

Article

Optimizing RWIS Locations with Wasserstein Distance and Geostatistics: A Case Study in South Korea

Nancy Huynh ¹, Jinhwan Jang ² and Tae J. Kwon ^{1,*}

¹ Department of Civil and Environmental Engineering, Faculty of Engineering, University of Alberta, Edmonton, AB T6G 1H9, Canada; loingan@ualberta.ca

² Korea Institute of Civil Engineering and Building Technology, Goyang-si 10223, Republic of Korea; jhjang@kict.re.kr

* Correspondence: tjkwon@ualberta.ca

Abstract: Road Weather Information Systems (RWISs) are essential components of modern Intelligent Transportation Systems (ITSs) deployed in cold regions to gather real-time data on winter weather and road surface conditions. Despite their benefits, the high cost associated with RWIS installations demands optimized placement strategies to maximize their utility and cost-effectiveness. Geostatistics-based RWIS location-allocation methods, particularly those involving semivariogram modeling to quantify underlying spatial characteristics, have gained international recognition. However, new locations require unique semivariogram models, a process that is time-consuming and constrained by the availability of comprehensive datasets, often rendering location analysis challenging or infeasible. Addressing these limitations, this study introduces an innovative approach using Wasserstein Distance (WD) to link semivariograms across different datasets. This method streamlines optimization by eliminating the need for repetitive semivariogram modeling in new study areas. Our findings demonstrate that WD-matched models replicate the location choices of original models with a high degree of similarity while ensuring that clean-slate locations remain proximate to those of original models, enhancing geographic equity in RWIS deployment. This validates the practicality of reusing developed semivariogram parameters for WD-matched highways, significantly reducing the need for new geostatistical analyses and enhancing the framework's applicability and accessibility for RWIS deployment across diverse geographic regions.



Academic Editor: Luigi dell'Olio

Received: 8 November 2024

Revised: 11 February 2025

Accepted: 18 February 2025

Published: 1 March 2025

Citation: Huynh, N.; Jang, J.; Kwon, T.J. Optimizing RWIS Locations with Wasserstein Distance and Geostatistics: A Case Study in South Korea. *Future Transp.* **2025**, *5*, 23. <https://doi.org/10.3390/futuretransp5010023>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: intelligent transportation systems; geostatistics; road weather information systems; location optimization; spatial equity; Wasserstein distance

1. Introduction

Intelligent Transportation Systems (ITSs) play a critical role in modern transportation engineering, improving transportation safety and mobility, especially in cold regions with adverse weather conditions. In many countries worldwide, the frequency of snow, ice, fog, and frost events are persistent problems that pose significant safety risks to traffic users. These events create challenging driving conditions for road users, necessitating transportation authorities to make interventions as needed. For these agencies to ensure that their roads remain in optimal conditions, timely and efficient Winter Road Maintenance (WRM) operations are critical.

1.1. Understanding RWISs: Uses, Types, and Limitations

To support the WRM decision-making process, the integration of ITS technologies, such as Road Weather Information Systems (RWISs), stands as a crucial component. RWISs are an advanced ITS technology that gathers, processes, and disseminates real-time road surface and weather condition information. This information can include air temperature, road surface temperature, road friction, and road surface status, amongst a variety of other weather-related variables. In practice, there are two types of RWISs available: mobile and stationary. Mobile RWISs consist of sensors and dash cameras installed on vehicles, which collect data along their travel routes with each data point geotagged with its precise location, whereas stationary RWISs consist of remote sensors installed along the roadway, collecting data at their installed location. Information collected and distributed by these RWISs can be used by transportation personnel to make reliable decisions regarding WRM and by road users to make informed decisions on trip routing. RWISs provide real-time measurements, facilitating the dissemination of immediate warnings to road users about upcoming hazardous conditions such as icy or snowy roads and flooding. This timely information can significantly reduce the risks associated with adverse road conditions by giving users ample time to prepare for them. As a result, the implementation of RWISs can have significant positive effects on traffic safety, mobility, level of service, and WRM. El Esawey et al. [1] performed a study to assess the safety benefits of integrating RWISs and Variable Message Signs in British Columbia, Canada, and found a 32.7% reduction in serious winter collisions after the implementation.

However, each type of RWIS has its own set of limitations. Although mobile RWISs provide spatially continuous measurements, these measurements are temporally discrete. Additionally, mobile RWISs often encounter problems of sensor damage and improper mounting, resulting in inaccurate data collection [2]. Similarly, while stationary RWIS provides highly temporal and accurate measurements, it still retains disadvantages such as cost. A typical RWIS installation with basic functionalities can still cost upwards of USD 50,000, even before adding extra sensors. Another disadvantage associated with stationary RWIS lies in its provision of point-based measurements, which often fail to account for the diversity and variability of the surrounding road surface conditions, especially in regions with microclimates. Therefore, strategic placement of RWIS stations is crucial to optimize coverage and ensure accurate representation of road conditions while simultaneously minimizing installation costs.

1.2. Literature Review

A multitude of studies have contributed to the optimization of density and locations of RWIS stations, demonstrating a variety of methodologies addressing this critical challenge. For instance, Jin et al. [3] strategically determined RWIS locations using weather-related crash data, providing a direct approach to minimizing weather-related accidents. Eriksson and Norrman [4], on the other hand, developed a method for optimal RWIS allocation based on road slipperiness classified from RWIS meteorological data to best identify hazardous road conditions. Zhao et al. [5] developed a two-stage model that optimized RWIS density and locations by considering both regional weather severity and associated cost and benefit. Additionally, Singh et al. [6] proposed a cell-based methodology to optimize RWIS benefits by identifying weather-related crash patterns in discretized urban areas. Fetzer et al. [7] developed a multi-objective RWIS location optimization methodology that accounts for crash data, vehicle miles traveled, geographic area coverage, power and maintenance access, and existing environmental sensor stations. Similarly, Al-Kaisy and Ewan [8] introduced a merit-based ranking system that prioritizes RWIS locations based on weather severity, crash rates, and vehicle miles traveled.

In parallel, our research has made significant strides in enhancing RWIS station optimization through a geostatistics-based approach. This method specifically focused on minimizing kriging variance to maximize monitoring capabilities, effectively reflecting the demand for RWIS stations and gaining international recognition [9]. We enhanced our methodology by integrating multiple weather variables, both spatially and temporally, which increased the precision and utility of RWIS deployments initially in the state of Iowa [9]. This approach incorporated spatial equity, aiming to ensure equitable access to critical weather monitoring capabilities across diverse geographic areas. This broader application has been implemented in real-world settings, demonstrating the practical effectiveness and adaptability of our models.

1.3. Challenges in RWIS Location Optimization Methods

Despite the advancements brought by our well-established geostatistics-based work on optimal RWIS location and density, certain limitations persist. Specifically, the frameworks we developed necessitate comprehensive geostatistical analyses, notably semivariogram modeling. This process is not only technically demanding and time-consuming but also requires a new model for each study area. The challenge becomes particularly pronounced when adequate datasets are unavailable, rendering location analysis impossible. While existing RWIS optimization methods are robust, they fall short of addressing the practical constraints where limited resources pose challenges. Such constraints underscore the need for a more accessible and efficient modeling approach to ensure practical implementation of the proposed RWIS location-allocation frameworks.

1.4. Research Objectives

To overcome the limitations associated with traditional semivariogram modeling in RWIS optimization, particularly the complexity and extensive time required for multi-variable analysis, the primary objective of this study is to streamline the geostatistical processes necessary for effective RWIS station placement. This is achieved through the innovative application of Wasserstein Distance (WD), a mathematical metric that is computationally efficient for low-dimensional data, especially with modern programming tools and libraries. By leveraging WD, this research attempts to establish a direct link between semivariograms across different datasets, enabling the reuse of previously developed semivariogram models for new study areas without the need for their reconfiguration or repetitive analysis. This approach not only diminishes the technical challenges often faced in geostatistical modeling but also expedites the application of RWIS optimization frameworks to new regions. Ultimately, this study aims to deliver a robust, scalable method that can be uniformly applied across various geographic settings, thereby enhancing the strategic deployment of RWIS stations and supporting more informed decision-making in WRM.

