

UNIVERSITY OF OKLAHOMA  
GRADUATE COLLEGE

ANALYSIS OF DYNAMICS OF ROAD WEATHER INFORMATION SYSTEM DATA

A THESIS  
SUBMITTED TO THE GRADUATE FACULTY  
in partial fulfillment of the requirements for the  
Degree of  
MASTER OF SCIENCE  
in  
Electrical and Computer Engineering

By  
MOHAMED ALI HABIB  
Norman, Oklahoma  
2023

ANALYSIS OF DYNAMICS OF ROAD WEATHER INFORMATION SYSTEM DATA

A THESIS APPROVED FOR THE  
SCHOOL OF ELECTRICAL AND COMPUTER ENGINEERING

BY

---

Dr. Hazem Refai, Chair

---

Dr. Samuel Cheng

---

Dr. Choon Yik Tang

© Copyright by MOHAMED ALI HABIB 2023  
All Rights Reserved.

## **Dedication**

*To my parents who shaped me into the man I am today  
Kadri & Manar*

*To my beloved brother and sister  
Mohamed Nour & Joud*

*And to my dear friends*

*This work is dedicated to you all  
Ali*

## Acknowledgments

First, I would like to extend my gratitude and thanks to my advisor *Dr. Hazem Refai* for his support and guidance. I would also like to state my appreciation to the distinguished members of my committee, *Dr. Samuel Cheng*, and *Dr. Choon Yik Tang* for their time and effort in reviewing this thesis. I would like to extend my gratitude to my colleagues, in addition to staff members including *Denise Davis*, and the *Student Affairs* team. Thank you all for your continuous support and assistance throughout my journey at the *University of Oklahoma*.

Heartfelt thanks and appreciation go to the *OU-Tulsa* family of students, professors, and staff for everything they have provided me with.

# Table of Contents

<b>1</b>	<b>Introduction.....</b>	<b>1</b>
1.1	Thesis Contributions .....	3
1.2	Content Summary.....	4
<b>2</b>	<b>Background .....</b>	<b>5</b>
<b>3</b>	<b>System Components &amp; Data Sources.....</b>	<b>7</b>
3.1	Road and Weather Information System (RWIS).....	7
3.2	HERE .....	9
3.3	Maintenance Activity Truck Tracking (MATT) .....	10
3.4	Road Surface Condition (RSC).....	11
3.5	Summary .....	13
<b>4</b>	<b>Classification of Road Surface Conditions Using Weather and Speed .....</b>	<b>14</b>
4.1	Building a Labeled Dataset .....	15
4.2	Classification of Surface Conditions using Auto labeled Data .....	15
4.3	Classification of Surface Conditions using Ground Truth .....	16
4.4	Summary .....	19
<b>5</b>	<b>Investigating Data-driven Efficiency Measures of Road Clearance and its Effects on Speed During Storms.....</b>	<b>20</b>
5.1	Statistical Temporal Speed Change Measures.....	21
5.2	Efficiency with Sliding Window Recovery Detection .....	24
5.2.1	Efficiency Metric .....	25
5.2.2	Recovery Detection Algorithm .....	26
5.3	Segment-level Speed Efficiency Evaluation .....	29
5.4	Summary .....	32
<b>6</b>	<b>Importance of Subsurface Probes in RWIS .....</b>	<b>33</b>
6.1	Exploratory Data Analysis (EDA).....	34

6.2	Hypothesis Testing .....	36
6.2.1	Correlation Analysis.....	37
6.2.2	Analysis of Variance (ANOVA).....	38
6.3	Regression Analysis .....	39
6.4	Relationship with Clearance Time .....	42
6.5	Machine Learning Model Development .....	44
6.6	Seasonal Variation and Trend Analysis through Parametric Modeling.....	50
6.6.1	Regression Parametric Modeling .....	50
6.6.2	Temporal Segmentation .....	51
6.6.3	Seasonal Trend of the Rate of Change .....	53
6.6.4	Daily Patterns of the Rate of Change.....	59
6.7	Event-Based Analysis of Rate of Change .....	64
6.7.1	Spatial Analysis.....	65
6.8	Summary .....	66
<b>7</b>	<b>Conclusion and Future Work.....</b>	<b>68</b>
7.1	Future Work Remarks .....	69
<b>8</b>	<b>References .....</b>	<b>71</b>

## List of Tables

Table 3-1 ODOT Hazardous Road Surface Conditions .....	12
Table 6-1 Summary of Models, Weather-Event F1 Scores on Minority Classes.....	49

## List of Figures

Figure 3-1 ODOT RWIS stations on I-35 in Oklahoma .....	7
Figure 3-2 RWIS Station Components .....	8
Figure 3-3 HERE Segments in Oklahoma .....	9
Figure 3-4 MATT Illustration .....	10
Figure 3-5 ODOT Field Divisions with County Names .....	11
Figure 3-6 Distribution of Road Surface Conditions Per County.....	12
Figure 3-7 Proportional Distribution of Road Surface Conditions.....	13
Figure 4-1 Illustration for Building a Labeled Dataset.....	15
Figure 4-2 Auto-labeling Policy for Road Surface Conditions Using RWIS .....	16
Figure 4-3 Confusion Matrix on Test Set for Model Trained Using Auto-labeled Data .....	16
Figure 4-4 Daily Number of Data Samples by Surface Condition.....	17
Figure 4-5 Correlation of Speed and Weather with RSC for Various Stations .....	18
Figure 4-6 Results of Models for Station 35ST136.....	18
Figure 4-7 Aggregated Results of Models across Stations .....	19
Figure 5-1 Illustration of Temporal Measures .....	21
Figure 5-2 Aggregated Speed over Time for Trip 49558.....	23
Figure 5-3 Illustration of Sliding Aggregated Speed Variation During Hazardous and Dry Conditions.....	25
Figure 5-4 Illustration of Efficiency Measure using Clearance and Recovery Times .....	26
Figure 5-5 Illustration of Recovery Time Detection.....	26
Figure 5-6 Illustration of Skewness and Kurtosis Roles in Capturing the Distribution Shape.....	27
Figure 5-7 Illustration of Variable Sliding Speed Distribution.....	27

