors? These questions represent the main concern of this research.

Objectives and Scope

The main objective of this research is to evaluate the performance of RWISs in terms of their accuracy to forecast pavement temperatures, and to identify and quantify the effects of the possible factors affecting this performance, such as locational attributes, forecast time and length, and seasonal variation. Due to time constraints and data availability, the scope of this research is limited to examining the following five research hypotheses:

1. ***Climatic patterns:*** RWIS forecasting accuracy is affected by different climatic characteristics (e.g., maritime versus continental)

2. ***Locational attributes:*** RWIS forecasting accuracy is dependent on various geographical and topographical settings on which each station is located

3. ***Seasonal variations:*** RWIS forecasting accuracy would vary by different months (e.g., shoulder months versus non-shoulder months)

4. ***Diurnal trends:*** RWIS forecasting accuracy would change with respect to daily temperature variations (e.g., high during the daytime and low during the nighttime)

5. ***Forecast length:*** RWIS forecasting errors would become larger as the forecast length increases

2. RWIS ROAD SURFACE TEMPERATURE FORECASTING

There are several road temperature forecasting models currently in operation in North America, such as the Model of the Environment and Temperature of Roads (METRo), the Fast All-Season Soil Strength (FASST) model, and the Snow THERmal Model (SNTHERM). Among these three models, the Canadian numerical model, METRo, is being implemented in most of the provinces in Canada.

METRo is superior to the other two competitors in forecasting pavement temperature and surface conditions due to its robust surface condition forecasting capability. Literature that compares the three pavement temperature forecast models states that although METRo requires a longer computing time to generate forecasts than the other two comparison models, it performs better under a variety of different conditions (which is very critical, especially on winter days) and is very easy to acquire, install, and use (NCAR 2007).

Furthermore, METRo was built in a user-friendly environment so that the developers, together with an ever-growing community of end users, can communicate interactively to rectify the problems or issues in a timely manner. In recognizing the model's robustness and versatility, the report from the National Center for Atmospheric Research (NCAR) recommended METRo be integrated into future generations of the Maintenance Decision Support System (MDSS) to maximize the productivity and efficiency in support of winter maintenance operations.

METRo utilizes the surface observation data together with weather forecasts (air temperature, cloud cover, precipitation rate, etc.) to predict how pavement surface temperatures and the accumulative precipitation amount will evolve in liquid and solid forms during a forecasting period (Linden and Petty 2008).

METRo contains three modules: an energy balance module for the road surface, a heat conduction module for the road material, and another module to take account for water, snow, and ice accumulation on the road. The surface energy balance is determined by analyzing various energy fluxes including incoming infrared radiation flux, emitted flux, the sensible turbulent heat flux, and the latent heat flux. Moreover, METRo also incorporates the flux related to the phase change of precipitating water. A one-dimensional heat diffusion equation is utilized to compute the evolution of the subsurface temperature profile by taking two main parameters, heat capacity and ground heat flux, into account (Linden and Drobot 2008).

Two numerical grids are available to METRo and they are selected on the basis of several road characteristics, such as whether they are normal roads or bridges/overpasses (Crevier and Delage 2001). A variable-resolution grid and a uniform-resolution grid are implemented typically to better account for normal roads, and bridges and overpasses, respectively.

Lastly, METRo uses various elements including precipitation, evaporation, and runoff to effectively simulate the accumulation of water, snow, and ice on the road. METRo is also capable of estimating the snow removal that is caused either by traffic or maintenance operations.

Generally, the road condition/temperature forecasting process follows three steps as shown in Figure 1: initialization of road temperature profile, coupling of the forecast with historical observations, and the forecast phase itself.

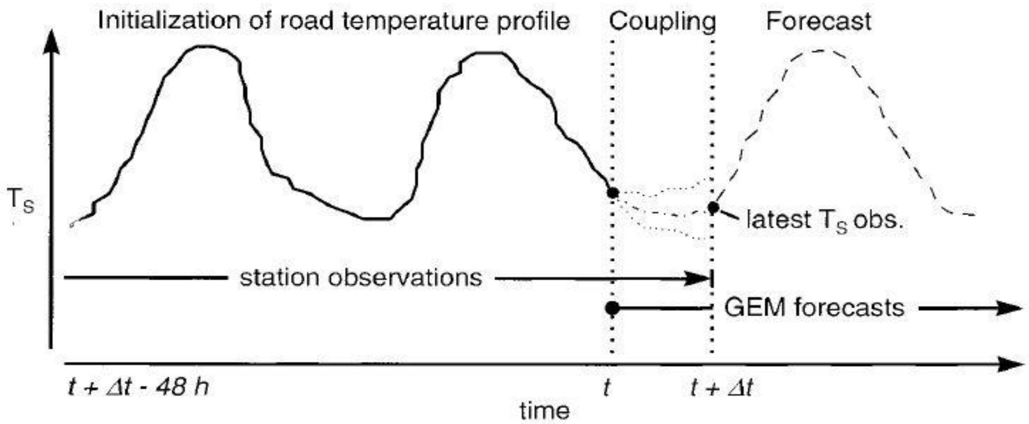


Figure 1. The three stages of the forecast process (adapted from Crevier and Delage 2001)

During the first phase, an initial road temperature profile is created using the historical pavement temperature observations. If METRo detects surface observations that are missing for more than four consecutive hours, an analytical approach is employed to fill in those missing values. Given RWIS forecasts are issued twice a day (i.e., at t = 0000 and 1200 UTC) at different hours than when atmospheric forecasts are issued, METRo uses these overlapped hours, represented as Δt, to create a reliable temperature profile.

Once the profile is created, the profile at the tail of the initialization phase is used in the coupling phase as the initial condition. The main goal in this phase is to adjust the atmospheric forecast generated from the GEM model through iterative processes to actual RWIS observations, thereby creating a highly accurate (i.e., within 0.1 degrees of the actual pavement temperature) "adjusted" pavement temperature to be used in the forecast phase.

At the end of the coupling phase, the forecast phase follows to check the forecasted parameters formed in the coupling phase and to compare them with the latest observations to minimize discrepancies.

Using the RWIS observations along with other meteorological data, METRo delivers near-future pavement temperature forecasts. Although METRo is believed to produce accurate pavement temperature forecasts, its performance on actual datasets has rarely been quantified. Especially in the context of weather forecasting, it is almost inevitable to have discrepancies between observations and forecasts, as there are many uncertainties that affect the accuracy of RWIS temperature forecasts. It is hypothesized that such discrepancies are caused by many factors including but not limited to the following:

- Instrumental errors
- Model errors
- Data errors
- Frequency of forecast issuances
- Climatic trends
- Locational attributes
- Seasonal variations
- Diurnal variation
- Forecast horizon
- Cloud cover, snow cover, precipitation, etc.
- Human errors

Instrumental errors can be, for instance, from the use of miscalibrated sensors. Human errors are always present given meteorologists are often involved in modifying the forecasting parameters and/or models using their discretion based on their past experience.

The models used to forecast may not be robust enough to account for all the possible scenarios that can occur in reality. Performance of models can be significantly poor when dealing with unexpected and extreme weather events such as hurricanes and snowstorms.

Not only can the models create errors, but the data itself can be erroneous due to sensor malfunctions. The errors associated with observation data may consequently affect the overall quality of forecasts because observations are fed into the models to produce the forecasts. Another important factor is the number of forecasts generated in a day, which can significantly influence the overall quality of outputs. Normally, forecasts are issued twice a day unless there is a substantial change in the temperature.

It is equally important to note that the overall forecast accuracy changes with respect to locational attributes and the corresponding climate. Climatic regions can be categorized into the following six groups: moist tropical climates, dry climates, moist climates with mild winters, moist climates with severe winters, polar climates, and highland climates (Ahrens 2009).

Most regions located between latitudes 40^o and 60^o in Canada have a temperate climate, meaning that summer is not as hot as the subtropical climate and milder than the polar climate. There are two sub divisions within temperate climates: maritime and continental. Cities adjacent to the sea or surrounded by the sea have maritime climates as the large mass of the sea takes a much longer time to warm up and cool down. As a result, less temperature variation is observed. On the other hand, continental climates have larger temperature variations given less time is required for air to warm up and cool down due to the lack of significant water bodies nearby.

