Canadian Journal of Civil Engineering
Revue canadienne de génie civil

Spatiotemporal Variability of Road Weather Conditions and Optimal RWIS Density – An Empirical Investigation

Journal:	<i>Canadian Journal of Civil Engineering</i>
Manuscript ID	cjce-2017-0052.R1
Manuscript Type:	Article
Date Submitted by the Author:	20-Apr-2017
Complete List of Authors:	Kwon, Tae J.; University of Alberta, Civil and Environmental Engineering Fu, Liping; Dept. of Civil Engineering,
Is the invited manuscript for consideration in a Special Issue? :	N/A
Keyword:	RWIS, spatial simulated annealing, kriging, variograms, optimization

**Spatiotemporal Variability of Road Weather Conditions and Optimal RWIS
Density – An Empirical Investigation**

Tae J. Kwon¹

Assistant Professor

Department of Civil & Environmental Engineering

University of Alberta,

Edmonton, AB, T6G 2R3, Canada

Email: tjkwon@ualberta.ca

Liping Fu

Professor

Department of Civil & Environmental Engineering,

University of Waterloo, Waterloo, ON, Canada, N2L 3G1

Intelligent Transportation Systems Research Center,

Wuhan University of Technology, Mailbox 125, No. 1040 Heping Road, Wuhan, Hubei 430063.

Email: lfu@uwaterloo.ca

¹ Corresponding author

26 ABSTRACT:

27 This paper presents a study aiming at understanding the relationship between the
28 spatiotemporal characteristics of road weather conditions and a number of RWIS stations using
29 real-world case studies. Semivariogram models are constructed to determine the spatial
30 variability of road weather conditions, especially, autocorrelation range which describes a
31 separation distance at which the measurements are no longer correlated to each other. An optimal
32 RWIS density is then determined through an optimization process that minimizes the total
33 inference errors across the underlying road network. The findings suggest that the regions with
34 less varied topography tend to have a longer spatial correlation range than the regions with more
35 varied topography. The study further reveals that the range of spatial autocorrelation is related to
36 the optimal density of RWIS network – the region with a longer range requires a less number of
37 RWIS stations, than the region having a shorter range.

38 **Keywords:** RWIS, spatial simulated annealing, kriging, variograms, spatiotemporal variability

39 1. INTRODUCTION

40 Road weather information systems (RWIS) can be defined as a combination of advanced
 41 technologies that collect, transmit, process, and disseminate road weather and condition
 42 information to help winter road maintenance (WRM) personnel make timely and proactive
 43 winter maintenance decisions. An RWIS collects data using environmental sensor stations (ESS),
 44 makes roadway-related nowcasting and forecasting, and disseminates the resulting information
 45 to the users. An RWIS ESS (henceforth called an RWIS station) connotes a stationary station,
 46 which usually consists of atmospheric, pavement, and/or water-level monitoring sensors.
 47 Measurements from a typical RWIS station includes air/surface temperature, visibility, wind
 48 speed, and road surface condition data, collected at 10-20 minute intervals. Availability of these
 49 data not only enables the use of cost-effective WRM, but also helps motorists make more
 50 informed decisions for their travel. Likewise, highway maintenance personnel are better assisted
 51 in their decision making process to facilitate timely and effective deployment of labor, material,
 52 and equipment during the course of current or upcoming snow storm events. For example,
 53 information on road surface temperature (RST) allows winter maintenance personnel to predict
 54 when and where ice or frost is likely to form on a roadway (Shao et al. 1997)

55 Despite these benefits, RWIS stations are expensive to deploy and maintain. The average
 56 installation cost could be as high as \$100k per station (depending on the type and number of
 57 sensors included). With budget limitation, RWIS stations must be installed strategically so that
 58 the total costs are minimized while the data from the network is still representative enough under
 59 various road weather and conditions (Manfredi et al. 2005). Hence, designing of a suitable RWIS
 60 network is of critical importance, determining the cost-effectiveness of the end total system. A

few studies have been conducted in the past to develop models for locating a given number of RWIS stations (Jin et al 2014; Kwon et al. 2014; Kwon and Fu 2013; Zhao et al. 2015). However, there are very few guidelines currently available providing information on the optimal density and spacing of RWIS stations for any given region. A single reference most widely being adopted is the RWIS sitting guideline made available by FHWA in 2008, which recommends an average spacing of 30-50km along a roadway (Manfredi et al. 2008). However, this recommendation appears to be originated from the existing practice and experience with little scientific justifications. Intuitively, the number of RWIS stations required for a region depends on the spatiotemporal variability of the region. Regions with winter weather conditions of high spatial variability would require a higher number of RWIS stations than those with uniform weather conditions. Currently, authorities responsible for RWIS planning have no reference available to assist them deciding the optimal density for their regions.

The objective of this study is to investigate the hypothesis that the optimal RWIS density or spacing of a region is dependent of the spatial variability of the road weather conditions of the region. To examine this hypothesis, a geostatistical approach is introduced to characterize the spatial variability of a variable of interest over a given region, which is subsequently applied to determine the corresponding optimal RWIS density or spacing for the region. The hypothesis is then assessed by comparing the correlation between the optimal RWIS density and the variability parameter.