Figure 5-8 Distribution of Efficiency .....	28
Figure 5-9 Change Ratio Per Segment in Speed Average and Variation for Trip 48741 .....	30
Figure 5-10 Observed Weather Condition by Nearby RWIS Station 213 Pre, During, and Post Clearance.....	30
Figure 5-11 Distributions of Change Ratio in Speed Average and Variation .....	31
Figure 6-1 Precipitation Type Table Mappings to Events of RWIS Weather Sensor [10] .....	34
Figure 6-2 Illustration of Temperature Data Over Time and Observed Variation During Storms	35
Figure 6-3 Monthly Distribution of Surface Temperature .....	36
Figure 6-4 Monthly Distribution of 2" Subsurface Temperature.....	36
Figure 6-5 Monthly Distribution of 6" Subsurface Temperature.....	36
Figure 6-6 Correlation Heatmap of Ambient, Surface, and Subsurface Temperatures.....	38
Figure 6-7 Distribution of Temperature per Surface Condition at 2" and 6" levels .....	38
Figure 6-8 Subsurface Temperature Distribution at 2" and 6" for Various Weather Events.....	38
Figure 6-9 Surface (left) and Ambient (right) Temperature Distribution for Various Weather Events .....	39
Figure 6-10 Regression Slope per RSC for Daytime and Nighttime.....	41
Figure 6-11 Regression Slope per Weather Event for Daytime and Nighttime .....	42
Figure 6-12 Distribution of Clearance Times .....	44
Figure 6-13 Distribution of Subsurface Temperature over Clearance Time Before and During Clearance.....	44
Figure 6-14 Regression Slope for Clearance Time over Mean Subsurface Temperature, Before and During Clearance .....	44
Figure 6-15 Illustration of the Machine Learning Models Experiment.....	45

Figure 6-16 RSC Class Distribution for Training and Testing Datasets .....	47
Figure 6-17 Model 1, RSC Results: Feature Importance, Confusion Matrix, PR Curve .....	47
Figure 6-18 Model 2, RSC Results: Feature Importance, Confusion Matrix, PR Curve .....	47
Figure 6-19 Model 3, RSC Results: Feature Importance, Confusion Matrix, PR Curve .....	47
Figure 6-20 Weather-Event Class Distribution for Training and Testing Datasets.....	48
Figure 6-21 Model 1, Weather-Event Results: Feature Importance, Confusion Matrix, Per class: Precision, Recall, F1 Score .....	48
Figure 6-22 Model 2, Weather-Event Results: Feature Importance, Confusion Matrix, Per class: Precision, Recall, F1 Score .....	49
Figure 6-23 Model 3, Weather-Event Results: Feature Importance, Confusion Matrix, Per class: Precision, Recall, F1 Score .....	49
Figure 6-24 Signal Dynamics of Ambient and Probes Temperature over Time .....	51
Figure 6-25 Increasing and Decreasing Temperature Patterns for Daytime and Nighttime.....	52
Figure 6-26 Regression Lines for 2" and Ambient: Unsegmented, Nighttime, and Daytime .....	52
Figure 6-27 Monthly Regression Lines for 2" and Ambient: Nighttime, and Daytime.....	53
Figure 6-28 Monthly Rate of Change (Slope) for Daytime and Nighttime .....	55
Figure 6-29 Monthly Temperature Distribution for Daytime and Nighttime, for Ambient, 2", and 6" Subsurface Probes .....	55
Figure 6-30 Monthly Behavioral Pattern of Rate of Change for Day/Night for Ambient and 2" Subsurface Probe for Various Stations.....	56
Figure 6-31 Monthly Behavioral Pattern of Rate of Change for Day/Night for Ambient and 6" Subsurface Probe for Various Stations.....	57
Figure 6-32 Monthly Rate of Change for Day/Night for 213 and 136 .....	58

Figure 6-33 Monthly Correlation Coefficient for Day/Night for 213 and 136.....	58
Figure 6-34 Daily Rate of Change for Day/Night, for both Subsurface Probes for Various Stations .....	59
Figure 6-35 Distribution of Daily Rate of Change for Day/Night for 2" and 6" Subsurface Probes .....	60
Figure 6-36 Daily Rate of Change for Day/Night for 2" and 6" Subsurface Probes .....	60
Figure 6-37 Monthly Distribution of Rate of Change for Day/Night for 2" Subsurface Probe ...	61
Figure 6-38 Monthly Distribution of Rate of Change for Day/Night for 6" Subsurface Probe ...	61
Figure 6-39 Spatial Distribution of Rate of Change for Day/Night for 2" Subsurface Probe .....	62
Figure 6-40 Spatial Distribution of Rate of Change for Day/Night for 6" Subsurface Probe .....	62
Figure 6-41 Wavelet Scalogram for Night/Day for 2", and 6" Subsurface Probes.....	64
Figure 6-42 Temperature Progression Through Storms (shaded).....	64
Figure 6-43 Slope Distribution for Event (Storm/Non-Storm): Unsegmented (left), and Segmented (right) .....	65
Figure 6-44 Slope Distribution per Station for 2" (left) and 6" (right).....	65
Figure 6-45 Segmented Slope Distribution per Station for 2" Probe .....	66
Figure 6-46 Segmented Slope Distribution per Station for 6" Probe .....	66