In addition, many mountainous regions are often found in continents such that weather conditions tend to change dramatically from one hour to the next and from one location to another. For instance, a thunderstorm can occur even during a perfectly clear day, and temperatures can drop significantly from hot temperatures to freezing temperatures within just a few hours.

Furthermore, discrepancies can vary by seasons. It is common that shoulder months such as September/October or April/May experience a large range of temperature variation as the weather changes dramatically within those months due to occurrence of convective disorganized precipitations (Ahren 2009). This variation makes the overall forecasting process much more difficult. Given the temperature variation is relatively higher in shoulder months, the forecasting models experience difficulties in taking high variability of temperature into account.

Temperature variation often shows noticeable diurnal trends. For example, on clear sunny days, the temperature of the road surface can rise at a tremendous rate in the afternoon hours and this can be very difficult to forecast. The effect of clouds and precipitation can make this even more challenging. During evening hours, pavement temperatures plummet from their afternoon peak due to rapid radiational cooling. When the rate of temperature change decreases, so does the forecast error. It is important to emphasize that the diurnal range of temperature relies heavily on the presence of cloud cover as it changes the daily temperature ranges as shown in Figure 2.

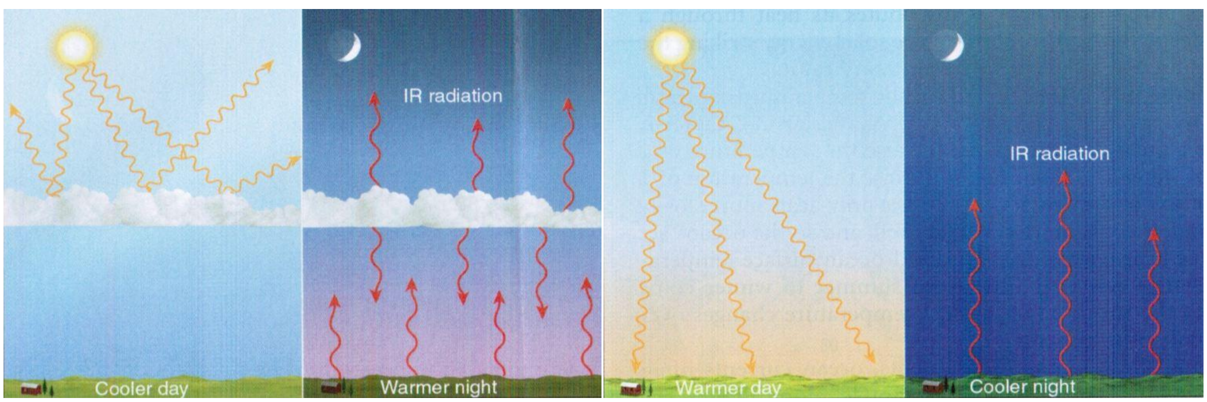


Figure 2. Small daily temperature range (left) and large daily temperature range (right) (adapted from Ahrens 2009)

3. STUDY AREA AND DATA DESCRIPTION

A request for RWIS forecasts and observations was sent to the affiliated members of the Aurora Program. However, only a few agencies were able to provide the requested data because many agencies do not archive their RWIS forecasts.

In addition, data from some agencies were not sufficient to serve the purpose of this project. For instance, Finland provided their RWIS observations and forecasts; however, the dataset covered only three stations for two hours of forecast horizon. For this reason, the analysis for Finland data could not be done in this effort.

For confidentiality, the regions that are included in this study have been labeled and categorized as their climatic types and these are maritime, mixed (i.e., maritime and continental), and continental. Table 1 provides a detailed data description for the three different temperature climates.

Table 1. Summary of the RWIS data

Climate Zones	Number of RWIS Stations Used	Forecast Valid Time (UTC)/ Forecast Period	Data Range
Maritime	9	07:30 &19:30/ Next 12 hours	2009-10-03 to 2010-04-30
Continental/Maritime (Mixed)	15	08:00 & 20:00/ Next 12 hours	2009-10-03 to 2010-04-30
Continental	13	10:00 & 22:00/ Next 6 hours	2009-12-01 to 2010-03-30

- **Climate Zones:** As mentioned earlier, three climatic groups, which exhibit different weather patterns, were used in this study.

  Continental climate exhibits extreme weather patterns. In summer, continental air masses could produce a temperature as high as 40° C; whereas, in winter, arctic air masses could decrease temperatures to as low as -54° C.

  On the other hand, maritime climate can be found in areas that are surrounded by oceans or large bodies of water and the temperature rarely goes above 20° C or below -10° C, so the range of temperature variation is relatively small.

  Climate classified as mixed has both continental and maritime climates. These regions exhibit continental characteristics in terms of their high temperature variations but their temperatures do not vary as much as maritime regions. Such unique characteristics of their climates result in experiencing both continental and maritime climates.

It is also important to note that maritime climate zones, as compared to continental ones, experience more days per year with snow and freezing rain, greater annual snowfall, and a large number of freeze-thaw cycles. Maritime climate zones are also moister and cloudier with both more frost days and weather variability day-to-day than in continental climate zones. However, mountainous areas in continental climate regions are expected to contribute to the degradation of RWIS forecasting qualities as they exhibit highly varying weather patterns as explained earlier in this report.

- **Number of RWIS Stations:** This column shows the number of RWIS stations from which the data were obtained for this analysis.

- **Forecast Valid Time (UTC) and Forecast Period:** Generally, all provinces issue their computed pavement temperature forecasts twice a day. It is important to understand that actual forecasts are usually prepared and generated prior to those hours. Hence, the earlier hours benefit from having the newest forecasts that were just issued. In addition, the data provided by the different regions have different forecast horizons. For a fair comparison, a same forecast horizon of six hours was considered. Note that forecasts have been made on an hourly basis; whereas, observations have been made on approximately a 20 minute interval (i.e., three observations per an hour).

- **Data Range:** The data range for forecasts and observations are shown in this column.

4. EVALUATION METHOD

Figure 3 shows the procedure proposed to evaluate the hypotheses for this research and to develop a benchmark. Detailed steps are described in the following section.