2. Methodology

2.1. Overview of Research Procedures

The first stage in this study was data collection, where mobile RWIS data, including road surface temperature (RST), dew point temperature (DPT), air temperature (AT), as well as geographic data about the study area, were gathered to capture the inherent nature and spatial variability of conditions along the studied highways. The next step was data processing to remove any missing or erroneous data. Finally, Min–Max normalization was applied to scale all variables across routes to a range of zero to one. This normalization step ensures the minimization of potential bias in the WD analysis.

After that, a geostatistical analysis was performed. The first step was to construct empirical variograms from processed data to optimize parameter estimations for unobserved locations and to account for the possible autocorrelation related to RWIS variables [9]. Variogram clouds for all variables were joined together and binned, resulting in a combined semivariogram incorporating all three variables.

The next stage was to perform WD analysis to compare the data distributions of different routes so that similar routes could be matched. Following the highway matching process, the semivariograms developed for each highway could then be used to run the optimization process for its matched counterpart. Subsequently, optimal RWIS locations were determined using a model-based approach. This approach employs kriging to obtain unbiased estimates while minimizing kriging variance. Two phases of location-allocation analysis were performed and examined: preferred location selection and clean slate optimization. Both phases also consist of two scenarios: the first uses original semivariogram parameters, while the second uses WD-matched semivariogram parameters. Finally, the effectiveness of incorporating WD into the location-allocation framework was assessed using matching percentage values and weighted separation distance. An overview of the research procedures is shown in Figure 1.

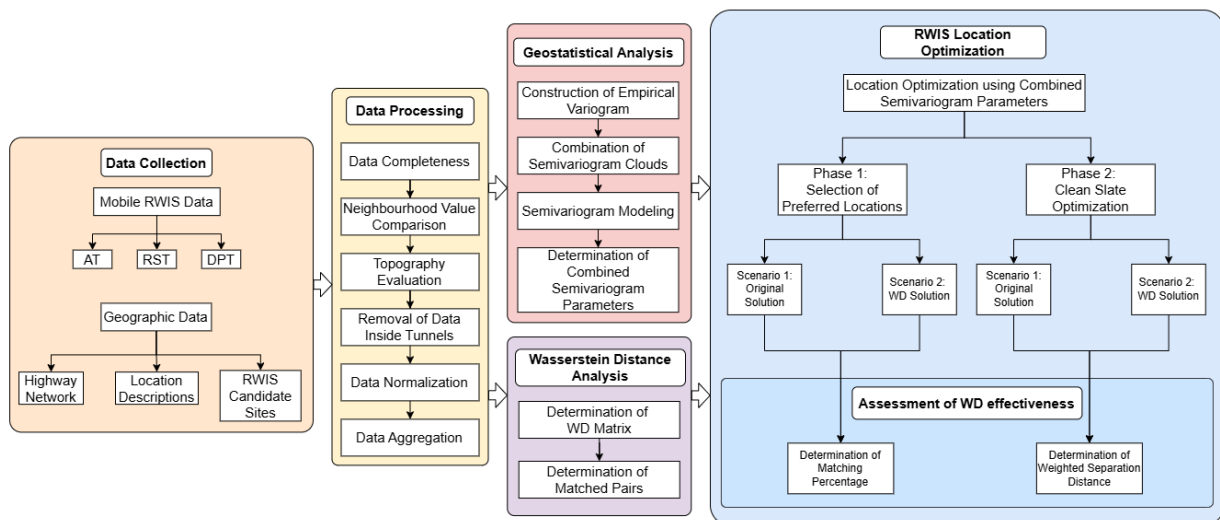


Figure 1. Overview of research procedure.

2.2. Geostatistical Analysis

For this study, geostatistical analysis was performed to maximize the likelihood of capturing the spatial variations in RWIS data while still minimizing potential bias. To perform this analysis, a key tool, spatial semivariograms, which quantify the dissimilarity between pairs of measurements based on their spatial separation distance and the direction of the line connecting the sampling pair, was applied [9]. The semivariogram has three basic parameters: sill, nugget, and range. The sill is the semivariance value of a semivariogram where it plateaus, representing the total variation in the dataset. The range is the distance at which the semivariogram reaches its sill, beyond which there is no correlation among the measurements. Lastly, the nugget is the semivariogram value at zero distance, which is expected to be zero. While theoretically expected to be zero, measurement and sampling errors can result in a non-zero nugget value, leading to what is termed the “nugget effect”. The difference between the actual sill and the nugget effect is referred to as a partial sill. A typical semivariogram is shown in Figure 2.

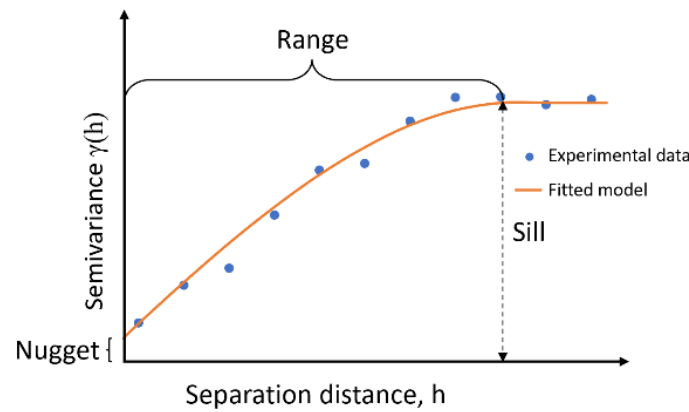


Figure 2. A typical semivariogram (adapted and modified from [9]).

Geostatistics inherently deals with uncertainties, meaning the true semivariogram cannot be directly determined [10]. Hence, only an estimate of the semivariogram can be known. The following unbiased estimator is predominantly used in common practice [10]:

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

where $z(x_i)$ is an observed value at a particular location, and $n(h)$ is the number of pairs at a distance h apart. $\hat{\gamma}(h)$ is the experimental semivariogram, and h is the lag distance. Prior to incorporating semivariogram parameters into the kriging estimation process, it is necessary to determine the most appropriate theoretical model that best represents the experimental semivariogram [11]. Kriging is a geostatistical interpolation technique that estimates unknown values at unsampled locations using the weighted averages of surrounding measurements, which are assigned based on the semivariogram model. A basic kriging estimation is defined as follows:

$$Z^*(X_0) = \sum_{i=1}^n \lambda_i Z(X_i) \quad (2)$$

where $Z^*(X_0)$ is the estimate at location X_0 , λ_i is the kriging weight, $Z(X_i)$ is the known measurements at sampling sites X_i , and n is the number of sampling neighborhood points. The weights are derived from the parameters of the fitted semivariogram models and the relative distance between the measured points and the prediction locations [12].

In this study, multi-variable semivariograms were employed to characterize the complex spatial structure of crucial weather variables, providing a more comprehensive representation of road weather conditions. Compared to single-variable models, they enhance robustness and accuracy by integrating cross-correlated data, thereby leading to more reliable location optimization. Individual semivariograms were constructed for each variable (i.e., RST, DPT, and AT) and subsequently joined together to form a combined semivariogram. This approach was inspired by Walter et al. [13], who first obtained variogram clouds for each year and then combined them into one comprehensive variogram cloud. The parameters obtained from the fitted multi-variable semivariogram model would then be used as critical inputs for the location optimization framework.

As presented in this section, geostatistical analysis, particularly semivariogram modeling, is a complex and time-intensive process, often requiring extensive computational resources for each new development. By leveraging WD, the proposed framework enables the reuse of existing semivariograms, eliminating the need for repetitive modeling and significantly reducing computational overhead. This not only enhances efficiency but also

simplifies implementation, making RWIS optimization more accessible. With this approach, transportation authorities can strategically deploy RWIS stations using existing datasets, facilitating broader adoption of data-driven decision-making in road weather monitoring.