## **Abstract**

Road and Weather Information System consists of a network of roadside Environmental Sensor Stations (EES) collecting meteorological data. Equipped with a variety of sensors, these stations gather data including ambient temperature, subsurface temperature, precipitation level and type, brightness, and more. Consequently, RWIS systems have been critical for increasing road safety over the years by providing valuable weather information helpful in anticipating and preparing for adverse weather conditions and reducing traffic collisions [1]. The advancement of Artificial Intelligence (AI) and Machine Learning (ML) methodologies enhances our ability to leverage these systems through the analysis of temporal datasets, unraveling dynamic behavioral patterns over time, and constructing more accurate and dependable predictive models.

The research study reported in this thesis focuses on understanding the dynamics of weather data to enhance road safety and the utilization of RWIS, especially during winter exhibiting hazardous road surface conditions. Utilizing ambient and subsurface temperature data from RWIS stations along Interstate 35, managed by the Oklahoma Department of Transportation (ODOT), the study methodically investigates the climatological patterns across different times of day and seasons. Parametric regression modeling is conducted to characterize the behavioral patterns of weather data for diurnal and nocturnal periods through seasonal progression. A notable aspect of this research is the demonstration of how incorporating data from subsurface temperature probes enhances the performance of machine learning models for the classification of road surface conditions and weather events. The study explored the influence of road clearance time on traffic speed during snowstorms employing statistical data-driven techniques of change in speed pre- and post-clearance. Machine learning classification techniques are employed to automate the detection of hazardous road surface conditions like ice and snow based on weather and speed data.

# 1 Introduction

The need for monitoring road weather conditions amidst adverse winter storms through Road and Weather Information (RWIS) roadside stations is becoming increasingly important. In 2013 [2], the National Weather Service (NWS) reported a yearly estimate in the United States of 580,000 crashes due to snowy, icy, and slushy roads. These crashes, unfortunately, cause around 180,000 injuries, and 2,100 fatalities. To account for these conditions, states allocate an estimated \$2.3 billion annually for winter road maintenance efforts aimed at snow and ice management. Moreover, the Federal Highway Administration (FHWA) in [15], based on National Highway Traffic Safety Administration (NHTSA) data from 2007 to 2016, estimated that 1,235,000 yearly crashes are due to adverse weather conditions – 21% of total annual crashes. Crashes that each year result in approximately 5,000 deaths and 418,000 injuries. Reflecting on this, it's clear that hazardous weather conditions on roads not only lead to significant accidents and economic expenditures but also disrupt and decelerate the flow of traffic.

Road weather conditions are monitored by Road and Weather Information System (RWIS) stations deployed on the I-35 highway in Oklahoma. Collecting various weather data which can be utilized in a variety of applications, from analysis to model relationships and enhancing our understanding of certain behavioral trends to building prediction and forecasting models. Such advancements and research studies can help improve our utilization of existing RWIS systems by highlighting the importance of some components (i.e., subsurface probes) and their direct or indirect impact on various weather conditions and further optimizing different prediction models.

Existing research pursuing to enhance safety on roads during adverse conditions tends to mostly focus on the development and improvement of forecasting models for surface temperature. For instance, adopted surface (or pavement) temperature forecasting models like METRo,

SNTHERM, and FASST, use a mixture of weather attributes representing atmospheric weather changes and heat induction to forecast surface temperature ahead of time. These attributes include temperature and precipitation levels from RWIS stations. Forecasting the temperature is highly important for Departments of Transportation (DOTs) across the nation to prepare for any hazardous conditions during cold winters, especially on the roads. However, such predictive models face challenges including maintaining robustness against diurnal trend variations (i.e., relationship during the day is different than at night), seasonal patterns, and spatial and climatic changes. Therefore, enhancing our utilization of RWIS systems is critical for further advancements in such prediction and forecasting models.

In this study, I conducted analytical experiments and developed predictive machine learning models to emphasize the importance of subsurface temperature sensors (i.e., probes) in RWIS measured at two levels – 2” and 6”. Experiments included regression-based parametric modeling to characterize the behavioral patterns and trends between subsurface temperature and ambient temperature concerning seasonal progression, especially during winter months. A relationship that is worth investigating since temporal variations in ambient, surface, and subsurface temperature lead to hazardous weather on the roads as they can, for example, cause the formation of ice pallets. Daytime and nighttime patterns were examined distinctly due to their contrasting characteristics, with temperatures typically ascending by day and descending at night, leading to divergent behaviors associated with these cycles. The variation in the rate of change in the relationship was studied across different road surface conditions and weather events. Hypothesis testing was implemented to study the changes across these conditions utilizing correlation coefficients (i.e., Pearson’s Correlation), analysis of variance (ANOVA), and regression analysis. We also characterized the behavior between subsurface temperature and road clearance time by snowplow

trucks which was seen to increase as temperatures decrease. Road clearance data was processed and analyzed before, during, and after clearance to measure the efficiency and impact of road clearance on traffic speed. Furthermore, machine learning models were developed to predict road weather conditions and weather events. Lastly, incorporating subsurface temperature in prediction models has shown an experimental improvement in performance. Data analyzed in the research experiments throughout this study come from multiple systems currently in use by the Oklahoma Department of Transportation (ODOT) including weather data from RWIS stations, clearance time data from snowplow trucks through Maintenance and Truck Tracking (MATT), and traffic speed data from HERE.