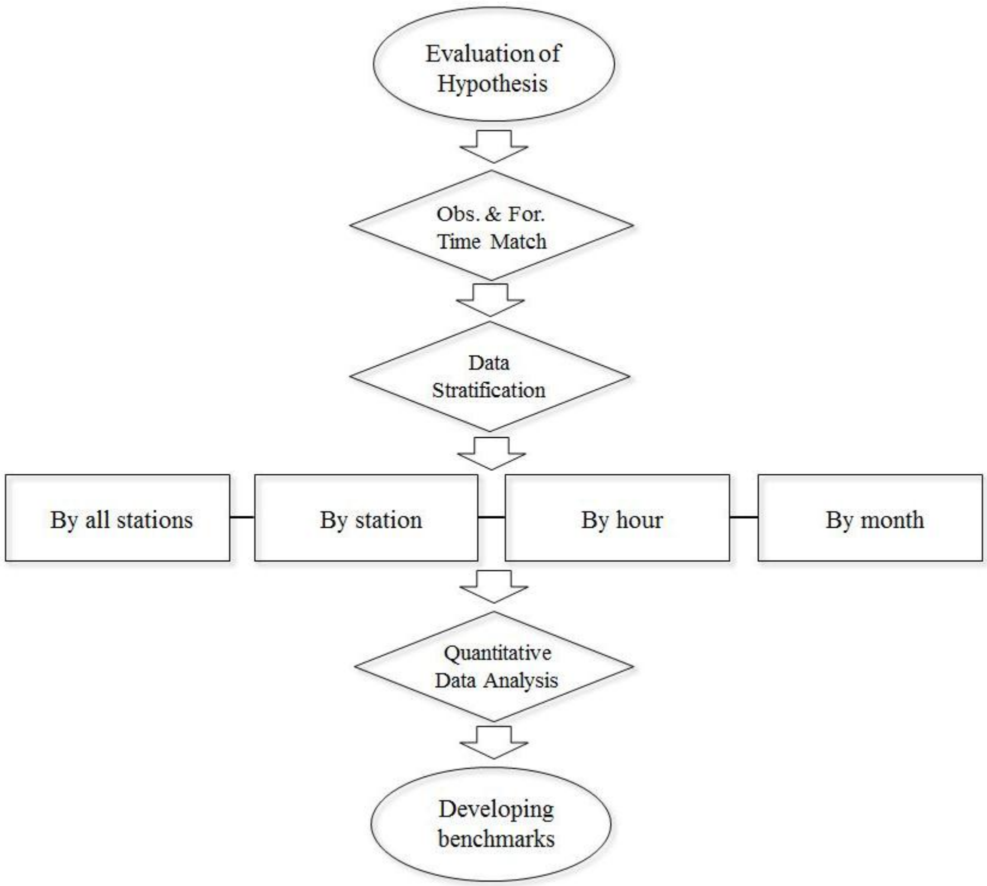


Figure 3. Workflow of the proposed study

4-1. Match of Observations and Forecasts

Pavement temperature data from two different climate groups (maritime and mixed) have a 20 minute temporal resolution; whereas, forecasts are made for every hour over the forecasting horizon. As a result, there are three observations available for each forecast and the time for 20 minute interval observations had to be matched with the corresponding forecast hour.

For example, if the observations were made at 05:05:00, 05:25:00, and 05:45:00, the observation time, 05:05:00, would be selected as the closest match of the forecast time, 05:00:00 to compute the absolute temperature difference. This closest-time matching was done as long as the observations were within 20 minutes of the model valid time (i.e., forecast time).

This matching was not done for the continental group, which provided the forecasts and observations already matched with the model valid timestamp in the raw data. It is important to note that such an exception on time-matching may introduce bias, given it is not known what was done to match the forecast and observation hours.

4-2. Data Stratification

Once all the data (forecasts and observations) were paired correctly for the timeframe available, they were stratified in four different ways so that the established individual hypotheses could be evaluated.

4-2-1. Data Stratification by All Stations and All Hours

Data stratification was done in a way to produce suitable performance measures to describe the overall quality of pavement temperature forecasts. To fulfill this, all available stations and hours for each climate group were stratified so that performance measures representing each climate group could be constructed to compare the accuracy of pavement temperature forecasts between the three different climate regions. This stratified dataset is to be used to test the hypothesis of the possible dependency of RWIS forecasting errors on the climatic pattern of the region in which the RWIS stations are located.

In addition, forecasting errors obtained from three different groups will be compared with a performance standard called acceptable tolerance level. Given the continental climate group only provided four months and 6 hour forecast data, the same periods and same forecast hours for the two other climate groups, maritime and mixed, were utilized to conduct a fair comparison.

4-2-2. Data Stratification by Station

Data stratified by station were intended for an analysis to show the variability, if any, between stations, regardless of time. This analysis is to identify the locational variability with respect to individual stations situated in different types of geographical and topographical settings (i.e., whether the station is sited in a high-attitude mountainous area or near a maritime area).

Mountainous areas have very complex terrain that may cause the climate to vary over short distances. During daytime when sunlight heats the mountains, the temperature of the mountain surface could rise quickly, forcing air to move upward along the slopes to eventually form anabatic clouds (Stull 2000). This unique phenomenon increases the amount of precipitation on the windward side and creates dry areas on the opposite side of the mountain. Thus, mountain areas typically have very localized weather patterns.

On the other hand, as discussed previously, maritime areas are affected by the thermal mass of large water bodies that trigger slower heating and cooling. As a result, maritime regions have a relatively low temperature variation when compared to mountainous areas. However, there tends to be more precipitation in maritime regions, which produces more days with snow and freezing rain.

The hypothesis on the effect of locational attributes on RWIS forecasting performance would be evaluated by examining the variability of forecasting errors between stations.

4-2-3. Data Stratification by Hour

Data stratification by hour was intended to show the variability of RWIS forecasting performance within a diurnal cycle, regardless of station. Each hour encompasses all minutes in that hour (i.e., 00 would be 000000-00:59:59, 01 would be 01:00:00-01:59:59, and 23 would be 23:00:00-23:59:59). If the model valid time began with 00, it was put in the 0 hour bin, 01 in the 1 hour bin, and so on.

It is important to note that the local times have been converted to Coordinated Universal Time (UTC) to compare the resulting statistics between all climate groups. Such stratification is to show the daily variation of temperature discrepancies as well as the forecast-aging trends.

The greatest variation in daily temperature occurs at the surface of the earth (Ahrens 2009). This also implies that the temperature at the surface undergoes a high variation in daily temperature (i.e., T_{max}-T_{min} is high). This typical diurnal variation may be reflected in the quality of pavement forecasts.

In addition, forecast-aging trends can be verified by showing the discrepancies that may become greater as the forecast horizon increases. Given the data provided by continental stations were insufficient (i.e., only 6 hour forecasts were available) to thoroughly examine the effect of diurnal variation, this group was excluded from the comparison.

4-2-4. Data Stratification by Month

Data stratification by month is to show the variability of the accuracy of the forecasts made over different months. As described earlier, weather over the shoulder (transitional) months usually undergo a variety of different weather patterns due to seasonal changes. For this, it is hypothesized that pavement temperature forecasts are (partially) affected by the extreme seasonal weather changes during shoulder months. By stratifying the data on a monthly basis, changes in forecasting discrepancies with respect to seasonal variations can be determined.

Given the data provided by continental provinces were insufficient (data for only four months were available) to comprehensively observe the effect of diurnal variation, the continental climate group was once again excluded from the comparison.

4-3. Performance Metrics

Two metrics are selected to capture the forecasting performance of a RWIS. The first one is the mean absolute error (MAE) defined by Equation 1:

$$MAE = \frac{1}{n}\sum_{i=1}^{n} | f_i - o_i | \qquad \textbf{(1)}$$

where: n, f, and o are the total number of matched pairs between forecasts and observations, forecast, and observation, respectively. MAE is a conventional measure of discrepancy used widely in science and engineering fields.

The second performance measure, which is currently used by the RWIS service providers to report the overall performance of their RWIS pavement temperature forecasts, is called percent of acceptable forecasts (PAFs), as shown in Equation 2:

$$PAF = 100\,\% \times \frac{Number\ of\ Forecasts\ Made - Accuracy\ Exceptions}{Number\ of\ Forecasts\ Made} \qquad \textbf{(2)}$$

where: number of forecasts made is equivalent to n noted above, and accuracy exceptions are determined based on the following two criteria:

- 2° C or greater difference when the observed pavement temperature is between -3° C and + 3° C
- 3° C or greater difference when the observed pavement temperature is between -20° C and -3° C or between +3° C and +10° C

After obtaining the MAE and the PAF that describe the overall quality of pavement temperature forecasts, a t-test is conducted to make a statistical inference between every pair of climate groups to confirm that the calculated performance measures, either by MAE or PAF, are truly reliable and thus statistically significant. Hypothesis testing for the difference between two means using MAE and the difference between two proportions using PAF were conducted. The underlying hypotheses can be formulated as follows:

$$H_o : \mu_1 = \mu_2 \ or \ p_1 = p_2 \qquad H_1 : \mu_1 \neq \mu_2 \ or \ p_1 \neq p_2 \qquad \textbf{(3)}$$

where: μ_1 and μ_2 are population means (MAE); whereas, p_1 and p_2 are population proportions (PAF).

Then, hypothesis testing on two independent variables can be performed using the following test statistics:

$$t_{calc} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \ (means) \tag{4}$$

$$t_{calc} = \frac{\bar{p}_1 - \bar{p}_2}{\sqrt{\frac{\bar{p}_1(1-\bar{p}_1)}{n_1} + \frac{\bar{p}_2(1-\bar{p}_2)}{n_2}}} \ (proportions) \tag{5}$$

where: $\bar{x}_1$ and $\bar{x}_2$ are the sample means of MAE, S_1 and S_2 are the sample standard deviations of MAE, and $\bar{p}_1$ and $\bar{p}_2$ are the sample proportions of PAF.