2.3. Wasserstein Distance

Wasserstein Distance (WD), a key concept in this study, is a mathematical metric used to quantify the dissimilarity between two probability distributions. Intuitively, it represents the minimum amount of ‘work’ required to transform one distribution into another [14]. In the context of RWIS optimization, this metric facilitates the comparison of the statistical properties of different highways, enabling a more nuanced approach to location-allocation. The mathematical definition of p-WD between probability distributions μ and ν on $\mathbb{R}^d$ [15] is defined as follows:

$$W_p(\mu, \nu) = \inf(\mathbb{E}X - Y^p)^{\frac{1}{p}}, p \geq 1 \quad (3)$$

where the infimum is taken over all joint distributions of d-dimensional random variables X and Y with marginals μ and ν , respectively. WD values among study routes are obtained using the `wasserstein1d` function in RStudio—version 4.3.1, which computes WD between their empirical distributions. The WD in 1-d [15] is given as follows:

$$W_p(X, Y) = \left(\int_0^1 |F_X^{-1}(\alpha) - F_Y^{-1}(\alpha)|^p d\alpha \right)^{\frac{1}{p}} \quad (4)$$

in which F_X and $F_X^{-1}(q) = \inf\{x : F_X(x) \geq q\}$, $q \in (0, 1)$ are denoted as the distribution and quantile functions of X . A lower WD value means that the two distributions are closer to each other. Hence, after obtaining the WD between every pair of highways, each highway can then be matched with the highway to which it has the lowest WD. A simplified illustration of the WD analysis performed in this study can be seen in Figure 3.

Wasserstein Distance (WD) Analysis

$$W_p = \left(\int_0^1 |F^{-1}(x) - G^{-1}(x)|^p dx \right)^{1/p}$$

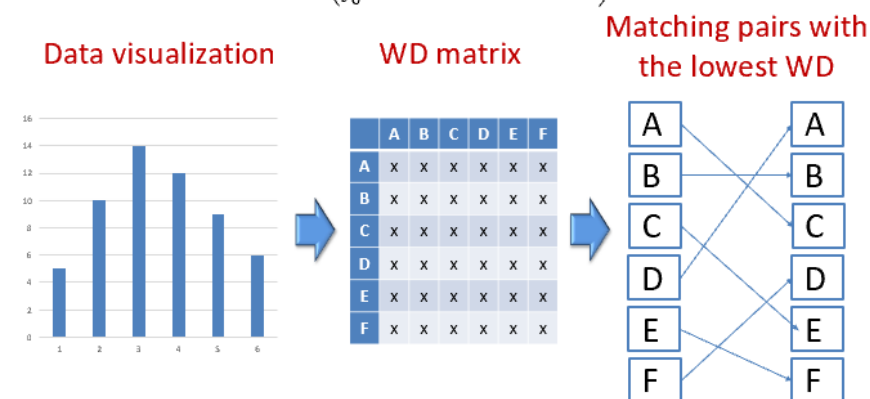


Figure 3. A simplified illustration of Wasserstein Distance analysis.

2.4. RWIS Location Optimization Framework

In this study, an RWIS location allocation framework was adopted from one of our well-adopted previous studies [9], in which the objective function is formulated to minimize the sum of mean ordinary kriging (OK) estimation variance. The equations of the objective function and relevant computation process [9] are shown below:

$$G = \begin{bmatrix} \gamma(x_1, x_1) & \gamma(x_2, x_1) & \dots & \gamma(x_k, x_1) & 1 \\ \gamma(x_1, x_2) & \gamma(x_2, x_2) & \dots & \gamma(x_k, x_2) & 1 \\ \vdots & \vdots & \dots & \vdots & \vdots \\ \gamma(x_1, x_k) & \gamma(x_2, x_k) & \dots & \gamma(x_k, x_k) & 1 \\ 1 & 1 & \dots & 1 & 0 \end{bmatrix} \quad (5)$$

where x_i ($i = 1, 2, \dots, k$) is the sampling site of a sample subset of size k , which is the number of RWIS stations. The semivariance between sampling site i and j is $\gamma(x_i, x_j)$.

$$g = [\gamma(x_0, x_k) \gamma(x_0, x_2) \dots \gamma(x_0, x_k) 1]' \quad (6)$$

Here, x_0 denotes estimation location. The minimum mean square error for ordinary kriging for the estimation location x_0 is as follows:

$$\sigma_{OKI}^2(x_0) = g' G^{-1} g \quad (7)$$

Based on those equations, the objective function can be formulated as follows:

$$f(w) = \frac{\sum_{i=1}^{n-k} \sigma_{OKI}^2(x_0)}{n} \quad (8)$$

where n is the total number of candidate RWIS stations (fishnet polygons) in the study area.

The optimization process in this study is divided into two distinct phases: preferred location selection and clean slate optimization. For the first phase, all the pairs of the proposed locations are divided into separate groups (each group containing one pair), with one location from each group selected to form a final set of solutions. The method of exhaustive search is then implemented to continuously pick out different combinations of locations and calculate their criterion value until an optimal solution is obtained. As for the second phase, Spatial Simulated Annealing (SSA), an effective and robust method for finding solutions to optimization problems, is implemented to find the optimal solutions for an objective function. This phase of the optimization implicitly considers geographic equity by aiming to minimize kriging variance, which in turn maximizes monitoring capability and ensures the equitable distribution of monitoring resources across various geographic locations. SSA is a widely used heuristic algorithm that operates by making slight perturbations to previous sampling points through random search techniques [16]. To avoid being stuck in local minima, not only does SSA accept improving solutions, but it also accepts worsening solutions based on a certain probability [9]. Worsening solutions are accepted at an initial probability of 0.2, which decreases exponentially to zero as a function of the iteration number. The maximum number of iterations is set to 10,000, above which the optimization process will stop. Similarly, the algorithm will also stop if no improvements are made after 200 iterations.

3. Study Area and Data Details

3.1. Study Area

The study area consists of eight highways in South Korea: Jungbu line, Gwangju Daegu line, Seoul Yangyang line, Suncheon Wanju line, Namehae line 1, Namhae line 2, Donghae line 1, and Pyeongtaek Jecheon line. The study highways, ranging from 100 km to 200 km in length, encompass topographically distinctive regions in South Korea, such as mountains, lowlands, and coastlines. The diverse array of topography also covers varying weather conditions across the study area. These roadlines currently do not have any RWIS stations installed, so to inform their eventual placements, weather and road surface

data were collected using mobile RWIS along the roadways. Data collection occurred on various nights between November 2022 and January 2023, when the temperatures did not change vastly over the data collection period, ignoring the temporal variability in weather conditions induced during the daytime. Figure 4 presents a map of South Korea, highlighting the eight study highways.



Figure 4. Study area—eight highway segments in South Korea.

3.2. Data Description and Processing

The mobile RWIS data provided by Mirae Climate Co., Ltd., South Korea, included, among other parameters, RST, DPT, and AT. These variables were collected using mobile RWIS every second along its travel route. As some measurements appeared to be erroneous with missing longitude, latitude, and temperature values, they were consequently removed from the dataset. Similarly, data completeness was then checked to see if the proportion of missing data was less than 10% of the total number of data points and that the remaining data were suitable for use. If more than 10% of the total number of data points are deemed erroneous (i.e., missing geographical coordinates, missing data values, etc.), the dataset was not used for analysis. The three types of weather data (RST, DPT, and AT) were then plotted and cross-checked with one another to detect any values that deviated from expected patterns. Abnormalities in data were also examined by superimposing the data points onto the world map in ArcGIS [17] and identifying locations with sudden peaks and troughs within the data distribution. These anomalies were primarily associated with data collected inside tunnels along the road networks and were subsequently removed to ensure the continuity of data distribution.