The rest of the paper is organized as follows. The following section explains the spatial variability of road weather conditions. The third section explains the semivariogram modelling approach and its application for determining the optimal density, followed by a discussion of

83 case studies in the fourth section. The last section highlights the conclusions and provides
84 recommendations for future work.

85 2. SPATIAL VARIABILITY OF ROAD WEATHER CONDITIONS

86 As described previously, the first requirement for the proposed investigation is to quantify the
87 spatial variability of a variable of interest representing the road weather conditions. Thermal
88 mapping (TM) is a conventional technique that aims to quantify spatial variations of road surface
89 conditions on any given stretch of roads (Gustavsson 1999). TM makes it possible to precisely
90 identify common trouble spots that may require more frequent monitoring and additional
91 maintenance treatments (Zwahlen 2013). Nevertheless, it requires substantial amount of time
92 and effort, particularly for cities that are in need of a large-scale implementation, posing a
93 significant limitation of its applicability at the regional level.

94 In recent decades, geostatistics, or the study of regionalized variables (ReV), has gained
95 popularity for use in quantifying the spatial structure by modelling a semivariogram based on the
96 information provided by a sampling of ReV of interest (Journel and Huijbregts 1978; Goovaerts
97 1997; Goovaerts 1999). The main difference between the classical statistics and geostatistics is
98 the assumption that lies on the notion of existence of spatial autocorrelation – measurements
99 collected nearby tend to be similar and their differences increase as their separation distances
100 increase (Olea, 2006). Numerous past studies have implemented the geostatistical technique
101 known as kriging to appreciate the spatial autocorrelation (spatial continuity) of ReV of interest,
102 and to provide better estimates than other conventional statistical methods (Prakash and Singh
103 1998; Cameron and Hunter 2002; Nunes et al. 2004; Yet et al. 2006).

Without loss of generality, we select road surface temperature (RST) as the variable of interest. RST is considered to be one of the most important and widely used variables necessary for improved winter road maintenance services. Accurate estimation of such is critical for generating reliable road weather conditions forecast and predicting black-ice potentials as it can help dramatically reduce the risk of motor vehicle accidents during winter months. The spatial variability of RST can be modelled by semivariogram using available observations recorded at a set of sampling sites. A commonly used semivariogram estimator is as follows (Olea 2006):

$$\hat{\gamma}(h) = \frac{1}{2n(h)} \sum_{i=1}^{n(h)} [z(x_i + h) - z(x_i)]^2 \quad (1)$$

Where $\hat{\gamma}(h)$ is the sample semivariogram, $z(x_i)$ is a measurement taken at location x_i , and $n(h)$ is the number of pairs of observations separated by the lag h with a given orientation. The number of pairs included in semivariogram estimation should, at least, be equal to 30 as set by Journel and Huijbrets (1978). Likewise, the lag distance of the sample semivariogram should be constrained to half the diameter in the sampling domain for all directions of analyses (Journel and Huijbrets 1978). An important assumption of the above estimator is the absence of any systematic variations; hence if there exist any spatial patterns, then they must be removed first to be trend-free. An example of the sample variogram is illustrated in Figure 1.

Fig. 1. An example of a typical semivariogram

Where,

- *nugget effect* is the micro-scale variation or measurement error that is estimated from the empirical semivariogram as the value of $\gamma(h)$ for $h = 0$ (i.e., any spatial variability that exists at a distance smaller than the shortest distance of two measurements),
- *Sill* is the $\lim_{h \rightarrow \infty} \gamma(h)$ representing the variance of the random field, and

- *range* is the distance at which data are no longer autocorrelated or the level of the plateau (if it exists).

Typically, a functional model is fit to the sample semivariograms, and the three most commonly employed models considered in this study are as follows (Olea 2006).

Exponential:
$$Ex(h) = C \left(1 - e^{-\frac{3h}{a}} \right) \tag{2}$$

Gaussian:
$$G(h) = C \left(1 - e^{-3 \left(\frac{h}{a} \right)^2} \right) \tag{3}$$

Spherical:
$$Sp(h) = \begin{cases} C \left(\frac{3h}{2a} - \frac{1}{2} \cdot \left(\frac{h}{a} \right)^3 \right), & 0 \leq |h| < |a| \\ C, & |a| \leq |h| \end{cases} \tag{4}$$

where *a*, *h*, and *C* represent range, lag, and sill, respectively.

Since it is critical to select the model that best replicates the shape of the spatial variability over the region of interest, one needs to assess the goodness of fit for each model. Cross validation is a verification process in which each observation is removed with replacement to produce an estimate at the same site of the removal (Olea 1999). The error incurred in this process is measured by taking the difference between the actual and the estimated values. This process continues until all observations are tested. There are some statistical measures reported in this study to confirm the validity of the fitted model, and some are root-mean-square-error (RMSE), which indicates how closely the fitted model predicts the measured values (i.e., smaller the better), and mean standardized errors (MSE) which represent the average of the prediction standard errors, and the mean of the standardized errors (i.e., closer to zero the better),

respectively. For more information on semivariogram modelling, readers are referred to a comprehensive guide made available by Olea (2006).