The level of significance was set to $\alpha=0.05$ indicating a 95 percent confidence interval. Based on the outcome of t_{calc}, a decision can be made if $|t_{calc}| > t_{crit}$, then reject H_o in favor of H_1; otherwise, reject H_1 (i.e., fail to reject H_o). After calculating necessary statistics, several benchmarks, which depict the observed variations of forecast discrepancies, would be developed using the nonlinear regression analysis, accordingly.

5. RESULTS AND DISCUSSIONS

5-1. Effect of Climatic Pattern

In this section, the data that has been stratified by all stations are used to see the overall quality of RWIS pavement temperature forecasts. The calculated statistics are used subsequently to evaluate the hypothesis stating that different climates or climatic patterns would affect the forecasting ability. Figure 4 illustrates pavement temperature forecast MAE and pavement temperature accuracy PAF calculated for each individual province.

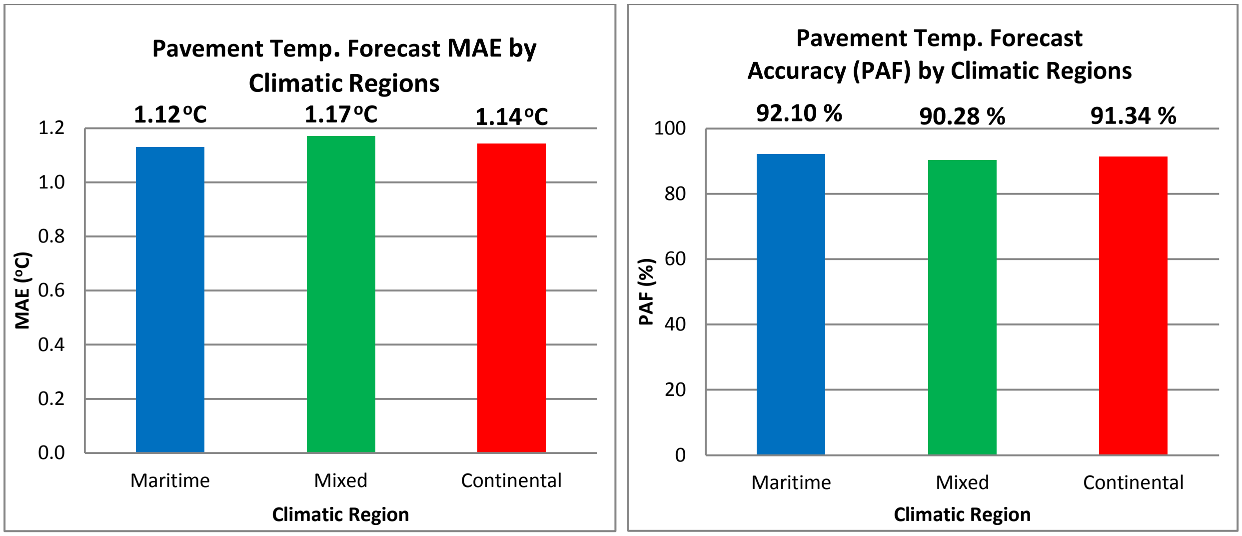


Figure 4. Pavement temperature forecast MAE by climate region (left) and pavement temperature forecast accuracy performance by climate region (right)

Table 2 provides a summary of the statistics.

Table 2. Summary of calculated statistics

Climate zones	Pavement MAE/ Tolerance (^{o}C)	Accuracy Performance PAF (%)
Maritime	1.12 / 3.00	92.10
Mixed	1.17 / 3.00	90.28
Continental	1.14 / 3.00	91.34

From Figure 4 and Table 2, it can be seen that the overall quality of RWIS pavement temperature forecasts are all within its acceptable tolerance levels set by each corresponding province. The maritime climate group has an overall discrepancy of 1.12^{o}C, having the lowest error, while the continental and mixed climate groups have an overall discrepancy of 1.14^{o}C and 1.17^{o}C, respectively. In addition, PAF values for the three different climate groups show similar results to their corresponding MAE values.

One important inference that can be made by observing the resulting statistics is that there is a correlation between the amount of pavement temperature discrepancies (MAE/PAF) and climatic patterns. It was hypothesized earlier that different climatic trends would affect the overall forecasting capability. As explained previously, a maritime climate region usually exhibits relatively stable weather patterns due to the presence of large water masses, thereby producing fewer forecast errors as the temperature does not fluctuate in a high range. This could provide a plausible explanation on maritime climate having lower forecasting errors while continental climate regions had higher forecasting errors.

Furthermore, given the mixed climate group is partially maritime and partially continental, it was found to produce the highest errors of all. One reasonable speculation for such a phenomenon could be that having a mixture of both climates makes it even more challenging to produce accurate forecast values. Thus, it can be concluded that RWIS pavement temperature forecasts are affected by and correlated with climatic trends.

However, it is equally important to note that such marginal differences in the overall performance could have been caused by many factors. For instance, as mentioned earlier, the continental climate group provided forecasts and observations with model valid times already being matched. It is not known as to what internal process was undertaken to match the time, but it can be a critical factor that influences the overall quality of forecasts.

Another hidden factor can be the underlying methodology being employed by the forecasters. For instance, information as to which forecasting models are used and the degree of meteorologist intervention during the forecasting process (i.e., most of forecasts are made using semi-automatic schemes) was not fully disclosed, thereby making it very difficult to draw conceivable conclusions. A further investigation is necessary to verify the concealed grounds.

By observing the comparison statistics, the hypothesis stating that different climatic trends would affect the overall forecasting capability can still be proved, given the three different climate groups have produced the different values. To make such claims and inferences based on the calculated values for MAE and PAF, it is important to see if the outcomes are truly reliable. To do so, the t-test is done as summarized in Table 3 for mean (MAE) and proportion (PAF).

Table 3. t-test results using means (MAEs) and proportions (PAFs)

t_{calc} using means				t_{calc} using proportions			
Climate	**Continental**	**Maritime**	**Mixed**	**Climate**	**Continental**	**Maritime**	**Mixed**
Continental	0	4.65	-7.58	**Continental**	0	2.70	-2.98
Maritime	-4.65	0	-10.51	**Maritime**	-2.70	0	-4.68
Mixed	7.58	10.51	0	**Mixed**	2.98	4.68	0

Given all $|t_{calc}|$ values are greater than t_{crit} = 1.96 (i.e., 95 percent confidence interval), the null hypothesis (H_0) is rejected. Thus, from the empirical investigations, it can be concluded that the difference in the accuracy of the pavement temperature forecasts between the three different climate groups are statistically significant. However, it should be noted that the difference appears to be insignificant from a practical point of view when considering their marginal differences.

5-2. Effect of Locational Attributes

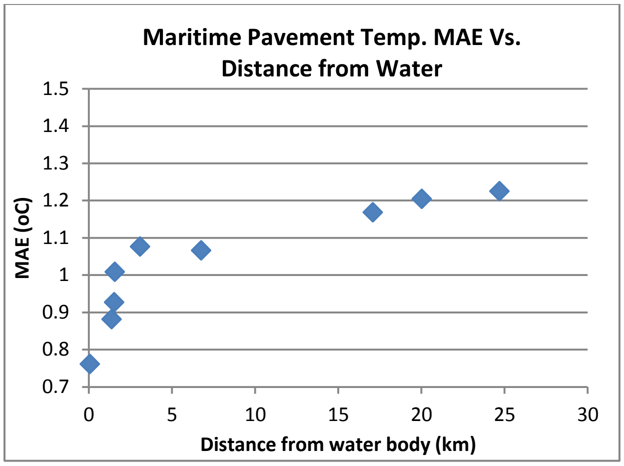
To capture the effect on forecast accuracy caused by locational attributes related to geographical and topographical settings, all stations need to be analyzed individually. For this reason, the data were stratified by individual station.