## **1.1 Thesis Contributions**

This thesis was developed with the objective of modeling the relationship between subsurface and ambient temperature to understand the behavioral patterns as the colder weather progresses. The relationship was studied over the course of 27 months spanning from January 2021 up to March 2023. The rate of change between ambient and subsurface temperatures, at the 2” and 6” levels, was modeled by the slope of a linear regression model – for interpretability and simplicity. The methods encompassed statistical and regression analysis as well as machine learning model development. The research detailed in this study aims to provide findings that can help enhance overall road safety during adverse weather events and prompt the effective usage and adoption of subsurface probes into RWIS.

The main contributions of this thesis are summarized below:

- [1] Characterize the behavioral relationship patterns between ambient and subsurface temperatures at 2” and 6” inches over seasonal progression for daytime and nighttime across different RWIS stations set in diverse climatic zones.

- [2] Investigate the relationship changes for various road weather conditions – provided by ODOT and weather events – captured by RWIS sensor.
- [3] Develop machine learning models to showcase the improvement of incorporating probes in decision-making to classify road weather conditions and weather events.
- [4] Develop machine learning models to classify road weather conditions leveraging weather data – capturing atmospheric fluctuations, and traffic speed data – capturing local traffic variations.
- [5] Investigate statistical data-driven measures of efficiency manifested by the impact of clearance time on traffic speed and its variation.

## **1.2 Content Summary**

In alignment with the objectives previously outlined, this thesis is structured into several segments. Section 1 offers a succinct overview of the objectives, research, and analyses conducted within this thesis. Section 2 summarizes the background and details some of the related work. Section 2 outlines the systems providing the data utilized in this study. Section 3 describes the development of machine learning classification models that utilize weather and traffic speed data to identify road weather conditions – attempting to automate the labeling of road weather conditions. Section 4 demonstrates the experiments aiming to investigate efficiency measures of road clearance efforts and study their impact on traffic speed by analyzing data before, during, and after clearance. Section 5 showcases a set of experiments highlighting the importance of subsurface probes in RWIS and modeling the temporal parametric relationship through seasonal progression. Finally, Section 6 provides a summary of the work accomplished in this thesis and proposes directions for potential future research venues.

## 2 Background

It has been historically proven that adverse road weather conditions are directly related to accidents related to factors such as rainfall, mean temperature, and more importantly, snowfall – which has a more significant relationship with accidents [5]. Over the years, academic researchers and transportation agencies have been working on developing a road infrastructure equipped with RWIS stations and prediction models to mitigate these dangerous effects. For example, the California Department of Transportation installed an icy curve warning system (ICWS) which has been shown to reduce crashes by 18% on a highway known for numerous accidents [7]. Therefore, improving our resource exploitation of road weather systems can be pivotal to enhanced safety. In the pursuit of improved road safety, we attempt to underscore the significance and influence of subsurface temperature probes within RWIS for predictive accuracy. In addition, we characterize subsurface temperature with ambient over time, a focal point of investigation in this research thesis.

The authors in [4], developed a statistical approach using a stepwise linear regression model to forecast road surface temperature using various inputs including subsurface temperature at 30 cm (~12 inches). A modest positive correlation between the model's surface and ambient prediction errors by the model was noted as a constraint. The authors in [3] introduced a model that integrates RWIS weather data with roadside camera images to classify road surface conditions into dry, wet, snowy, and icy. The method included understanding feature correlations through Principal Component Analysis (PCA) and then implementing a regression model with a reported accuracy that ranges between 91% and 100% - around a 10% error margin. The authors in [8] evaluated three machine learning models in the classification of road surface conditions during the 2014-

2015 winter period. The researchers suggested a solution to advance RWIS systems by incorporating imagery data collected by a phone-based system deployed in vehicles. These images are integrated with RWIS weather data, including surface temperature, to classify surface conditions – an integration resulting in an 18% average increase in accuracy. Among the tested algorithms – Random Forest (RF), Random Trees (RT), and Artificial Neural Network (ANN). Tree-based methods consistently surpassed neural networks in accuracy –with RT achieving the highest accuracy rate at 78%. The authors in [9], presented a compact alternative to conventional RWIS for tracking road conditions. The researchers identified air and surface temperature as key predictors and employed non-contact sensors for data acquisition, essentially creating a miniaturized version of an RWIS station. They utilized the established METRo model for prediction with an overall Root Mean Squared Error (RMSE) of 5.47 °C, noting an observed variation with an RMSE of 7.28 °C and 1.15 °C for daytime and nighttime, respectively.

In brief, there is a critical need for expanded research to understand the interplay between adverse weather conditions and traffic flow, which is essential to advance the reliability of road weather prediction tools, such as Road and Weather Information Systems (RWIS), and improve the robustness and efficacy of Intelligent Transportation Systems (ITS) infrastructures [6].