The descriptive statistics of weather parameters for all eight highways are shown in Table 1, providing values of minimum (Min), average (Avg), maximum (Max), and standard

deviation (SD) of the collected data. Upon closer inspection of Table 1, it is observed that RST shows a range of -9.515°C to 10.316°C with varying averages for all highways, whereas DPT varies from -21.462°C to -3.886°C with inconsistent averages. AT ranges from -13.264°C to 15.356°C with averages from -11.754°C to -1.335°C . These statistics provide an insightful synopsis of the temperature variations across the eight highways. Since the temperatures vary greatly, the RST, DPT, and AT variables were then normalized to minimize bias in WD analysis. Min–Max normalization was used for its ability to scale all data to a common range and preserve the shape of the original distribution while maintaining relative relationships between data points. The normalization method applied to the data has the formula as shown in Equation (9):

$$v' = \frac{v - \min}{\max - \min} \quad (9)$$

where v' is the normalized value, v is the original value, $\min$ and $\max$ are the minimum and maximum values of the data. Lastly, the data points on each route were also aggregated every kilometer, reducing the number of data points used for the construction of semivariograms. Aggregated sections without any data points were deleted. The numbers of data points before and after data cleaning and the numbers of aggregated sections for each route are seen in Table 2.

Table 1. Descriptive statistics of weather parameters for 8 highways.

		Jungbu	Gwangju Daegu	Seoul Yang Yang	Suncheon Wanju	Namhae Line 2	Namhae Line 1	Donghae Line 1	Pyeongtaek Jecheon
RST ($^{\circ}\text{C}$)	Min	−9.085	−3.570	−9.515	−9.323	−4.812	−3.439	−1.499	−7.418
	Avg	−6.398	−0.835	−5.141	−5.305	−1.300	−0.802	2.206	−4.517
	Max	−2.369	4.120	−0.933	1.422	7.854	4.251	10.316	0.610
	SD	0.978	1.056	1.462	1.464	1.168	0.986	1.743	1.142
DPT ($^{\circ}\text{C}$)	Min	−15.256	−9.306	−14.897	−14.020	−8.388	−9.027	−21.462	−12.108
	Avg	−13.248	−7.150	−11.918	−11.367	−6.652	−7.825	−17.045	−10.342
	Max	−10.722	−3.886	−8.175	−7.926	−4.122	−6.115	−10.751	−7.556
	SD	1.017	1.262	1.243	1.289	0.859	0.626	1.816	0.738
AT ($^{\circ}\text{C}$)	Min	−13.264	−7.647	−12.975	−13.001	−6.962	−6.749	−4.291	−11.357
	Avg	−11.754	−5.051	−7.830	−9.328	−3.703	−4.152	−1.335	−8.841
	Max	−9.573	2.673	15.356	6.037	6.008	3.718	9.761	1.611
	SD	0.718	1.492	2.876	2.491	1.944	1.804	2.080	1.153

Table 2. Number of data points of raw data, post-data cleaning, and post-data aggregation.

Highway	Length (km)	Raw Data	After Data Cleaning	Number of Aggregated Sections
Jungbu Line	119	13,883	13,803	94
Gwangju Daegu	171	27,104	24,075	166
Seoul Yang Yang	151	66,358	12,153	118
Suncheon Wanju	118	17,734	12,532	114
Namhae Line 1	166	26,148	22,874	164

Table 2. Cont.

Highway	Length (km)	Raw Data	After Data Cleaning	Number of Aggregated Sections
Namhae Line 2	107	17,493	13,533	102
Donghae Line 1	126	20,179	18,268	121
Pyeongtaek Jecheon	126	19,905	15,693	120

The large discrepancy between the number of data points before and after the data cleaning of the Seoul Yang Yang line was a result of the rerouting of an RWIS vehicle taking a detour outside of the designated study area.

4. Results and Discussions

4.1. Semivariogram Modelling

Spatial semivariogram modeling was performed to examine the range of spatial autocorrelation for the study area. All three variables, AT, RST, and DPT, were selected and examined to fully represent the road surface conditions. Previous studies have suggested that the construction of a reliable semivariogram requires 30 or more sampling points [10], and this requirement was met as we have a large sample size. Variogram clouds for each variable were obtained using the R statistical package gstat [18]—version 4.3.1—and then combined to acquire the multi-variable variogram for all three variables. Subsequently, the average semivariance for every determined lag distance was calculated to construct a combined experimental semivariogram for the eight highways. The utilization of combined multi-variable semivariograms was predicted to provide a more reliable location solution by capturing the fluctuation of several weather variables. By investigating the general shape and trend of the experimental semivariograms, the Gaussian model was chosen to fit the plots. The formula for the Gaussian model [11] is as follows:

$$\gamma(h) = C_0 + C_1 \left[1 - \exp\left(-\frac{3h^2}{a^2}\right) \right] \quad (10)$$

where C_0 is the nugget; C_1 denotes the sill; h is the lag distance; and a is the range of influence. The generated nugget, sill, and range for each line are shown in Table 3:

Table 3. Semivariogram parameters.

Route	Nugget	Sill	Range (m)
Jungbu Line	0.004768	0.04662	26,880.34
Gwangju Daegu	0.009627	0.03894	35,233.74
Seoul Yang Yang	0.005744	0.04708	60,084.41
Suncheon Wanjū	0.007561	0.02477	17,677.41
Namhae Line 2	0.01026	0.01690	20,400.53
Namhae Line 1	0.009895	0.04301	59,446.34
Donghae Line 1	0.005802	0.07855	75,679.68
Pyeongtaek Jecheon Line	0.002962	0.08381	95,301.42

As Donghae Line 1 and Pyeongtaek Jecheon have a range longer than 50% of the road length, they are not used for the RWIS location optimization stage. The semivariograms of all six highways are shown in Figure 5.

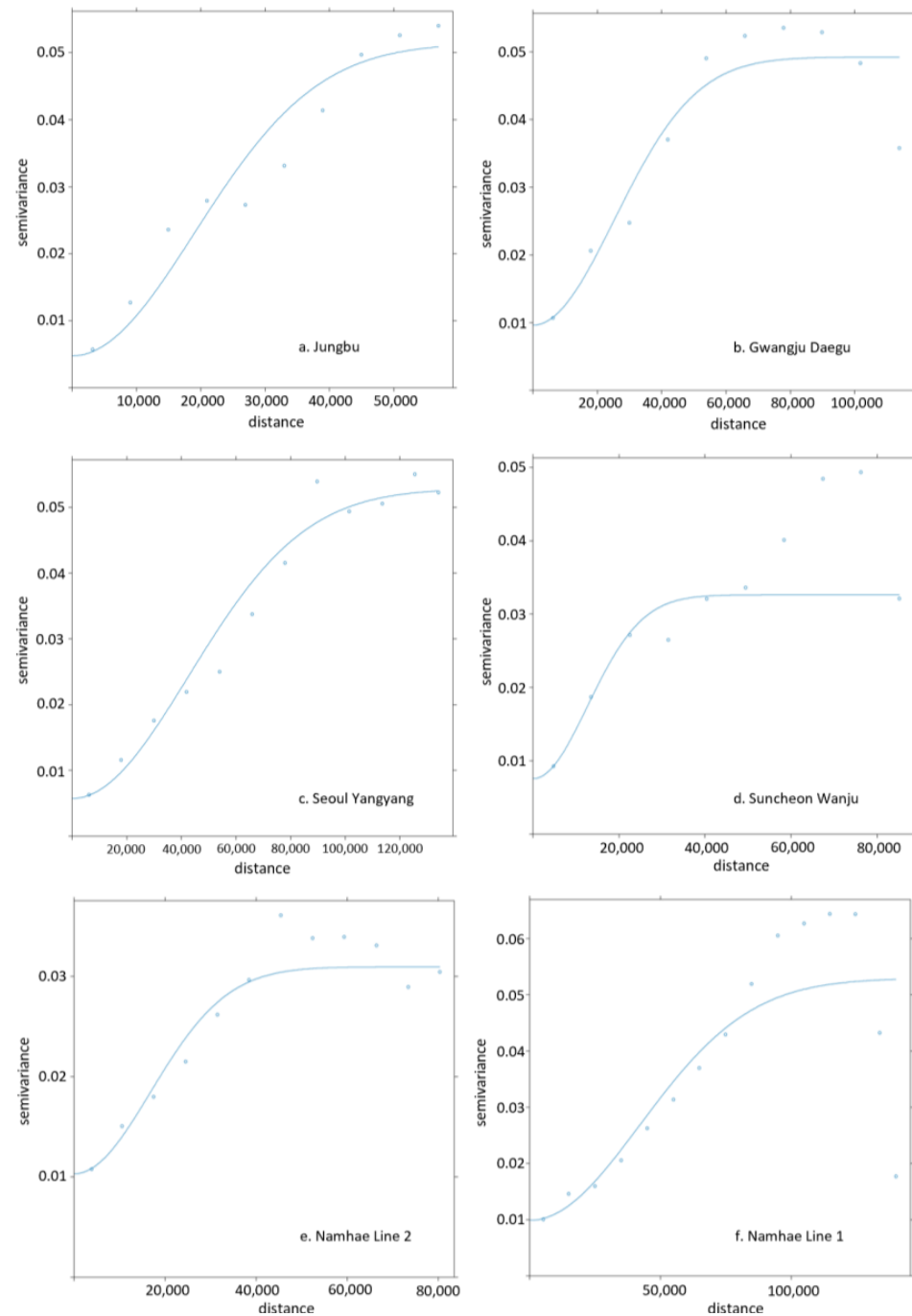


Figure 5. Semivariogram models for six highways.