To carry out the analysis, three different graphical information system (GIS) layers from the map library at the University of Waterloo were used to extract the following information for all four provinces:

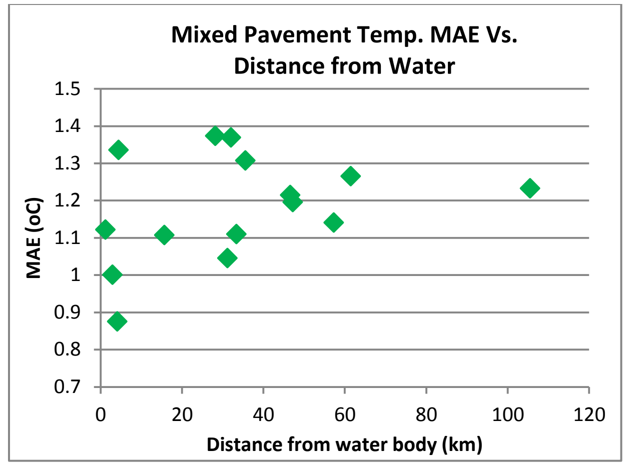
- Local and regional lakes and seawater
- Digital Elevation Model (DEM)
- Parks and forests

Again, the quality of the forecasts made by individual RWIS stations could be affected by three locational features, including existence of a nearby large water body, altitude, and land use.

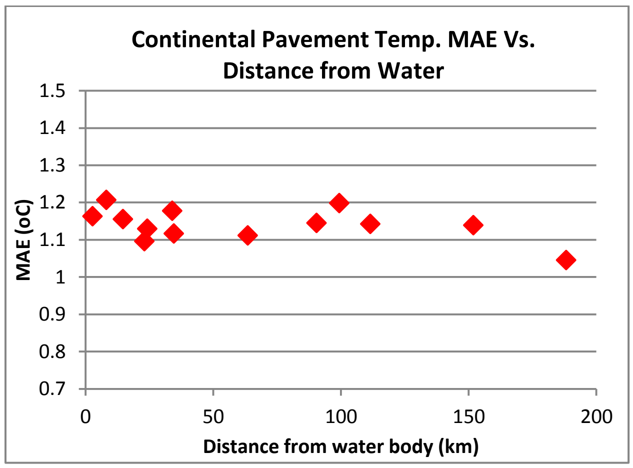
Figure 5 summarizes the RWIS forecast errors of individual stations arranged in order of distance from a water body.



(a)



(b)



(c)

Figure 5. Data stratified by individual stations in order of distance from water body

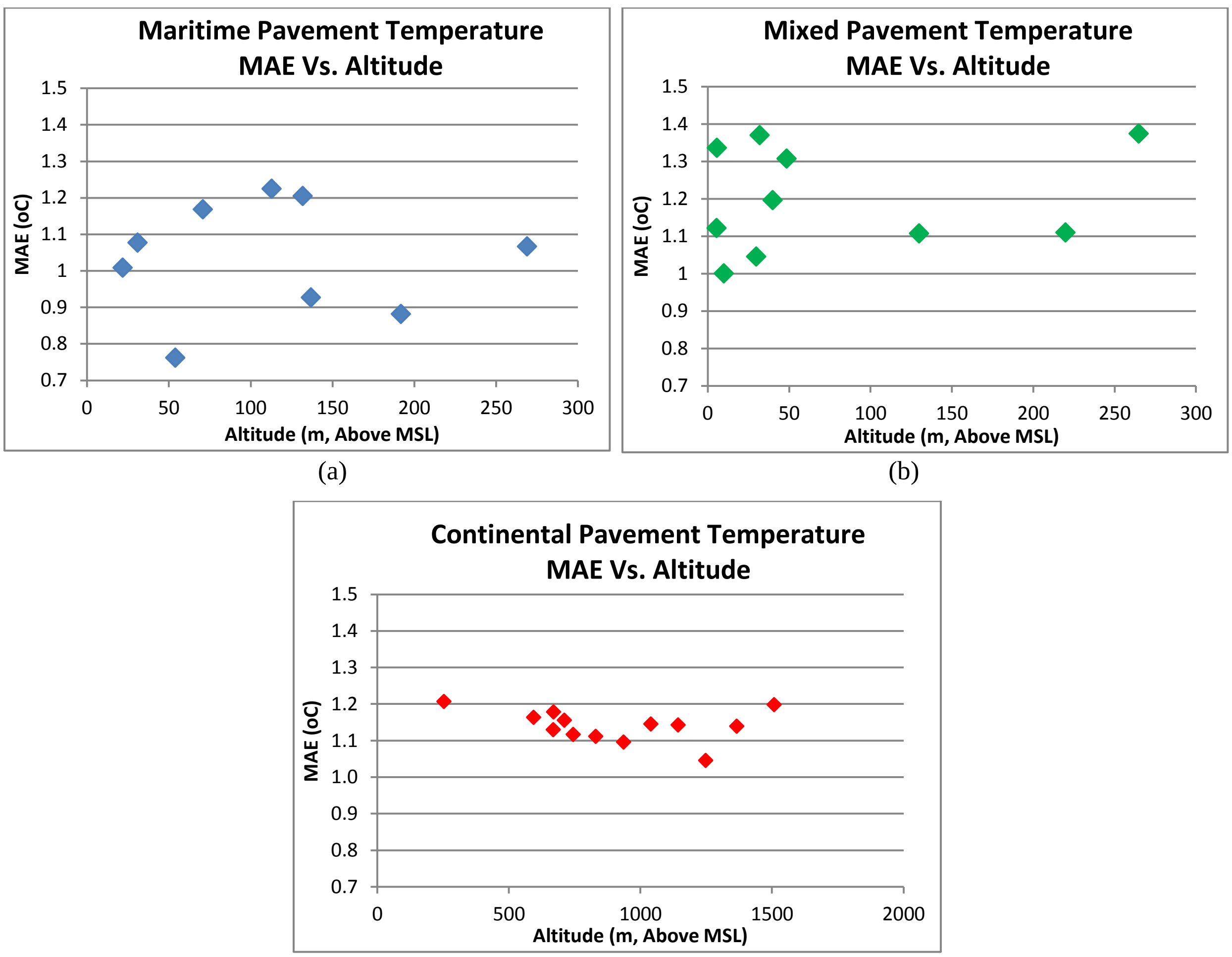
By analyzing the statistics shown in Figure 5, a correlation was found only in the maritime climate group where forecast errors (MAE) increased by the distance from a water body. A possible inference as to why such a correlation exists with respect to the location of RWIS stations is that, from a meteorological point of view, a station situated nearby a large water body is expected to exhibit relatively stable temperature variations.

As mentioned earlier, air temperature can be a critical factor that controls the overall quality of road surface temperature estimation when fed into METRo to produce the subsequent pavement temperature forecasts. Making the inference that stations located on the coastlines tend to produce lower forecasting errors is also supported by the maritime climate group results shown in Figure 5(a).

On the other hand, the mixed and continental climate groups did not seem to have any correlation in this regard. This is probably because the continental group is less affected by lake effect as most of the RWIS stations are situated inland. Notice in the results shown for the mixed climate group in Figure 5(b) that the errors are not distributed uniformly, but randomly. Such a random error distribution may have occurred due to the mixed characteristics of maritime and continental climates, which make them much more challenging to determine the exact confounding grounds.

Another effect of locational attributes can be that air and surface temperatures near parks and forests usually vary in a significantly different pattern from those at an urban center. Given the data that were available did not contain much data extracted for such areas, it was not possible to show the proposed correlation.

The quality of forecasts can also be affected by elevation or land use. Highways located either nearby or in a high-altitude mountainous area are more likely to experience dramatic temperature changes, making it more difficult to predict their surface temperatures. Figure 6 illustrates the MAE values plotted against their corresponding altitudes.