3. EFFECT OF RWIS DENSITY ON SPATIAL INFERENCE ERRORS

To determine an optimal density of an RWIS network, it is important to first choose appropriate criteria so that the suitability of any given density can be evaluated. In our previous effort (Kwon et al. 2016), we had attempted to address the problem of finding the optimal location for a given set of RWIS stations. The problem is formulated as an integer programming problem with the objective of minimizing the weighted total of two objectives, namely, minimizing total network-wide inference errors and maximizing the total traffic exposure. The most intriguing part of this method lies with the way in which inference errors are calculated via one of the most renowned geostatistical methods known as kriging. Kriging is a geostatistical technique that provides a best linear unbiased estimator (BLUE) for variables that have tendency to vary over space and time (Yeh et al. 2006). This so-called hybrid interpolation method not only provides estimates but also kriging variance (i.e., estimation errors) at unvisited sites based on a set of available observations by characterizing and quantifying spatiotemporal variability over the area of interest (Goovaerts, 1997). For this reason, use of kriging variance has long been favored by many researchers for optimizing various sampling design strategies concerning various environmental and meteorological parameters (van Groenigen et al. 1999; Nunes et al 2007; Yeh et al. 2006; Melles et al 2011). Considering the nature of the optimization problems mentioned above, only heuristic algorithms are viable to solve large-sized problems (Revelle et al. 2008). Heuristic approaches such as genetic algorithm and spatial annealing combine the search of good or fair quality solutions with the limited computation time for solving complex

166 and large-scale problems (i.e., non-deterministic polynomial-time hard) by removing the
167 constraint of achieving a globally optimum solution (Amorim et al. 2012).

168 In this study, we focus only on the variable of interest - road surface temperature (RST) with
169 the goal of minimizing the total spatial inference errors over the road network. Because this error
170 *generally* follows a decreasing function of the total number of RWIS stations or density
171 (although there may exist some peaks in its pattern), the optimal density cannot be decided solely
172 based on this error minimization criterion. Instead, the opposing force – total installation costs
173 must be considered. However, due to the challenge of converting the inference error to monetary
174 benefit, we decided to identify the optimal density indirectly from the relationship between
175 inference error and the number of RWIS stations. The process includes the following key steps:

- 176 • *Estimate the semivariogram model* based on available observations following the process
177 described in the previous section;
- 178 • *Find the optimal location* for any given number of RWIS stations: as described in Kwon
179 et al. (2016), kriging variance of RST is calculated to reflect the needs for locating RWIS
180 stations for improved winter road maintenance operations (i.e., locations with higher
181 estimation errors require more attention than others with lower errors), and sum of
182 average kriging variance should consequently be minimized. There are a few different
183 kriging models available and the model considered in this study is called universal
184 kriging, a hybrid geostatistical technique, which forms a basis of point observations and
185 regression of the target variable on spatially exhaustive auxiliary information (Hengl et al.
186 2007);

- The optimization problem is solved using a combinatorial optimization algorithm called spatial simulated annealing (SSA). SSA is a spatial counterpart to simulated annealing (SA, Kirkpatrick et al. 1983), specifically designed to optimize sampling designs of environmental variables using kriging. SA is a stochastic metaheuristic search algorithm first proposed by Metropolis et al. (1953) and mimics the annealing of metal. SA is fundamentally same as Monte Carlo annealing, probabilistic hill climbing, statistical cooling, and stochastic relaxation (Aarts and Korst 1989). The term “annealing” is related to the metallurgical process of metal alloy heating and relaxed cooling to increase toughness and reduce brittleness (Goovaerts 1997). In searching for the optimality, this algorithm utilizes a vector h to control the direction and the length, with which random perturbations are made iteratively until some user-defined stopping conditions (i.e., number of iterations) are met. It is this unique generation mechanism that has made the algorithm more attractive when dealing with optimizing a sensor network in a two-dimensional geographic space (Brus and Heuvelink 2007; Mohammadi et al. 2012; Amorim et al. 2012; Pereira et al. 2013). Also, following the recommendations made by the prior researchers, the study area will be discretized into a grid system so that computational effort can be significantly reduced (ensure convergence) while maintaining an acceptable level of spatial representativeness. If the reader needs more information how the algorithm works, see Kirkpatrick et al. (1983), and van Groenigen et al. (1999).

4. CASE STUDIES

4.1 Study Sites and Data Processing

The proposed approaches are examined via four case studies covering one Canadian province (Ontario), and three US states (Utah, Minnesota, and Iowa) using various dataset provided by each region under investigation. These four regions are a good candidate as they already have a well distributed and dense RWIS network with distinctive and unique meteorological (lake effect) and topographical (mountainous) characteristics, from which more reliable assessments (e.g., analysis of spatial variability from RWIS measurements) can be realized. The findings from each region should provide sensible guidelines and measures as to how the optimal location and density would vary from one region to another.