4.2. Wasserstein Distance Analysis

The function `wasserstein1d` in the `transport` package in RStudio [19] was used to obtain the WD value for each of the variables RST, DPT, and AT among the eight highways. In total, 28 pairs of highways were formed with three variables each, resulting in 84 calculated WD values. Subsequently, the WD values of three variables of each pair were averaged to obtain the average WD for each pair. Then, pairs of highways with the lowest WD values were identified, as shown in Table 4. These paired highways were considered to have the most similar characteristics to each other in terms of probability distributions. Hence, the semivariogram parameters of matched highways were then used to perform RWIS location optimization for their corresponding original highways in scenario 2 in each phase.

Table 4. Pairs of matched highways from WD analysis.

Original Highway	Matched Highway	Wasserstein Distance
Suncheon Wanju	Pyeongtaek Jecheon	0.0461
Jungbu	Namhae Line 1	0.0958
Gwangju Daegu	Namhae Line 1	0.0376
Seoul Yang Yang	Suncheon Wanju	0.0790
Namhae Line 2	Namhae Line 1	0.0531
Namhae Line 1	Gwangju Daegu	0.0376
Donghae Line 1	Namhae Line 1	0.0521
Pyeongtaek Jecheon	Suncheon Wanju	0.0461

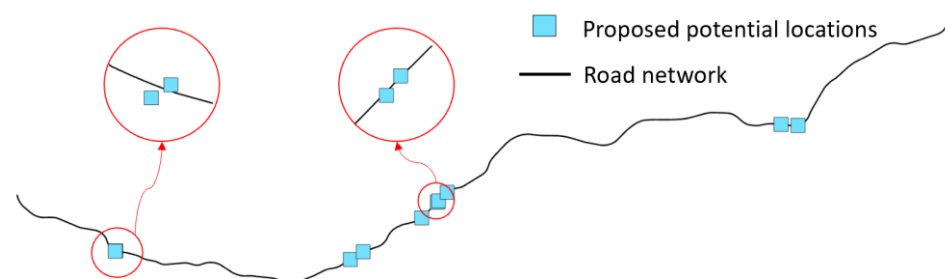
4.3. Optimal RWIS Location Analysis and Evaluation

The semivariogram parameters obtained previously were then used to determine optimal RWIS locations. The optimization problem was formulated and solved using the R statistical package [18]—version 4.3.1. A 200 m × 200 m prediction grid was generated in ArcGIS [17] to cover the entire road network for each line to create the study area, as well as the candidate sites for RWIS stations on each highway.

As mentioned earlier, this study is divided into two specific phases for building the RWIS network and validating the implementation of WD in the RWIS location optimization framework. The description of each phase is detailed below.

Phase 1—Selection of Preferred Locations

Multiple potential RWIS locations (candidate locations) have been chosen by Korea's transportation authorities. Due to limited resource constraints, only half of the potential locations could be chosen for RWIS installations. Thus, the first phase of our study was to select the preferred solution out of every pair of potential sites along a road network. For instance, given a road network with ten proposed sites, one out of every pair of potential locations should be chosen to obtain a total of five preferred locations for RWIS station placement. Figure 6 shows the potential locations for Namhae Line 2 proposed by Korea's transportation authorities as an example. The main goal was to maximize the monitoring coverage by minimizing the kriging variance. Therefore, the combination of locations that produced the lowest kriging variance would be chosen as the preferred locations. The number of optimal solutions needed to be selected for the remaining six highways, Jungbu, Gwangju Daegu, Seoul Yangyang, Suncheon Wanju, Namhae Line, and Namhae Line 2, were 16, 14, 18, 16, 14, and 5, respectively.

**Figure 6.** A demonstration of candidate locations for Namhae Line 2.

To evaluate the validity of using WD to utilize previously developed semivariograms in the location optimization framework, two scenarios were considered for the selection of preferred locations.

1. Using original semivariogram models
Here, the semivariogram model developed for a specific highway was used to perform the location optimization process for that same highway.
2. Using WD-matched semivariogram models
To assess the effectiveness of applying WD, this scenario employed the semivariogram model of one highway to perform the location optimization process for its matched highway, as detailed in Table 4. For example, the WD solution for the Jungbu line would be found using Namhae Line 1's semivariogram parameters for the optimization process. Conversely, Suncheon Wanju's WD solution was generated using Gwangju Daegu's semivariogram parameters because Pyeongtaek Jecheon's semivariogram range exceeded half of Suncheon Wanju's total road length.

Therefore, two sets of preferred locations were generated for each highway. The solutions for scenario 1 were termed "original solutions", and those for scenario 2 were termed "WD solutions". The preferred solutions derived from this process are presented in Figure 7.