(a) (b) (c)

Figure 6. RWIS MAE versus RWIS station altitude

As can be seen, no correlation with respect to altitude values was found. However, it should be noted that the overall pavement temperature forecast accuracy for the three climate groups are all within the acceptable tolerance level, so it is possible that the magnitude of correlations with different locational attributes may be too small to be captured within the existing dataset. Another possibility is the existence of other factors that may have confounded the underlying relationships. Nonetheless, the exact hidden grounds need to be further scrutinized by using more station data, for instance, or by using the data stratified by routes to see if there are any locational correlations.

5-3. Effect of Diurnal Trends

The forecasting errors for the RWIS stations from the individual regions were stratified by hour to examine if any diurnal and time trends exist. Figure 7 shows the MAE by forecasting period.

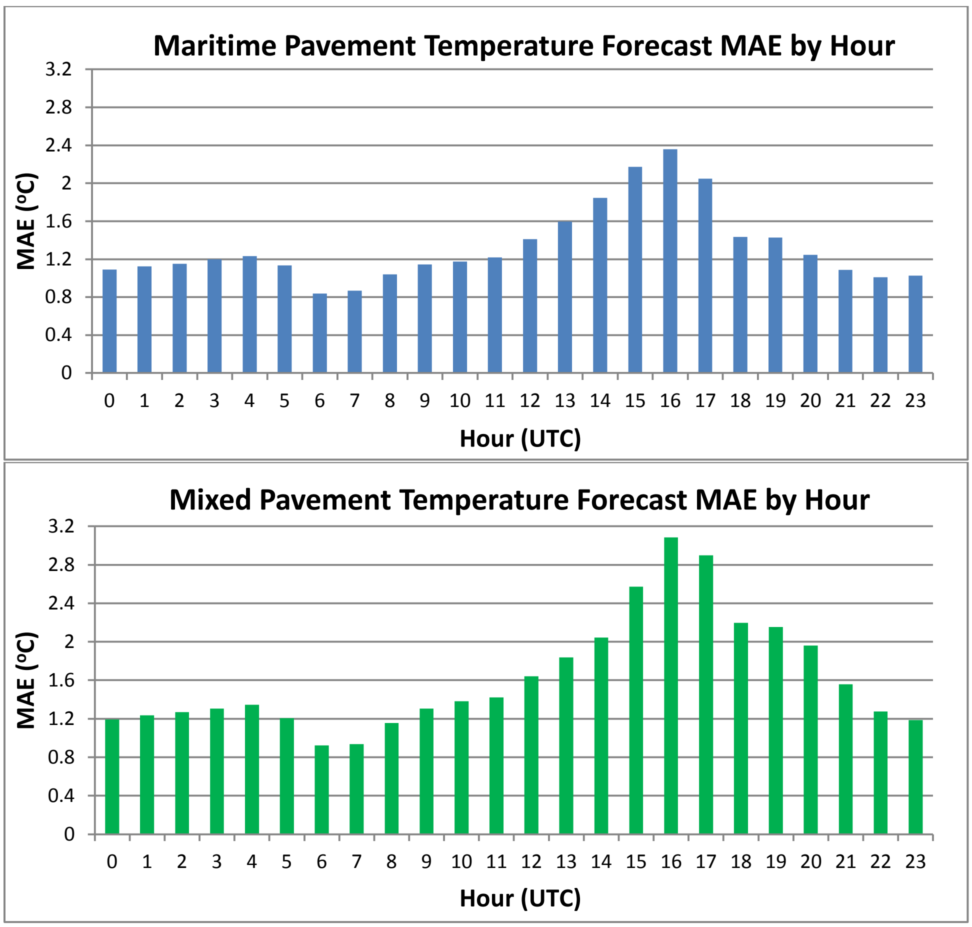


Figure 7. Data stratified by hour (UTC) for maritime (top) and mixed (bottom) climate groups

Note that all local times have been converted to UTC to have a valid comparison between different climate groups. The continental climate group is not included in this comparison due to insufficient data availability.

As mentioned in the previous sections, forecasts are made and issued (or become available) at 08:00 and 20:00 for the mixed climate group and 07:30 and 19:30 for the maritime climate group. Then, the lowest errors are expected to occur for the hours close to the time when forecasts were made (08:00 and 07:30 for the mixed and maritime climate groups, respectively). However, with the data we obtained, it was observed that the least errors occurred either at 06:00 or 07:00 and at 22:00 or 23:00, which gives approximately an average of one hour time difference.

To explain this phenomenon, the actual forecasting process needs to be understood. It was found that there is usually a time lag between when the forecasts are generated and when the forecasts are issued. Given the forecasts are typically generated prior to the official valid forecast time, the earlier hours can benefit from having the newest forecasts that were just issued. Also, for example, in the maritime climate group, the greatest jumps are seen between 05:00 and 06:00 and again between 17:00 and 18:00. These jumps are due to the forecaster adjusting the forecast to match the initial atmospheric and pavement conditions just before the forecasts are released.

Another phenomenon to clarify is that even after 20:00, MAE values continue to decrease. Such an abnormal trend could have been caused by a diurnal variation (as also supported by the graphs in Figure 8) where during the daytime, temperature tends to fluctuate greatly; whereas, temperature tends to fluctuate less during the nighttime.

Then, given the forecasts that are released in the morning have been calculated using the "less varying" historical observations, not much effort is needed in the forecasting phase to eliminate the discrepancy. Thus, it took only one hour to arrive at "convergence," which occurred at 07:00 for the maritime climate group. On the other hand, the forecasts that are released in the afternoon must have used the "highly varying observations" collected during the daytime, thereby taking longer adjustment times. As a result, it is speculated that an additional two to three hours are required until the pavement forecast error reaches its minimum at 22:00.

Finally, it needs to be pointed out that forecasting errors caused by climatic variations are also observed from Figure 7. The extreme MAE difference (i.e., T_{max}-T_{min}) is larger for the mixed than the maritime climate group. This is primarily due to the fact that these two groups have different climate types and hence produce different MAE values (See section 5-1).

5-4. Effect of Forecast Length

It is known that the quality of forecasts tends to deteriorate as the forecasting horizon increases (Crevier and Delage 2001), as was also confirmed in section 5-3. This means that forecast errors generally increase with the length of the forecast horizon. To investigate the error behavior, two separate models for two groups were developed for the hours between 07:00 and 16:00 using the hourly-stratified data. As shown in Figure 8, the forecasting performance deteriorates quickly.