### 3 System Components & Data Sources

In this section, we provide an overview of data systems used in the research studies of this thesis, setting the stage for the motivations and methodologies discussed in subsequent chapters by familiarizing the reader with the nature of the data and the corresponding systems in place.

#### 3.1 Road and Weather Information System (RWIS)

The systems include RWIS (Road and Weather Information System) – providing real-time weather data being collected through a variety of weather sensors and stored on the data servers made available for viewing or downloading through the corresponding website. [Figure 3-1] displays the fifteen weather stations deployed by the Oklahoma Department of Transportation (ODOT) on I-35 along with the corresponding counties in which they are located. The stations cover I-35 from the Texas to the Kansas border with over 200 miles of interstate road. Stations are typically situated 15 miles away from each other. This wide range provides a spectrum of climates suitable for exploration and analysis.

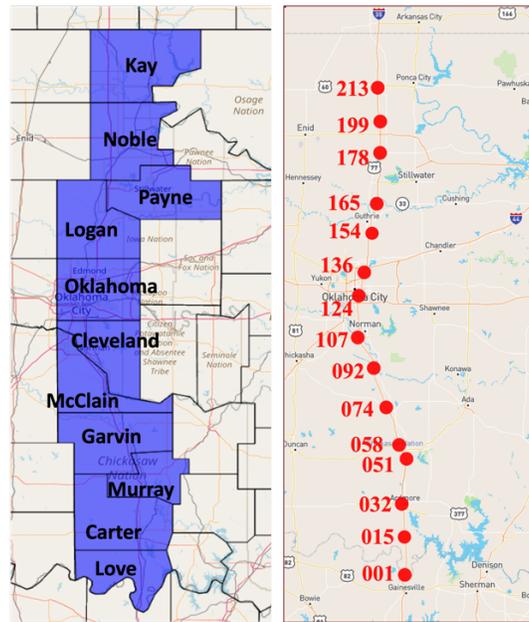
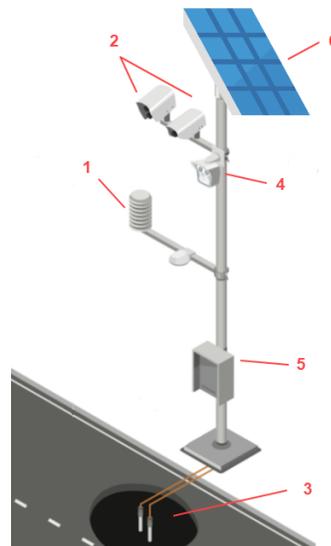


Figure 3-1 ODOT RWIS stations on I-35 in Oklahoma

Each RWIS station is equipped with multiple weather sensors collecting valuable data. [Figure 3-2] provides a 3D illustration of a weather station. Where (1) is the weather station sensor [10] collecting various types of weather data including precipitation levels and types, temperature, brightness, etc. The station is equipped, in (2), an infra-red surface temperature sensor [11] collecting surface and bridge temperatures. Subsurface temperature probes [12], in (3), are collecting underground temperature data at 2” and 6”. Each station has a camera, in (4), installed with a view of the road. The Internet of Things (IoT) cabinet represents the brain of the stations where the power source, the modem, and the embedded system are located managing data collection and communication to the servers. Finally, shown in (6), is the solar panel utilizing sun radiation to recharge the batteries in the IoT cabinet.



*Figure 3-2 RWIS Station Components*

We use RWIS data collected per station every 5 minutes to build a data table used to conduct investigative statistical and exploratory data analysis as well as building predictive machine

learning models in addition to parametric models that aid in understanding the temporal relationship of weather data.

### 3.2 HERE

Traffic speed data is provided by a GPS probe-based system, which collects data from in-vehicle transponders, navigation systems, and mobile devices. The data is provided on a segment level where, for example, each highway or road can consist of many segments traveling per direction. Speed is provided as an average representing the average speed from all the vehicles traveling through the designated segment at some point in time. The figure shows HERE segments in Oklahoma, each point corresponds to the starting point of the segment. Segments cover interstate, state, and US highways. As shown, in [Figure 3-3], most of the segments are condensed near the Oklahoma City and Tulsa metropolitan areas.

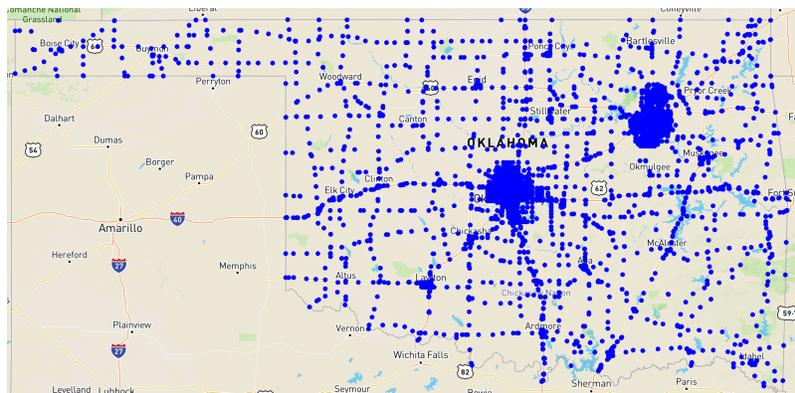


Figure 3-3 HERE Segments in Oklahoma

Traffic speed data is used to collect speed data near a weather station representing the traffic conditions. Speed data is either aggregated across segments or analyzed as a distribution. We analyzed speed conditions during storms and inspected them at different clearance times.