Ontario is the second largest Canadian province, situated in east-central Canada, and has a continental climate like most other provinces of Canada. Northern Ontario has long, very cold winters and short summers whereas the southern part enjoys the tempering effect of the Great Lakes. Southwestern Ontario is in the Mixedwood Plains that is typically flat with rolling hills. To its north contains the Hudson Plains consisting mainly of flat and wet surface. Utah is situated in the Mountain States (also called the Mountain West) from one of the nine geographic divisions of the United States. Because of its geographic location, Utah has extremely varied topography with a large portion of the State being mountainous. The lowest area is the Virgin River Valley in the southwestern part with altitude of 750m, while the highest points lies in the Uinta Mountains with altitude higher than 4000m. Utah is also known for very diverse climates – for instance, there are definite variations in temperature with altitude and with latitude. Average temperature differences between the southern and northern counties at around similar altitudes

typically range between 6 and 8 degrees with the northern counties having lower temperatures. The topographies of Iowa and Minnesota, on the other hand, consist mainly of rolling plain and flat prairie. The differences of their lowest and highest altitudes are also small, ranging from the lowest points of 183m and 146m to the highest points of 702m and 509m for Minnesota and Iowa, respectively.

As for the data used in this study, Ontario, Utah, Minnesota, and Iowa, which currently have 140, 96, 97, and 67 RWIS stations well distributed over their regional road network, respectively, provided historical RWIS observations collected at 10 to 15-min intervals over three consecutive winters (October to March). Figure 2 depicts the four study areas with existing RWIS stations and road network shown in red circles and black lines, respectively. The data came stratified by individual stations each containing measurements including the variable of interest – road surface temperature (RST). To assure the validity of the data, a data quality check was performed and definite outliers (i.e., -9999 °C) recorded as a result of sensor malfunctions were removed. The data from the four study areas together contained nearly 80 million rows of data; hence VBA scripts were written to process the entire datasets, returning a seasonal and a monthly average of RST stratified by station. These processed datasets will be an input to constructing a semivariogram model on two different temporal units, monthly and seasonal. It is worthwhile noting that use of finer temporal aggregation units (i.e., hourly / daily) could result in different correlation patterns, whereby making the results more suitable at the operational level. However, the aggregation units selected in this study are deemed more appropriate as the primary focus is on optimizing a long-term stationary RWIS network, which are more suitable at the network-wide planning level.

Fig. 2. Study areas under investigation and the existing RWIS networks: (a) Ontario, (b) Iowa, (c) Minnesota, and (d) Utah. (Map Source: Esri 2015)

4.2 Semivariograms of Seasonal and Monthly RST

To investigate the spatial correlation range of the RWIS measurement, the processed RST data were implemented as per the general modelling guidelines discussed earlier. Locational attributes (i.e., latitude and longitude) of each RWIS station were extracted and implemented to de-trend any existing patterns. Semivariogram models were calibrated using R with packages gstat (Pebesma 2004) and automap (Hiemstra et al. 2009).

Figure 3 shows the sample and fitted semivariogram models using the seasonal RST data and Table 1 provides a summary of semivariogram model parameters including sill, nugget, and range, and cross-validation results illustrating the accuracy of the fitted models for both monthly and seasonal data.

MAE and RMSE represents mean absolute error and root mean squared error. COR indicates the correlation between the predicted and observed values (ideally 1). As anticipated, spatial correlation of RST at Iowa and Minnesota (Figure 3b & 3c) having relatively less varied topography is shown to have a longer spatial correlation range suggesting that on average, the RST measurements at those two regions vary less (thus more predictable) when compared to those at Utah undergoing more varied topography. In addition, Ontario with a moderate topographic variability has a spatial correlation range falling between the ranges of other three regions. Likewise, the spatial structure of RST from Utah is less stable and tends to fluctuate in a greater range (in y-axis) as the separation distance increases, whereas the other two regions have a less fluctuation of semivariances contributing to the higher prediction power.

274

Draft

Table 1: Summary of Semivariogram Parameters and Cross-validation Using Monthly and Seasonal RST for All Study Sites

Fig. 3. Sample and fitted semivariogram models for four regions using the seasonal RST data.

Another inference that can be made by observing the resulting statistics is that for all four regions, the discrepancies tend to be relatively higher for shoulder months (i.e. October and March) than non-shoulder months (i.e. November, December, January, and February). This could be due to the fact that the weather patterns typically vary in a wider range over these shoulder months, making it more difficult to have accurate predictions. Furthermore, during these months, spatial continuity of the weather-related variable (RST) is also affected, resulting in a shorter range. The mean range using the monthly data for Ontario, Iowa, Utah, and Minnesota are found to be 78.72km, 101.91km, 34.93km, and 93.24km, respectively, which generally agree with the average ranges found using the seasonal data (72.84km, 90.48km, 40.47km, and 95.47km for Ontario, Iowa, Utah, and Minnesota, respectively). Slightly different ranges resulted from using the monthly data could be due to the generalization or aggregation of all monthly RST data.