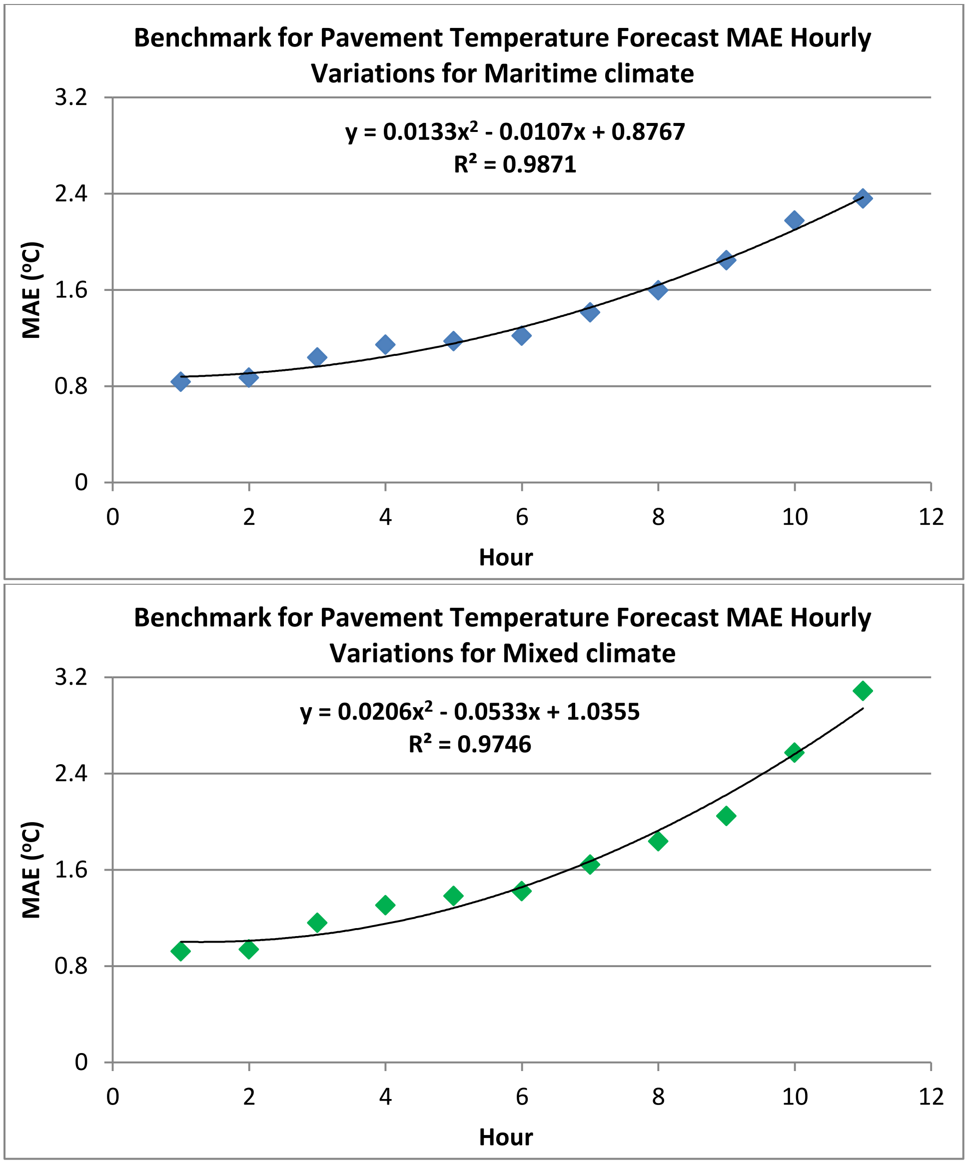
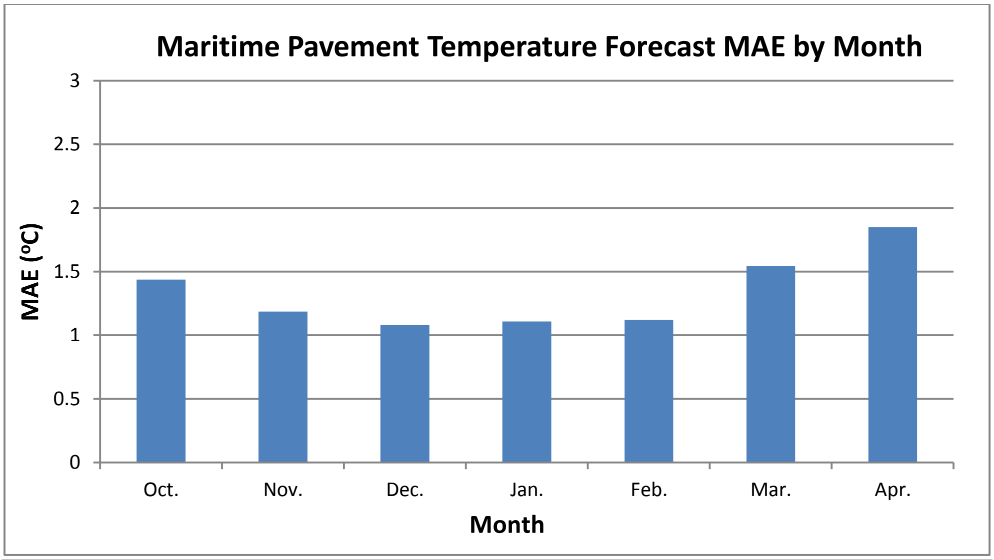


Figure 8. Forecast error as a function of forecasting lead time for maritime (top) and mixed climate (bottom)

For both climate groups, a quadratic function was found to fit the error pattern the best. High R^2 values from both models indicate that the error terms were very well explained by the models. The model coefficients were determined as shown in Figure 8. Note that all of the calculated coefficients were statistically significant at the 5 percent level.

5-5. Effect of Seasonal Variations

The pavement temperature data for both observations and forecasts that were stratified by month are summarized in Figure 9.



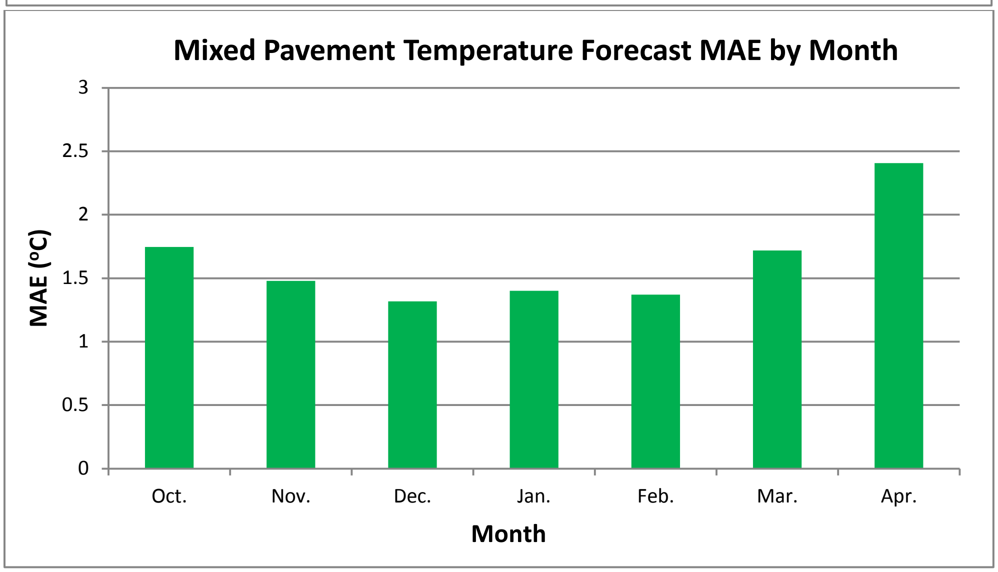


Figure 9. Monthly pavement temperature forecast MAE for maritime (top) and mixed climate (bottom)

As can be seen, there is a clear pattern associated with the monthly pavement temperature forecast MAE. The discrepancies tend to be relatively higher for shoulder months (i.e., October and April) than for non-shoulder months (i.e., December, January, and February). This could be because the weather patterns typically vary in a great range over these shoulder months, making it very difficult to develop accurate forecasts. Although the magnitudes of MAE are different for individual climate types, a general pattern appears showing that October and April have higher MAE values than those of in-between months.

The monthly variation patterns of the forecast errors for the two climate groups can be fit into quadratic functions using monthly-stratified data. As shown in Figure 10, the quadratic functions were superimposed onto the data.