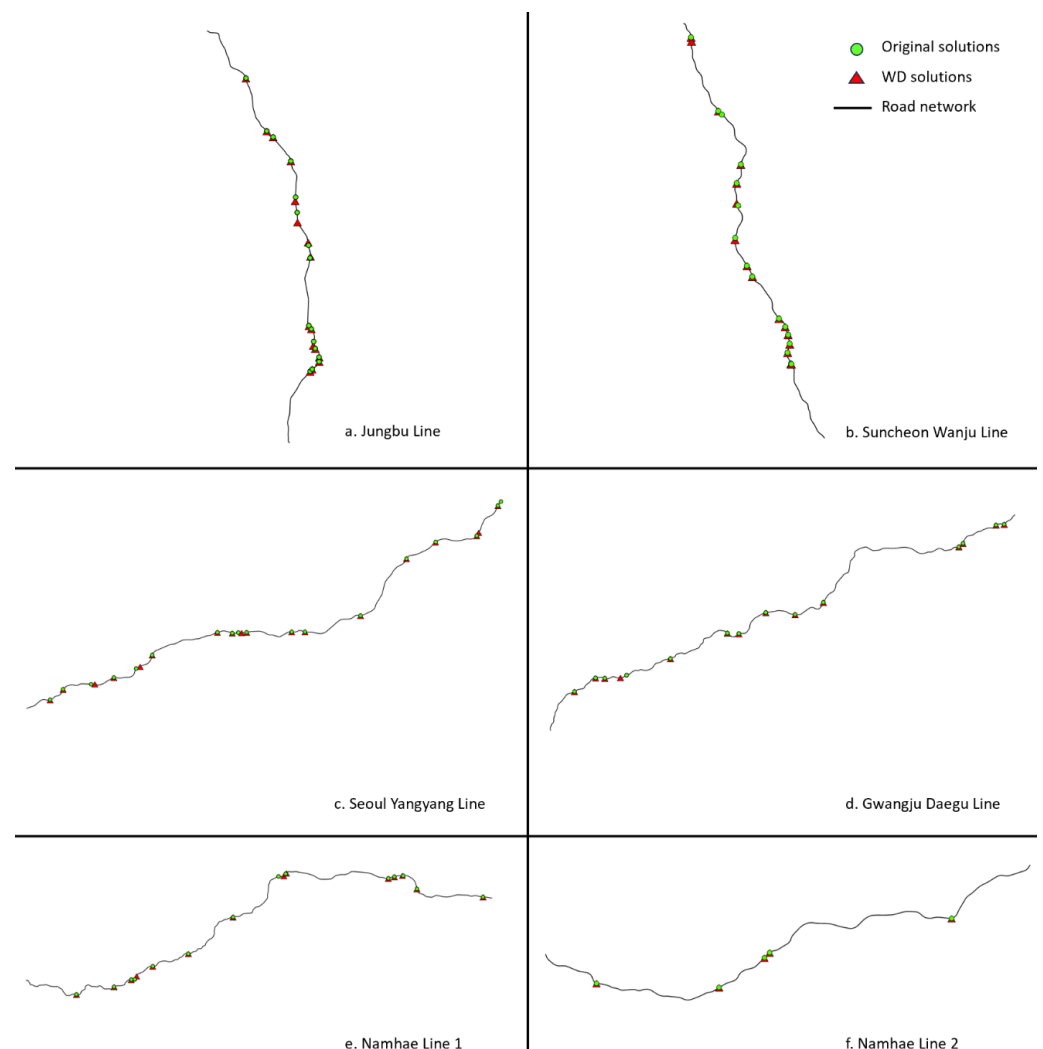


Figure 7. Preferred RWIS location sets for each highway under two scenarios showing the similarity between the two solutions.

Figure 7 shows the optimal locations for six highways, with green circles representing the preferred locations for scenario 1 and red triangles representing preferred locations for scenario 2 on the road networks. When the original solution locations and WD solution

locations coincide, the green circles are shown on top of the red triangles. As seen in Figure 7, many of the preferred locations generated in the two scenarios overlap, especially for Namhae Line 2, where all the selected locations were overlapped, and for Gwangju Daegu Line, where only one station was not overlapped. This indicates that WD-matched semivariogram models can yield solutions similar or identical to those produced by original semivariogram models. This shows the consistency of using WD-matched semivariogram models in identifying preferred locations. It signifies that the WD-matched semivariogram models are robust and reliable alternatives to the original semivariogram models.

In addition to the visual confirmation, the effectiveness of using WD is quantitatively assessed by computing the matching percentage for each highway, as shown in Table 5. The matching percentage for each highway was determined by calculating the proportion of overlapping solution locations between the original solutions and WD solutions relative to the total number of preferred locations selected. The average matching percentage was found to be 81%, indicating a relatively high alignment between the solutions generated using the original semivariograms and those using the WD-matched semivariograms. These findings demonstrate the effectiveness of incorporating WD into the RWIS location allocation framework.

Table 5. Matching percentage for two scenarios of preferred solutions.

Highway	Overlapping Stations	Number of Locations Selected	Matching Percentage
Jungbu Line	11	16	69%
Gwangju Daegu	13	14	93%
Seoul Yang Yang	14	18	78%
Suncheon Wanju	13	16	82%
Namhae Line 2	5	5	100%
Namhae Line 1	9	14	64%
Average:			81%

Although the WD-matched solutions and the original solutions exhibit an 81% match, several factors contribute to the remaining 19% mismatch. Circumstances such as discrepancies in highway lengths, which result in vastly different semivariogram ranges, may hinder WD's ability to approximate the original semivariograms. Additionally, differences in data density across different highways could also contribute to the mismatch. The presence of tunnels in the study areas, which caused highly fluctuating and abnormal data that were subsequently removed, may have introduced some discontinuities in the data distribution along the road networks, further affecting the WD's performance.

Phase 2—Clean slate optimization

In this phase, the candidate locations from *Phase 1* are instead expanded to cover the entire road network, excluding regions inside tunnels. The purpose of this analysis was to objectively determine the optimal use of available resources without the constraint of predetermined locations. For this optimization, SSA was employed, a well-established heuristic algorithm used to tackle spatial optimization problems. The two scenarios, as described in Phase 1, were also employed for this phase, namely the use of the original semivariograms and the WD-matched semivariograms. This was performed to evaluate how the WD method fares in a broader and less constrained optimization process. Therefore, this phase serves to further demonstrate the robustness of the proposed method by allowing for a clean-slate optimization, ensuring the most strategically advantageous sites are selected independently of any prior constraints. Moreover, three trials were conducted

to find multiple sets of solutions for each highway, ultimately selecting the solution with the lowest criterion value. In total, the clean-slate optimization was run 36 times. Figure 8 illustrates the optimal locations derived from this clean-slate optimization process, where the green circles represent the locations obtained from scenario 1, while the red triangles represent those obtained from scenario 2.

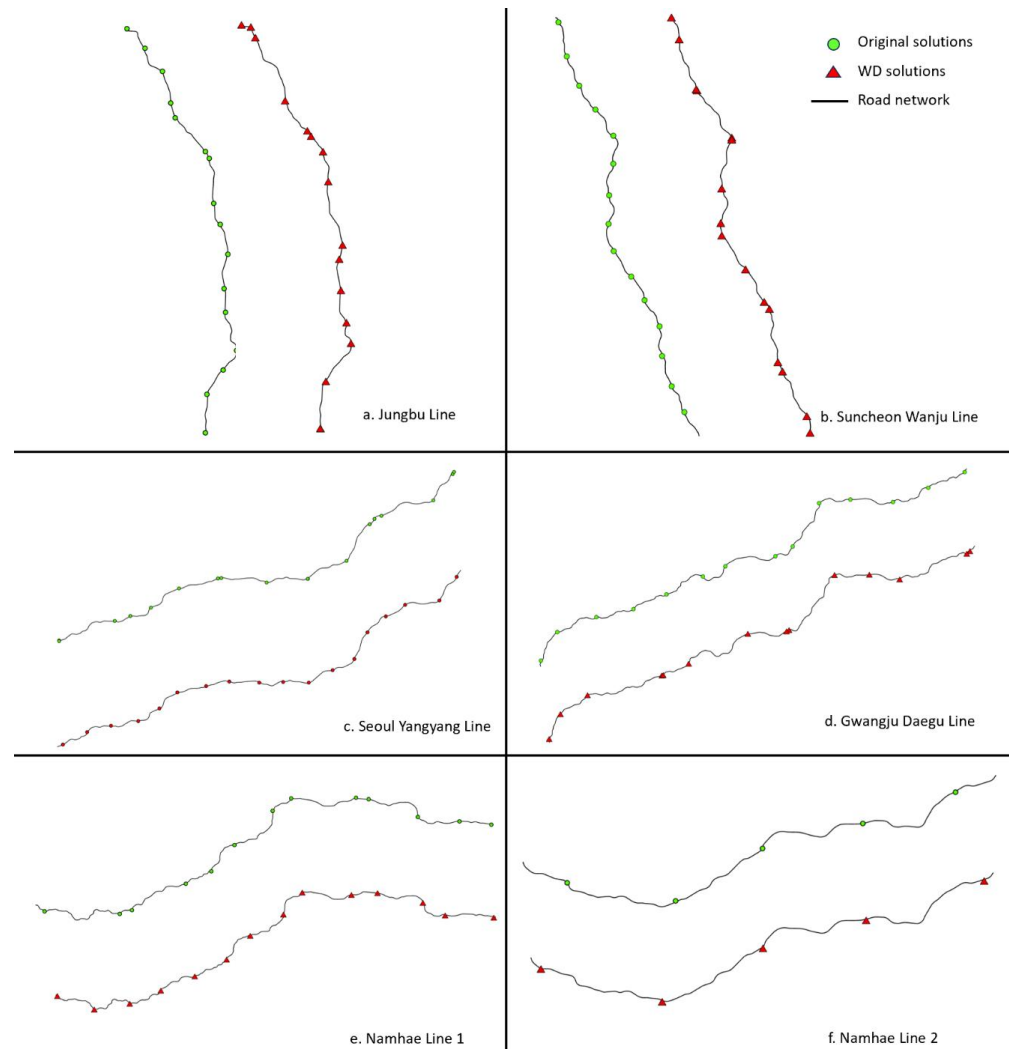


Figure 8. Clean-slate optimization RWIS location sets for each highway under two scenarios showing the similarity between the two solutions.