4.3 Optimal Density and Its Relationship to Spatial Variability

The proposed optimization model was first implemented in designing an optimal RWIS network using the kriging variance of the seasonal RST data as an optimization criterion (i.e., minimization of kriging variance). A demand surface was discretized into a uniform grid having a cell size of 1km by 1km. Since an RWIS station should not be placed anywhere other than a regional road network, the solution space was limited to the road network. For each region, the constrained optimization was run in an iterative fashion by adding one additional RWIS station

to the network and its corresponding fitness value was recorded. The optimization continued until the total number of stations reached 350 – an arbitrary number ensuring that the key pattern in error-density relationship is fully revealed. For each run, a total number of iterations was set to 5,000 but an additional stopping criterion was used to terminate the run whenever there was lack of progress in improving the fitness function (i.e., marginal difference between the current and previous solutions). In addition, probability of accepting a worsening solution was set to 0.2 as recommended by the prior studies (Kirpatrick 1984; Brus and Heuvelink 2007; Heuvelink et al. 2010) to avoid being trapped in local minima. For each study region, the average running time was approximately 3 days by Windows operating desktop computer running with a 3.39 GHz processor and 8.00 GB of RAM. R was used as a base platform for running all the optimization in this study.

To enforce a valid and fair comparison, the fitness values were normalized and the number of stations added to the network was converted to the number of station per unit areas ($10,000\text{km}^2$). This was necessary because the total area of each study area was different, thus comparing the fitness value directly to the number of stations added will not be considered valid. Figure 4 shows the comparison of the optimization results as a function of pre-specified density for all four regions.

A quick visual inspection of Figure 4 shows that Iowa and Minnesota both having a less varied topography requires a less number of stations per unit area of $10,000\text{km}^2$, while Utah (more varied topography) requires a considerably more number of stations to achieve the comparable objective function values. Likewise, Ontario with moderately varied topographic characteristics requires a more number than Iowa and Minnesota but a less number than Utah. Another important conclusion that can be drawn is that regions with a longer spatial continuity

(Iowa and Minnesota as defined in semivariograms) will require a less number of stations to cover the same area than a region with a much shorter spatial continuity (Utah). This makes intuitive sense since the measurements taken at a less varied topographic region will be able to represent and cover a larger area. Given the shape of the all four curves, it is quite challenging to pin point the optimal density. Instead, a rate of change was calculated for every point and when the change is around 3% (again, an arbitrary number selected for a comparison only), the corresponding density was considered as the optimal density. As a result, Iowa, Minnesota, Ontario would require 2.0, 2.2, and 2.9 stations per every $10,000\text{km}^2$, respectively, whereas Utah would need 4.5 stations to cover the same area, indicating that a topographically varied region will likely need about 2 times more the number of RWIS stations required by a less varied region. Using the numbers generated, it is also possible to evaluate and assess whether one will need to install additional RWIS stations to provide a good coverage in terms of the monitoring capability of road surface temperature. Considering the size of each study area covered herein, Utah, Iowa, Minnesota, and Ontario would benefit by expanding their network by installing additional 59, 16, 38, and 72 RWIS stations, respectively. It must be noted that such recommendations are based on optimization criteria used in this study, and can change when different variables are used in the optimization process and when the land use characteristics are considered. For instance, many northern areas in Ontario have no or less road network, thus in practice, it may not be necessary to install RWIS stations in those areas.

Fig. 4.: Normalized estimation errors as a function of RWIS density for the four study areas

To further test the aforementioned hypothesis, the relationship between the optimal number of RWIS station required per unit area ($10,000\text{ km}^2$) and the semivariogram parameter - range is

examined, as illustrated in Figure 5. Although the relationship relies on a small number of case studies, it reveals a clear linkage between the two measures, demonstrating the usability of the correlation range in any given area for conveniently determining the station density. For instance, if the analysis of interest is the number of stations per unit area, a region with 60km range (for the given variable of interest) would require, on average, 3.5 RWIS stations per every 10,000 km² so as to have an adequate coverage. There is no doubt that more case studies are required to obtain a promising result, it certainly provides valuable information, particularly for highway authorities initiating a state-wide RWIS implementation plan.

Fig. 5. A linear relationship of range vs density

5. CONCLUSIONS AND RECOMMENDATIONS

RWIS network represents a critical infrastructure for most transportation agencies. Design of an effective RWIS network for a region depends on a good understanding of the regional road weather patterns. In this research, we have investigated the hypothesized relationship between the spatiotemporal variability of road weather conditions of a region and the number of RWIS stations required to provide adequate coverage of the region. An empirical study was conducted using data from three US states and one Canadian province with different weather and topographical characteristics: Iowa and Minnesota (plain), Ontario (plain and rolling) and Utah (mountainous). One of the key road weather condition parameters - road surface temperature (RST) - was considered in the subsequent geostatistical analysis and RWIS location optimization. Their spatiotemporal characteristics were quantified by constructing semivariograms, which were examined to estimate the spatial correlation ranges. The respective spatial structures constructed based on seasonal RST measurements were first used to determine the optimal RWIS network (configuration) for each additional RWIS station added to the network. The