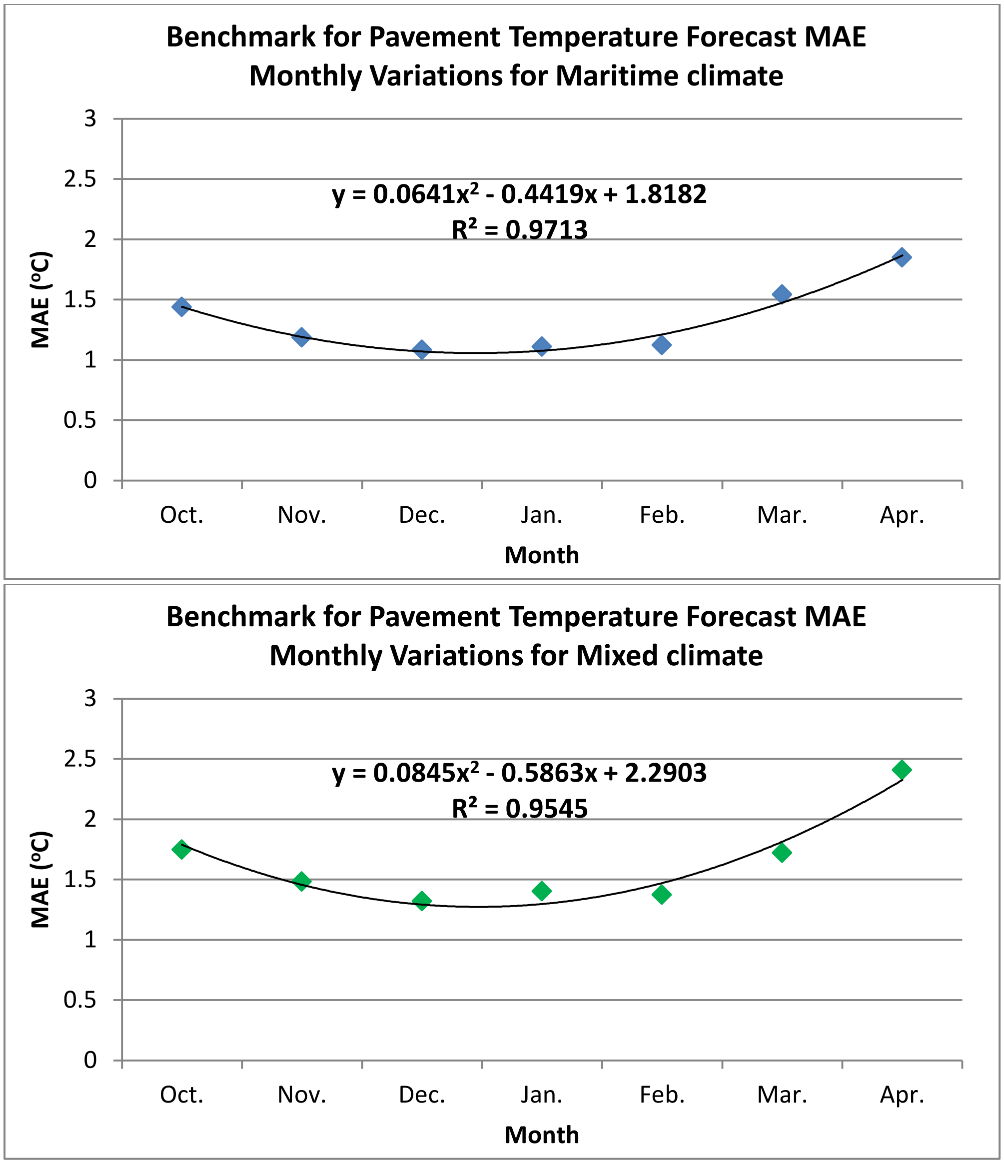


Figure 10. Benchmark for monthly variations for maritime (top) and mixed climate (bottom)

Using the derived equations shown on the two charts, the expected deviation of the forecast using the observations can be estimated. However, it needs to be pointed out that such results came from analyzing the limited number of data in terms of the data range (duration), number of stations, and so on. Thus, further analysis with many more data is essential to develop a more rigorous model.

6. CONCLUSIONS AND RECOMMENDATIONS

In this research, we examined factors that may affect the accuracy of RWIS pavement temperature forecasts: climatic trends, locational attributes, seasonal/monthly variations, diurnal trends, and forecast length. The main findings are summarized as follows:

- Pavement temperature forecasts from the maritime climate group had the highest quality and those from the mixed climate group had the lowest quality, both in terms of MAEs and PAFs. The MAEs were found to be 1.12^{o} C, 1.14^{o} C, and 1.17^{o} C for the maritime, continental, and mixed climate groups, respectively, while the corresponding PAFs were 92.10, 91.34, and 90.28 percent, respectively. The significant performance differences between the regions suggested that the RWIS forecasting performance may be affected by climatic trends, as in, the unique climatic patterns of the regions may have caused the differences in RWIS forecasting performance.

- The correlation between the forecasting accuracy of RWIS stations and their topographical features, such as altitude and amount of vegetation cover, and geographical features, such as the distance to local lakes/waters, were investigated within each region. The researchers found that the RWIS stations in the maritime climate region that are located near coastal areas produce fewer forecast errors and that the errors became larger moving farther away from the water. On the other hand, no correlation was found in either the mixed or continental climate groups. It should be noted that a more detailed statistical analysis with additional data is necessary to determine the exact rationales as to why such a correlation was/was not found and to arrive at a definitive conclusion.

- Daytime forecasts were less accurate than the ones generated for nighttime. Furthermore, as expected, the accuracy of forecasts was found to deteriorate quickly as the forecasting horizon increases.

- Forecast errors were found to exhibit seasonal variations with forecasts for the shoulder months (October and April) tending to be poorer than other months. This could be due to the presence of a mixture of two different weather extremes within those shoulder months.

- There was a clear quantitative relationship between forecast errors and forecasting time and length, suggesting that it is possible to quantify these errors based on the time a forecast is made and the time the condition is to be forecasted (forecasting horizon).

It must be emphasized that this research represents an initial effort to benchmark RWIS performance for monitoring and forecasting of road weather and surface conditions. Further research is needed in the following specific directions:

- A similar analysis should be performed on RWIS data from other regions and countries to validate the findings from this study and investigate other potential factors that affect RWIS performance. For example, knowing whether the forecasts are generated in an automated or manual mode, and the extent of meteorologists' intervention could provide a vital source that contributes to the quality of forecasts. The data can also be stratified by each station and by each hour so that the variations of forecasts for each station on an hourly basis can be studied.

- A further analysis is needed to identify spatial variation patterns of the RWIS measurement and forecasting errors, which are required to determine RWIS requirements such as density and location.

- Other statistical performance measures, such as root mean squared error (RMSE) or mean absolute percentage error (MAPE), are recommended to be utilized together with MAE to cross-validate the variation in the errors of the RWIS forecasts. Should there be a greater difference in two measures, the greater the variance there is in the individual forecasting errors.

- Although a strong correlation was not found between forecasting errors and various locational attributes, it is highly recommended to analyze the data in a more localized form (i.e., use of data that are stratified by routes) to investigate other error-contributing factors, which may have confounded the underlying relationships. For instance, more advanced models should be assessed to capture the joint effect of the factors.

- RWIS technology has become an increasingly ubiquitous and indisputable tool for winter road maintenance agencies to optimize their maintenance operations. It would be valuable to develop an international capacity and platform for automated quality assurance and performance benchmarking of RWIS measurements and forecasts. This could be an Internet-based data and application service system connected to different RWISs around the world. Another approach would be requiring the vendors to report all relevant performance measures.

By further investigating these untested factors together with the results drawn in this report, it will ultimately contribute to making more informed decisions in selecting RWIS vendors and locating RWIS stations.

REFERENCES AND BIBLIOGRAPHY

Ahrens, C. D. (2009). *Meteorology Today: An Introduction to Weather, Climate, and the Environment.* Belmont, CA: Brooks/Cole, Cengage Learning.

AMEC Earth & Environmental. (2007). *RWIS Forecasting Performance Report*, October 2006 to April 2007. Nova Scotia Department of Transportation & Public Works.

Crevier, L. P., and Delage, Y. (2001).” METRo: A new model for road-condition forecasting in Canada.” *Journal of Applied Meteorology*, 40: pp. 2026-2037.

Linden, S., and Petty, K. R. (2008). The Use of METRo (Model of the Environment and Temperature of the Roads) in Roadway Operation Decision Support Systems. 24th Conference on Interactive Information and Processing Systems (IIPS), American Meteorological Society. New Orleans, LA.

NCAR. (2007). *A Comparison of Road Temperature Models: FASST, METRo, and SNTHERM*. Report version 2.0. National Center for Atmospheric Research.

Stull, R. B. (2000). *Meteorology for scientists and engineers.* Belmont, CA: Brooks/Cole.

Ye, Z., Strong, C., Fay, L., and Shi, X. (2009). Cost benefits of weather information for winter road maintenance. Montana State University, Western Transportation Institute.

White, S. P., Thornes, J. E., and Chapman, L. (2006). *A Guide to Road Weather Systems*. Available at: http://www.sirwec.org/documents/rwis_web_guide.pdf