The resulting clean-slate optimization solution exhibited a noticeably more even distribution along the highways in both scenarios compared to the preferred locations, encompassing a more extensive coverage. This improvement in distribution explicitly supports geographic equity by removing potential location constraints, allowing the optimization process to explore the entirety of the road networks for optimal sites. Geographic equity refers to the fair and balanced distribution of RWIS stations to maximize monitoring coverage across road networks. By minimizing kriging variance (i.e., estimation uncertainty), the proposed approach implicitly considers all regions for RWIS deployment, ensuring that road users in both urban and rural areas benefit from improved road weather monitoring. It is seen in Figure 8 that the original solutions and the WD solutions are highly similar to each other, with WD solutions positioned in close proximity to the original solutions. The close positioning of these two solutions implies that the approach using the WD-matched semivariograms could effectively maintain the relevance of the optimal

solutions generated from implementing the original semivariogram models. This suggests that WD is a valid method that can be integrated within the RWIS location optimization framework, as it allows the preservation of the balanced and extensive coverage of optimal RWIS locations generated using the traditional approach.

Aside from the visual representation, the validity of using WD to adopt previously developed semivariograms was also quantitatively examined. The metric used for this assessment in Phase 2 is weighted separation distance, as detailed in Table 6. For clean-slate solutions, the separation distance was the sum of all the distances between every original location and its closest WD location within a highway, divided by the number of selected locations. The weighted separation distance was then derived by dividing the separation distance by the total road length. The reason for using weighted separation distance was to consider all factors that could influence the location solutions, including the proximity between two sets of solutions, the length of the highways, and the number of sites to be selected.

Table 6. Weighted separation distance between two scenarios of clean slate solutions.

Highway	Length (km)	Sum of Distances (km)	Number of Locations Selected	Separation Distance (km/Site)	Weighted Separation Distance (km/Site·km)
Jungbu	119	63.79	16	3.987	0.0336
Gwangju Daegu	171	56.64	14	4.046	0.0236
Seoul Yang Yang	151	85.23	18	4.735	0.0314
Suncheon Wanju	118	40.29	16	2.518	0.0213
Namhae Line 2	107	18.53	5	3.707	0.0347
Namhae Line 1	166	40.65	14	2.903	0.0175

Table 6 shows that the separation distance between the original solutions and the WD solutions is as low as 2.5 km per site and as high as 4.7 km per site. The weighted separation distance reveals that the average separation distance between each site does not exceed 3.5% of the entire road length. Figure 9 demonstrates a proportional relationship between WD and weighted separation distance, with a Spearman correlation value of 0.493. This correlation, while moderate, indicates a positive association between two variables. The trend shows that as the WD increases, there is a corresponding increase in the weighted separation distance, indicating that a lower WD value between two sets of data will produce solution locations closer to each other.

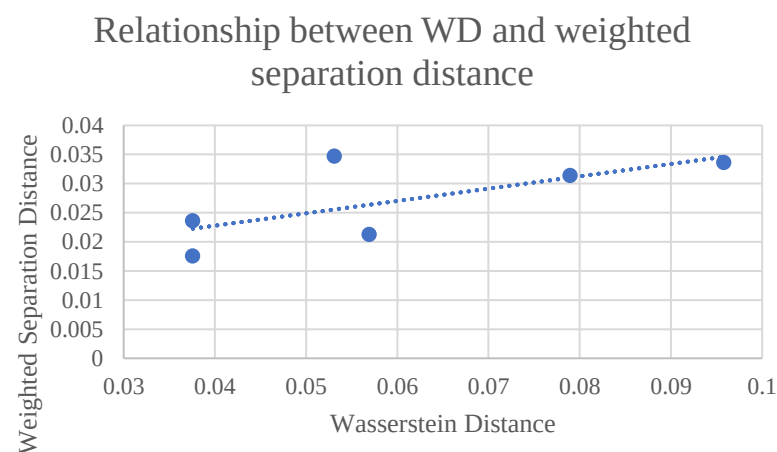


Figure 9. Graph of the relationship between WD and weighted separation distance.

This insight asserts the practicality of WD in the optimization process, particularly when integrating new datasets. When a new dataset is introduced, the WD between the new dataset and the datasets with previously developed semivariograms can be computed to identify the pair with the most closely aligned meteorological distribution. Subsequently, the semivariograms from this closely matching dataset can then be adopted for the new dataset's geostatistical analysis. This approach simplifies the location optimization process by utilizing existing semivariograms, eliminating the need to create a new semivariogram for each study area. Existing RWIS location optimization methods often require extensive geostatistical analysis, which is time-intensive and technically demanding. The proposed approach significantly reduces computational complexity, enhances efficiency, and broadens accessibility. Ultimately, this allows transportation authorities to implement RWIS optimization without technical expertise, thus bridging the gap between advanced analytics and operational decision-making while ensuring a more geographically equitable deployment of RWIS.

4.4. Recommendations for Practical Implementation

The research findings present a streamlined and practical approach showcasing the potential for RWIS location optimization frameworks, specifically those incorporating geostatistics, to be widely adopted beyond academic and specialized research settings. In particular, local administration and transportation agencies can follow the steps outlined below to implement the proposed methods in practice.

1. Gather existing road surface and weather conditions for relevant regions at the municipal, provincial, or national level;
2. Engage geostatistical experts to develop multi-variable semivariograms from these data and store them in a centralized repository;
3. When deploying RWIS stations in a new region, personnel with minimal training can perform a Wasserstein Distance analysis to identify the dataset most similar to the newly introduced one;
4. Reutilize the semivariogram of the matched dataset to perform RWIS location optimization in the new region.

These steps streamline RWIS deployment, reducing reliance on specialized expertise and enabling efficient, data-driven decision-making. By reusing semivariograms, agencies can minimize computational costs and ensure equitable monitoring coverage, even in regions with limited data availability.

Beyond technical efficiency, the proposed framework has key policy implications. Establishing a national or regional semivariogram repository would allow resource-limited agencies to optimize RWIS locations without conducting costly new analyses. Policymakers should integrate this approach into infrastructure planning, prioritizing RWIS deployment in high-risk areas to enhance road safety. Funding mechanisms, such as government grants or public-private partnerships, can further support equitable implementation, particularly in underserved regions.

Furthermore, regulatory frameworks could incorporate WD-based RWIS planning as a standardized practice. Strengthening inter-agency collaboration between national and local transportation departments will enhance data-sharing to ensure a more cost-effective and scalable expansion of RWIS networks.

5. Conclusions

This study introduced the use of Wasserstein Distance (WD) as an innovative approach to rationalizing the application of previously developed semivariograms for new datasets, thereby eliminating the need to conduct multiple geostatistical analyses for future

study sites. Utilizing WD, a sophisticated location optimization model was developed by integrating a multi-variable semivariogram composed of three types of mobile RWIS measurements (i.e., RST, DPT, and AT) for each road network. Location optimization was then performed using the exhaustive search method for Phase 1 and the SSA algorithm for Phase 2, generating a total of 24 sets of solutions. In Phase 1, the preferred locations derived from the original semivariogram parameters of each highway showed adequate alignment, with an 81% overall match rate, with those generated using WD-matched semivariograms. In Phase 2, clean-slate solutions showcased a more equitable distribution across the network, affirming the method's practical utility and contributing to geographic fairness in monitoring capabilities. The proximity of solution locations produced using the original and WD-matched semivariogram parameters, ranging from 2.5 km to 4.7 km per site, further validates that the proposed WD-based method can effectively optimize RWIS station placement, even in areas lacking extensive RWIS data.

The methodologies developed and insights gained from this study have resulted in several substantial contributions. A key achievement is the effective implementation of the WD, which has significantly streamlined the RWIS deployment process. Traditionally, semivariogram modeling is a resource-intensive and technically demanding task. By applying WD to link semivariograms across varying datasets, this innovative approach obviates the need for new model development for each study area, thereby saving considerable time and reducing the level of expertise required. This advancement is particularly advantageous for transportation agencies tasked with installing RWIS across extensive regions characterized by varied geographical and climatic conditions, as this approach does not necessarily require new data collection to conduct optimal location analyses. By demonstrating a simplified and practical approach, this study highlights the potential for geostatistical tools to be widely adopted beyond specialized research domains, enabling their application in real-world scenarios. This research recommends the development of a centralized repository of semivariograms derived from historical meteorological datasets across various regions. Thus, when new datasets are introduced, WD analysis can be used to match them with the most compatible existing data, allowing transportation agencies to reuse previously developed semivariograms for RWIS location optimization. This approach streamlines the geostatistical analysis, enhancing efficiency and broadening the accessibility of RWIS planning for agencies with varying technical capabilities.