RWIS station location optimization was formulated on a basis of minimizing the spatially averaged kriging estimation errors (kriging variance) of RST measurements. The optimization results were then used to produce the RWIS density curve for each region. The main findings are summarized as follows:

- An investigation of spatial autocorrelations of RST measurements on two temporal units revealed that the ranges for plain regions were found to be comparable - average ranges determined using the monthly and seasonal data were 101.91km and 90.48km for Iowa and 98.24km and 95.47km for Minnesota, respectively. In contrast, an average range for mountainous regions was found to be 34.93km (monthly) and 40.47km (seasonal), which are much shorter than those found for plain regions indicating lower spatial continuity. Furthermore, Ontario with mixed terrain (rolling plain) was found to have a moderate range - 78.72km and 72.84km using the monthly and seasonal data, respectively.
- A general pattern was observed for all four study sites that the ranges during shoulder months (October and March) were relatively shorter than the rest of the months, implying that the RST in these shoulder months had higher spatial variation.
- A visual inspection of the four density curves shows that the number of stations required per unit area ($10,000\text{km}^2$) is less in Iowa and Minnesota, than that in Utah, indicating that a larger number of stations would be required in a mountainous region to cover the same area located in a plain region. This finding has also confirmed the positive correlation between the optimal RWIS density and the semivariogram parameter – the average range.
- An optimal density was determined by assuming that it occurred when the percentage of the rate of changes were about 3%. As a result, Iowa and Minnesota would require, on average,

2 stations while Utah would require around 8 stations per unit area, suggesting that mountainous regions such as Utah are in need of 4 times more the number of RWIS stations than the plain regions. It was also found that Ontario with mixed topographical characteristics would require about 4 stations per unit area.

- The findings also suggest that there is a strong dependency between the RWIS density and the spatial correlation parameter – range. The regions with less varied topography tend to have a longer spatial correlation range than the region with more varied topography.

Further research is recommended in the following specific directions:

- First, more case studies should be conducted to regions with different climatic and topographical characteristics such that more conclusive results can be obtained. With a sufficient number of case studies, it is possible to develop a quantitative relationship between the optimal spacing and semivariogram range.
- Second, more datasets covering a larger temporal range (i.e., 10 years) would be required to improve the level of confidence on the findings presented herein.
- Third, other key road weather condition variables such as snow/ice cover, visibility, and friction should be investigated to evaluate their correlation ranges for their comparability with the ones obtained using RST measurements.
- Fourth, the optimization output was derived from using a single semivariogram that is assumed to be homogeneous over the entire area. This rather strong assumption may hold true in smaller regions but may not be true of larger regions with diverse climates (e.g., Ontario). To avoid such shortcomings, a more detailed analysis should be carried out to

410 identify if a region under investigation needs to be represented by two or more
411 semivariogram models.

- 412 • Lastly, traffic-related factors including accident rate, highway class, or traffic volume
413 should be considered when designing an optimal RWIS network so as to obtain a more
414 balanced network satisfying the needs of winter maintenance groups as well as the needs of
415 travelling public.

416 Nevertheless, the findings from this study should provide a useful reference for highway
417 authorities to better decide the average spacing (range) and the density (number of stations per
418 unit area) as per the topographical characteristics of their regions when developing a statewide
419 RWIS implementation and/or expansion plan.

420 ACKNOWLEDGEMENTS

421 The authors wish to thank Ontario Ministry of Transportation, and Iowa, Utah, and Minnesota
422 Department of Transportation for providing the data necessary to complete this study. The
423 authors would also like to thank our anonymous reviewers for providing constructive feedback
424 and helpful comments that have significantly contributed to improving the quality of this paper.
425 This research was funded by the Aurora Program (www.aurora-program.org), and was partially
426 funded by National Sciences and Engineering Research Council of Canada (NSERC).

427 REFERENCES

- 428 Aarts, E., and Korst, J. 1989. Simulated Annealing and Boltzmann Machines – A stochastic
429 approach to combinatorial optimization and neural computing. John Wiley & Sons, New York,
430 NY.
- 431 Amorim, A.M.T., Goncalves, A.B., Nunes, L.M., and Sousa, A.J. 2012. Optimizing the location
432 of weather monitoring stations using estimation uncertainty. *International Journal of*
433 *Climatology*, 32: 941-952.