### 3.3 Maintenance Activity Truck Tracking (MATT)

The Maintenance Activity Truck Tracking (MATT) system consists of a mobile application, web application, and embedded system collecting clearance data from snowplow trucks. Data include data collected from the sand deployment system through the embedded PCB, through the mobile application, and to a website where it is stored on a database available for viewing and downloading for analysis. The application facilitates communication between drivers and operators through text, audio, images, and video. The data is temporal and geospatial being collected per trip and truck. The figure provides a visual showing the components of MATT, where data collection, including GPS and embedded system, starts at the truck level. Then it is communicated through the embedded system deployed in the truck to the mobile application, in (2), which is in turn responsible for communicating the data to the MATT admin panel website where data is available for download and viewing, in (3) where we show a sample trip data with GPS latitudes and longitude with timestamps.

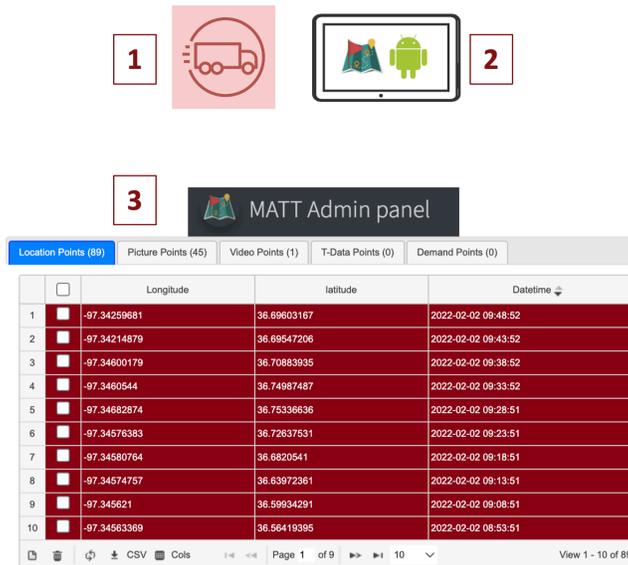


Figure 3-4 MATT Illustration

We use MATT data to analyze the clearance time of trips happening during snowstorms in an attempt to derive and quantify the efficiency of clearance on traffic speed. In other words, creating statistical indicators of the impact of clearance time on traffic speed.

### 3.4 Road Surface Condition (RSC)

Road Surface Condition (RSC) data is collected by ODOT operators through manual visual inspection of road weather conditions during storms to classify the type of hazard the road has. Surface condition data is collected per ODOT division or district, divisions are shown in [Figure 3-5]. In each division, the data is organized per county and the corresponding control sections. The raw data table includes starting and ending timestamps, condition type, control section, highway (i.e., I-35), and district of each hazardous event.

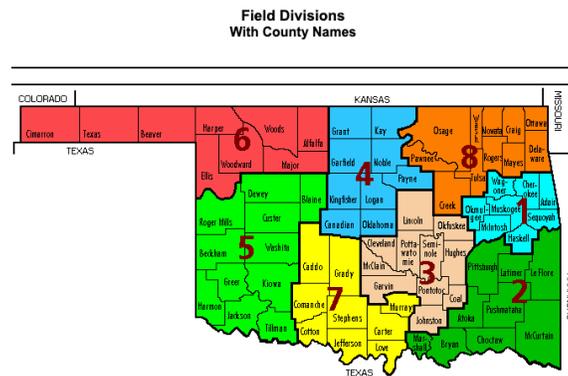


Figure 3-5 ODOT Field Divisions with County Names

The following table details the different hazardous weather surface conditions recorded in the dataset. As shown, the conditions consist of wet, snowy, and icy conditions.

Road Surface Conditions (RSC)
Slick Spots Highway
Slick Hazardous Highway
Light Snow
Moderate Snow
Heavy Snow
Snow Packed
Blowing Snow
Ice

Table 3-1 ODOT Hazardous Road Surface Conditions

[Figure 3-6] shows the distribution of surface conditions observed through the counties from North to South in 2021 (right) and 2022 (left). We notice that most snowy and overall extreme conditions are concentrated more in the northern counties – which is expected.

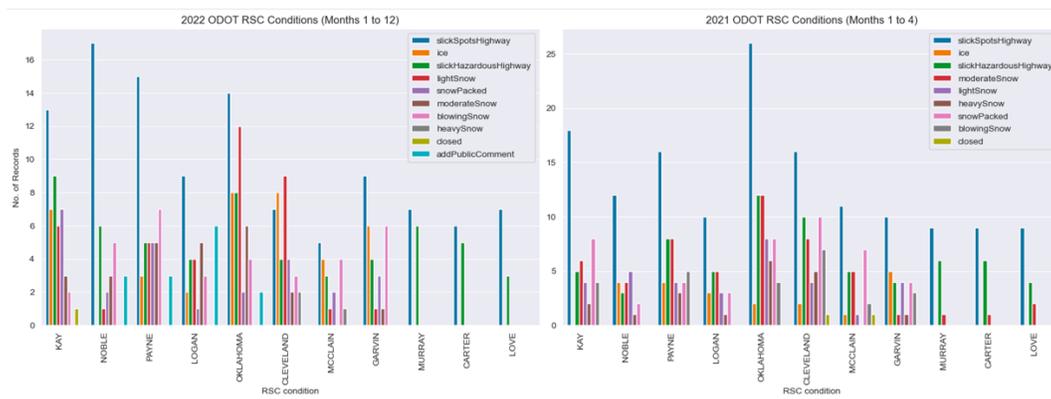


Figure 3-6 Distribution of Road Surface Conditions Per County

The next figure displays, as a pie chart, the proportions of the observed conditions for 2021 (right) and 2022 (left) on I-35 in Oklahoma.