While the study yielded valid results, opportunities for further improvement remain. First, although the proposed locations consisted of four types of RWIS, this study treated them under the same criterion. Future research should consider them separately, with criteria tailored to each type of RWIS to better reflect their unique functions and potential influence on optimal RWIS placement. Secondly, only three different variables (AT, RST, and DST) were used in this study. Expanding the analysis to include additional RWIS variables, such as humidity, precipitation intensity, and friction, would provide a more comprehensive assessment of optimal RWIS placement. Humidity could be examined along with DPT to identify conditions conducive to frost or black ice formation. For instance, precipitation intensity, such as snowfall and rainfall rates, directly impacts road surface conditions and visibility, making it a critical factor for RWIS optimization. Furthermore, friction levels influence driver maneuverability and safety, particularly in inclement winter conditions. The use of these variables may necessitate the development of specialized models tailored to their spatial and temporal characteristics. Additionally, a larger and more diverse dataset is needed to increase the robustness of the semivariogram models and validate the generalizability of the findings. Incorporating data from multiple regions with varying climatic and topographical characteristics would ensure that the proposed method remains applicable across different environments, reinforcing the conclusiveness

of the results presented in this study. Moreover, time should be considered in conjunction with space to examine the effect of spatiotemporal structure on RWIS location-allocation. The development of space-time semivariograms could enable a more dynamic analysis, capturing both spatial and temporal autocorrelation of road weather variables to support RWIS placement that accounts for spatiotemporal variability. Lastly, the effectiveness of incorporating WD in the location-allocation framework can be examined by performing spatial analysis using buffer zones and sensitivity analysis for buffer size. Nonetheless, the methods and findings presented herein deliver substantial advancements in RWIS deployment strategies, providing valuable insights and tools that significantly contribute to the field of road weather management.

Author Contributions: Conceptualization, N.H., J.J. and T.J.K.; Data curation, N.H. and J.J.; Formal analysis, N.H., J.J. and T.J.K.; Investigation, N.H. and T.J.K.; Methodology, N.H., J.J. and T.J.K.; Project administration, J.J. and T.J.K.; Resources, J.J. and T.J.K.; Software, N.H. and T.J.K.; Supervision, T.J.K.; Validation, N.H. and T.J.K.; Visualization, N.H.; Writing—original draft, N.H. and T.J.K.; Writing—review and editing, N.H. and T.J.K. All authors have read and agreed to the published version of the manuscript.

Funding: Support was received from the Natural Sciences and Engineering Research Council of Canada under grant agreement no. RGPIN-2024-05039, which has significantly contributed to the success of this research.

Institutional Review Board Statement: Not Applicable.

Informed Consent Statement: Not Applicable.

Data Availability Statement: Data collected and analyzed for the purpose of this study are not readily available due to privacy concerns as the data are completely owned by Mirae Climate Co., Ltd. Requests to access the datasets should be directed to Mirae Climate Co., Ltd.

Acknowledgments: The authors would like to thank Mirae Climate Co., Ltd. for providing the necessary data for this study.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. El Esawey, M.; Walker, S.; Sowers, C.; Sengupta, J. Safety Assessment of the Integration of Road Weather Information Systems and Variable Message Signs in British Columbia. *Transp. Res. Rec. J. Transp. Res. Board* **2019**, *2673*, 305–313. [\[CrossRef\]](#)
2. El-Rayes, K.; Ignacio, E.-J. Evaluating the Benefits of Implementing Mobile Road Weather Information Sensors. *FHWA-ICT-22-004* **2022**. [\[CrossRef\]](#)
3. Jin, P.J.; Walker, A.; Cebelak, M.; Walton, C.M. Determining Strategic Locations for Environmental Sensor Stations with Weather-Related Crash Data. *Transp. Res. Rec. J. Transp. Res. Board* **2014**, *2440*, 34–42. [\[CrossRef\]](#)
4. Eriksson, M.; Norrman, J. Analysis of station locations in a road weather information system. *Meteorol. Appl.* **2001**, *8*, 437–448. [\[CrossRef\]](#)
5. Zhao, L.; Chien, S.; Meegoda, J.; Luo, Z.; Liu, X. Cost–Benefit Analysis and Microclimate-Based Optimization of a RWIS Network. *J. Infrastruct. Syst.* **2016**, *22*, 04015021. [\[CrossRef\]](#)
6. Singh, A.K.; Li, J.; Murphy, M.; Walton, C.M. Optimizing Environmental Sensor Station Locations for Road Weather Management: Overview and a Generalized Modeling Framework. *Transp. Res. Rec. J. Transp. Res. Board* **2016**, *2551*, 18–28. [\[CrossRef\]](#)
7. Fetzer, J.; Caceres, H.; He, Q.; Batta, R. A multi-objective optimization approach to the location of road weather information system in New York State. *J. Intell. Transp. Syst.* **2018**, *22*, 503–516. [\[CrossRef\]](#)
8. Al-Kaisy, A.; Ewan, L. Prioritization Scheme for Proposed Road Weather Information System Sites: Montana Case Study. *Front. Built Environ.* **2017**, *3*, 45. [\[CrossRef\]](#)
9. Biswas, S.; Kwon, T.J. Development of a Novel Road Weather Information System Location Allocation Model Considering Multiple Road Weather Variables over Space and Time. *Transp. Res. Rec. J. Transp. Res. Board* **2022**, *2676*, 619–632. [\[CrossRef\]](#)
10. Olea, R.A. A six-step practical approach to semivariogram modeling. *Stoch. Environ. Res. Risk Assess.* **2006**, *20*, 307–318. [\[CrossRef\]](#)
11. Mert, B.A.; Dag, A. A Computer Program for Practical Semivariogram Modeling and Ordinary Kriging: A Case Study of Porosity Distribution in an Oil Field. *Open Geosci.* **2017**, *9*, 663–674. [\[CrossRef\]](#)

12. Moraga, P. *Spatial Statistics for Data Science: Theory and Practice with R*; Chapman & Hall/CRC Data Science Series; CRC Press: Boca Raton, FL, USA, 2023.
13. Walter, J.F.; Christman, M.C.; Hoenig, J.M.; Mann, R. Combining data from multiple years or areas to improve variogram estimation. *Environmetrics* **2007**, *18*, 583–598. [[CrossRef](#)]
14. Cohen, S. Finding Color and Shape Patterns in Images. Doctor of Philosophy's Dissertation, Stanford University, Stanford, CA, USA, 1999.
15. Panaretos, V.M.; Zemel, Y. Statistical Aspects of Wasserstein Distances. *Annu. Rev. Stat. Appl.* **2019**, *6*, 405–431. [[CrossRef](#)]
16. van Groenigen, J.W.; Siderius, W.; Stein, A. Constrained optimisation of soil sampling for minimisation of the kriging variance. *Geoderma* **1999**, *87*, 239–259. [[CrossRef](#)]
17. ArcGIS Pro. 2024. Available online: <https://www.esri.com/en-us/arcgis/products/arcgis-pro/overview> (accessed on 8 November 2024).
18. Pebesma, E.J. Multivariable geostatistics in S: The gstat package. *Comput. Geosci.* **2004**, *30*, 683–691. [[CrossRef](#)]
19. Schuhmacher, D.; Bähre, B.; Gottschlich, C.; Hartmann, V.; Heinemann, F.; Schmitzer, B. *Computation of Optimal Transport Plans and Wasserstein Distances*; The Comprehensive R Archive Network, 16 September 2024. Available online: <https://cran.r-project.org/web/packages/transport/transport.pdf> (accessed on 8 November 2024).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.