- 434 Brus, D.J., and Heuvelink, G.B.M. 2007. Optimization of sample patterns for university Kriging
435 of environmental variables. *Geoderma* 138, 86–95.
- 436 Cameron, K., and Hunter, P. 2002. Using spatial models and kriging techniques to optimize
437 long-term ground-water monitoring networks: a case study. *Environmetrics*. 13, pp. 629-657.
- 438 ESRI. 2016. ArcGIS Desktop: Release 10.4. Redlands, CA: Environmental Systems Research
439 Institute. Map Sources: Esri, GEBCO, NOAA, National Geographic, DeLorme, HERE,
440 Geonames.org, and other contributors.
- 441 Goovaerts, P. 1997. *Geostatistics for Natural Resources Evaluation*. Oxford University Press,
442 New York.
- 443 Goovaerts, P. 1999 Geostatistics in soil science: state-of-the-art and perspectives. *Geoderma* 89,
444 pp 1–46
- 445 Gustavsson, T. 1999. Thermal mapping – a technique for road climatological studies
446 *Meteorological Applications*. vol. 6, pp. 385-394.
- 447 Hengl, T., Heuvelink, G.B.M., and Rossiter, D.G. 2007. About regression-kriging: From
448 equations to case studies. *Computers and geosciences*. 33, pp 1301-1315.
- 449 Heuvelink, G.B.M., Brus, D., de Gruijter, J.J. 2006. Optimization of sample configurations for
450 digital mapping of soil properties with universal kriging. In: Lagacherie, P., McBratney, A.,
451 Voltz, M. (Eds.), *Digital Soil Mapping: An Introductory Perspective*. Elsevier, Amsterdam,
452 The Netherlands, pp. 1–17.
- 453 Hiemstra, P.H., Pebesma, E.J., Twenhofel, C.J.W. and Heuvelink, G.B.M. 2009. Real-time
454 automatic interpolation of ambient gamma dose rates from the Dutch Radioactivity
455 Monitoring Network. *Computers & Geosciences*.
- 456 Jin, P.J., Walker, A., Cebelak, M, and Walton, M. 2014. Determining strategic locations for
457 environmental sensor stations with weather-related crash data. In *Transportation Research*
458 *Record: Journal of the Transportation Research Board*, No. 2440, Transportation Research
459 Board of the National Academies, Washington, D.C. pp. 34-42.
- 460 Journel, A.G., Huijbregts, C.J. 1978. *Mining Geostatistics*. Academic Press, New York.
- 461 Kirkpatrick S, Gellat Jr. CD, Vecchi MP. 1983. Optimization by simulated annealing. *Science*
462 220. 671–680.
- 463 Kwon, T. J., Fu, L., & Jiang, C. 2014. RWIS Stations – Where and How Many to Install: A Cost
464 Benefit Analysis Approach, *Canadian Journal of Civil Engineering (CJCE)*, DOI:
465 10.1139/cjce-2013-0569.

- 466 Kwon, T.J., and Fu, L. 2013. Evaluation of alternative criteria for determining the optimal
467 location of RWIS stations. *Journal of Modern Transportation*, 21: 17–27, doi:10.1007/s40534-
468 013-0008-9.
- 469 Kwon, T.J., Fu, L., & Melles, S. 2016. Location Optimization of road weather information
470 system (RWIS) network considering the needs of winter road maintenance and the traveling
471 public. *Computer-Aided Civil and Infrastructure Engineering*. 1-15, DOI:
472 10.1111/mice.12222
- 473 Manfredi, J., Walters, T., Wilke, G., Osborne, L., Hart, R., Incrocci, T., and Schmitt, T. 2005.
474 Road weather information system environmental sensor station sitting guidelines. Publication
475 No. FHWA-HOP-05-026. Federal Highway Administration. April 2005.
- 476 Manfredi, J., Walters, T., Wilke, G., Osborne, L., Hart, R., Incrocci, T., Schmitt, T., Garrett, V.
477 K., Boyce, B., and Krechmer, D. 2008. Road Weather Information System Environmental
478 Sensor Station Siting Guidelines, Version 2.0. FHWA, Washington, DC.
- 479 Melles, S.J., Heuvelink, G.B.M., Twenhofel, C.J.W., van Dijk, A., Hiemstra, P.H., Baume, O.,
480 and Stohlker, U. 2011. Optimizing the spatial pattern of networks for monitoring radioactive
481 releases. *Computers and Geosciences*, vol. 37, pp. 280-288.
- 482 Metropolis N, Rosenbluth A, Rosenbluth M, Teller A, Teller E. 1953. Equation of state
483 calculations by fast computing machines. *Journal of Chemical Physics* 21, pp. 1087–1092.
- 484 Mohammadi, S.S., Hezarkhani, A., and Tercan, A. E. 2012. Optimally locating additional drill
485 holes in three dimensions using grade and simulated annealing. *Journal of Geological society*
486 *of India*, Vol. 80 pp. 700-706.
- 487 Nunes, L.M, Cunha, M.C., and Ribeiro, L. 2004. Optimal space-time coverage and exploration
488 costs in groundwater monitoring networks. *Environmental Monitoring and Assessment* 93,
489 103-124.
- 490 Olea R.A. 1999. *Geostatistics for engineers and earth scientists*. Kluwer, Boston.
- 491 Olea, R.A. 2006. A six-step practical approach to semivariogram modeling. *Stoch Environ Res*
492 *Risk Assess* 20, pp. 307–318
- 493 Pebesma, E.J. 2004. Multivariable geostatistics in S: the gstat package. *Computers &*
494 *Geosciences* 30 (7), 683–691.
- 495 Pereira, G. T., de Souza, Z. M., Teixeira, D. D. B., Montanari, R., and Marques, J. 2013.
496 Optimization of the sampling scheme for maps of physical and chemical properties estimated
497 by kriging. *R, Bras, Ci, Solo*, vol 27 no.5, pp 1128-1135.