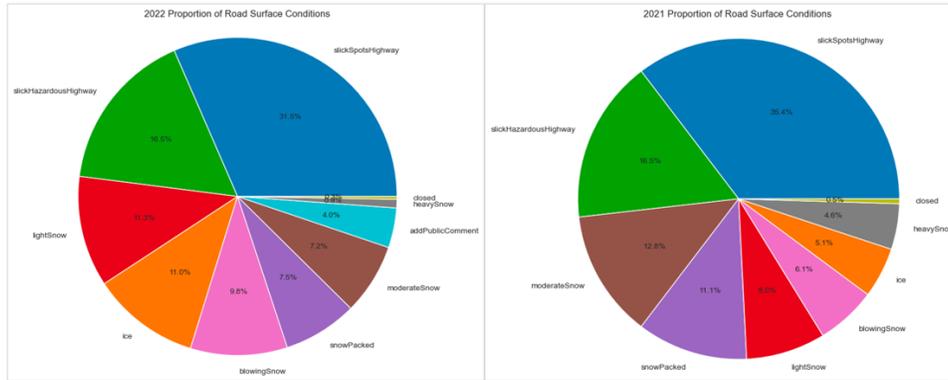


Figure 3-7 Proportional Distribution of Road Surface Conditions

Road surface conditions data is used to label weather data tables to identify and analyze data during storms and the various types.

### 3.5 Summary

In conclusion, the data sources outlined in this chapter present a rich and complex tapestry of geospatial and temporal datasets that are integral to the research presented in this thesis. The RWIS stations provide a comprehensive array of weather-related data, crucial for understanding and predicting road conditions. The IoT infrastructure underlying these systems allows for real-time, continuous data collection, which is pivotal in building reliable predictive models and conducting thorough exploratory analyses. Traffic speed data adds another layer of complexity, offering insights into speed patterns that are affected by weather conditions. MATT data give a detailed account of maintenance operations, providing valuable information on the efficiency of road clearance during snowstorms. Finally, the RSC data, although manually collected, offer a ground truth for validating the predictions made by models and offer a categorical perspective on the impacts of various weather events on road conditions. Together, these datasets not only demonstrate the intricacies involved in managing and analyzing large-scale weather and traffic data but also underscore the potential for sophisticated statistical and machine-learning tools to yield actionable insights for improving road safety and maintenance efficiency.

## **4 Classification of Road Surface Conditions Using Weather and Speed**

In this section, we will present weather road surface conditions machine learning classification models trained to identify hazardous conditions. Constructing a system that proficiently categorizes road weather conditions offers significant advantages. It allows transportation agencies to label surface conditions over a broader area in real time using RWIS stations. This not only enhances the efficiency of the task but also conserves operators' time, eliminating the need for on-site road inspections during storms to determine surface conditions. The experiments were conducted on data collected from 6 RWIS stations in the northern part of I-35 – as these stations tend to have more frequent hazardous conditions due to their colder climate. Stations include, in order from north to south, 35ST213, 35ST199, 35ST165, 35ST154, 35ST136, and 35ST124. A dataset, per station, was built consisting of weather features – gathered from the RWIS station including data describing precipitation, wind, ambient and underground temperature, speed features – gathered from nearby segments, and road surface conditions (RSC) labels – collected manually by visual inspection by ODOT. The analysis covered winter data including January and February of 2022. We will showcase findings from two distinct modeling experiments. In the initial experiment, we developed a model that classifies surface conditions into dry, wet, and snowy categories. This classification was achieved without utilizing ODOT RSC labels. Instead, we constructed labeled weather data based on a policy that considered surface temperature, total precipitation, precipitation type, humidity, and wind speed. In the subsequent experiment, we employed ODOT RSC labels, which are considered the ground truth, to create a dataset that classifies road surface conditions into hazardous and dry categories.

## 4.1 Building a Labeled Dataset

Let's start with the first experiment, the next figure provides a visual explanation of building the training dataset. There are several RWIS stations deployed on I-35, each of which provides various weather data. We combine weather data with speed data from two probe-based systems, namely INRIX and HERE, from a group of segments within 4 miles of a station – assuming the same weather conditions within this region.