- 498 Prakash, M.R., and Singh, V.S. 1998. Network design for groundwater monitoring – a case study.
499 Environmental Geology 39 (6), pp 628-632.
- 500 Revelle, C.S., Eiselt, H.A., and Daskin, M.S. 2008. A bibliography for some fundamental
501 problem categories in discrete location science. European Journal of Operational Research,
502 184, 817–848
- 503 Shao, J., Swanson J.C., Patterson, R., Lister, P.J., and McDonald, A.N. 1997. Variation of winter
504 road surface temperature due to topography and application of Thermal Mapping
505 Meteorological Applications, vol. 4, pp. 131-137.
- 506 Van Groenigen, J.W., Siderius, W., Stein, A. 1999. Constrained optimization of soil sampling for
507 minimization of the kriging variance. Geoderma 87, 239–259.
- 508 Yeh, M, Lin, Y.P, and Change, L.C. 2006. Designing an optimal multivariate geostatistical
509 groundwater quality monitoring network using factorial kriging and genetic algorithms.
510 Environ Geol, 50, pp. 101–121
- 511 Zhao, L., and Chien, S. 2014. Optimizing RWIS Sites considering Traffic and Regional Weather
512 Conditions. Transportation Research Board 93rd Annual Meeting, pp 18.
- 513 Zhao, L., Chien, S., Meegoda, J., Luo, Z. & Liu, X. 2015. Cost–benefit analysis and
514 microclimate-based optimization of a RWIS network, Journal of Infrastructure Systems, 22(2),
515 1–11, DOI: 10.1061/(ASCE)IS.1943-555X.0000278, 04015021.
- 516 Zwahlen, H.T., Russ, A., and Vata, S. 2013. Evaluation of ODOT Roadway/Weather Sensor
517 Systems for Snow and Ice Removal Operations- Part I: RWIS. Ohio Research Institute for
518 Transportation and the Environment. FHWA/OH-2003/008A.

Fig. 1. An example of a typical semivariogram.

Fig. 2. Study areas under investigation and the existing RWIS networks: (a) Ontario, (b) Iowa, (c) Minnesota, and (d) Utah. (Map Source: Esri 2016)

(a). Ontario

(b). Iowa

(c). Minnesota

(d). Utah

Fig. 3. Sample and fitted semivariogram models for four regions using the seasonal RST data.

Fig. 4.: Normalized estimation errors as a function of RWIS density for the four study areas

Fig. 5. A linear relationship of range vs density

1 **Table 1: Summary of Semivariogram Parameters and Cross-validation Using Monthly and**
 2 **Seasonal RST for All Study Sites**

State	Month/ Seasonal	Model	Model Parameters			Cross-validation		
			Nugget	Sill	Range (km)	MAE	COR	RMSE
Iowa	Oct	Sph	0.00	1.10	88.96	0.60	0.70	0.98
	Nov	Gau	2.20	3.20	142.71	1.37	0.51	1.70
	Dec	Exp	0.43	5.90	106.08	1.25	0.46	2.19
	Jan	Sph	0.00	2.12	100.16	0.63	0.67	1.02
	Feb	Sph	0.00	1.14	111.63	0.67	0.77	1.02
	Mar	Sph	0.00	1.66	61.95	0.90	0.69	1.33
	Seasonal	Sph	0.00	0.58	90.48	0.57	0.86	0.81
Utah	Oct	Gau	0.00	9.70	21.39	2.60	0.42	3.41
	Nov	Sph	0.46	3.42	50.29	1.34	0.55	1.69
	Dec	Gau	2.30	5.40	33.74	1.59	0.42	2.22
	Jan	Gau	0.68	2.40	37.31	1.10	0.55	1.46
	Feb	Gau	0.88	2.70	40.12	1.10	0.63	1.43
	Mar	Gau	1.10	5.90	26.71	1.55	0.62	1.94
	Seasonal	Exp	0.27	2.78	40.47	1.24	0.53	1.55
Minnesota	Oct	Sph	0.00	2.70	62.68	1.31	0.60	1.87
	Nov	Sph	0.37	5.00	133.60	1.41	0.25	2.65
	Dec	Gau	0.00	5.50	83.59	1.22	0.68	2.23
	Jan	Sph	1.60	3.50	116.26	1.02	0.60	2.02
	Feb	Sph	0.00	0.47	114.31	0.52	0.90	0.71
	Mar	Sph	0.27	2.80	78.99	1.20	0.59	1.94
	Seasonal	Sph	0.00	3.50	95.47	1.15	0.55	2.14
Ontario	Oct	Gau	0.00	6.50	43.21	1.83	0.55	3.07
	Nov	Gau	1.20	4.30	97.29	2.97	0.64	2.49
	Dec	Gau	0.65	3.40	102.50	1.43	0.87	2.01
	Jan	Sph	0.00	1.90	99.29	1.22	0.90	1.64
	Feb	Sph	0.42	1.90	70.28	1.19	0.88	1.76
	Mar	Gau	0.00	1.70	59.76	0.86	0.85	1.10
	Seasonal	Gau	0.20	0.27	72.84	1.15	0.72	1.64