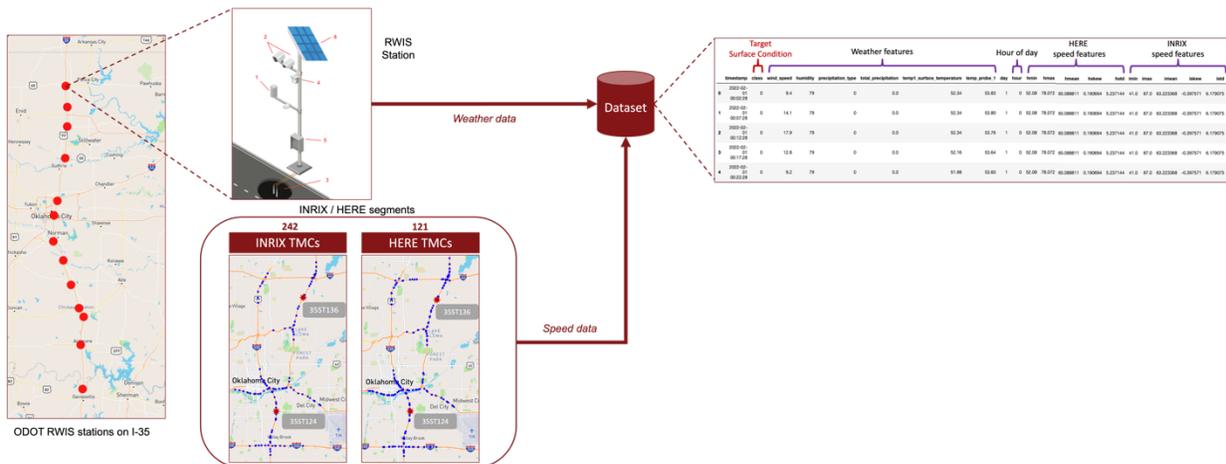


Figure 4-1 Illustration for Building a Labeled Dataset

## 4.2 Classification of Surface Conditions using Auto labeled Data

The following [Figure 4-2] shows the used labeling policy to label the built dataset with surface condition labels (i.e., wet, icy, wet, or dry). This aims to build a self-contained system that is independent of human input (i.e., RSC labels) and only uses sensor data to classify surface conditions. The policy labels the data as follows: When the surface temperature resides below the freezing point, several factors are considered. If there is evident precipitation, the humidity level surpasses 60%, the wind speed measures above 3 m/s – wind plays a factor in forming ice [18], and the type of precipitation is identified as snow, the surface receives an 'icy' label. Conversely, if the precipitation is present, the humidity exceeds 50%, and the type is categorized as rain, the surface is denoted as 'wet'. In scenarios that don't fit either criterion, the surface is labeled 'dry'.

For cases where the surface temperature is above the freezing threshold, the label is determined as 'wet' if there's active precipitation and humidity levels extend beyond 50%. In all other situations, the surface is characterized as 'dry'.

**Policy:**

- Surface Temperature  $\leq 32$ , then:
  - Icy if:
    - Total prec  $> 0$  (mm/d) and prec type  $\geq 70$  and humidity  $\geq 60$  and wind speed  $\geq 3$  (m/s)
  - Wet if:
    - Total prec  $> 0$  (mm/d) and prec type  $> 0$  and humidity  $\geq 50$
  - Dry OW.
- Surface Temperature  $> 32$ , then:
  - Wet if:
    - Total prec  $> 0$  (mm/d) and prec type  $> 0$  and humidity  $\geq 50$
  - Dry OW.

**Precipitation Type Table**

Synop Code w.w.	Meaning
0	No precipitation.
40	Precipitation present.
51	Light drizzle.
52	Moderate drizzle.
53	Heavy drizzle.
61	Light rain.
62	Moderate rain.
63	Heavy rain.
67	Light rain and/or drizzle with snow.
68	Moderate rain and/or drizzle with snow.
70	Snowfall.
71	Light snow.
72	Moderate snow.
73	Heavy snow.
74	Ice pellets.
89	Heavy hail.

Table 1: Synop Code Table

Figure 4-2 Auto-labeling Policy for Road Surface Conditions Using RWIS

The following figure presents the resulting normalized confusion matrix on the test set for the trained model, Naïve Bayes. The model was trained using an 80/20 test split, features were scaled in the range [-1,1] before fitting. We noticed that the model had high accuracy for dry and wet conditions with moderate accuracy for icy conditions – as some examples are misclassified as wet. The confusion between wet and icy in the model is explained by the similarity in the distributions for these two classes, plus, they also could mean a limitation in the labeling policy – as it does not provide ground truth and is only a proxy for guessing surface conditions.

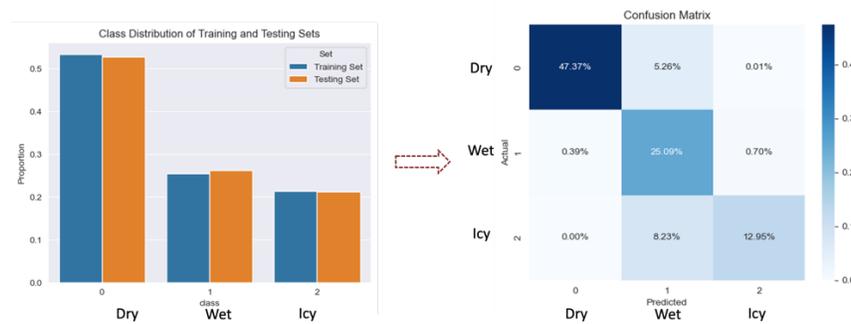


Figure 4-3 Confusion Matrix on Test Set for Model Trained Using Auto-labeled Data

### 4.3 Classification of Surface Conditions using Ground Truth

Moving on to the next experiment, classifiers were trained to distinguish between surface conditions using ground truth RSC labels. These models utilized, as features, weather data, HERE

speed data, and labels, the ODOT road weather surface conditions. [Figure 4-4] shows the daily number of data samples per surface condition, that are either hazardous or dry. We notice that most hazardous conditions happened in early and late February while some in early January – this pattern applies to the other stations.

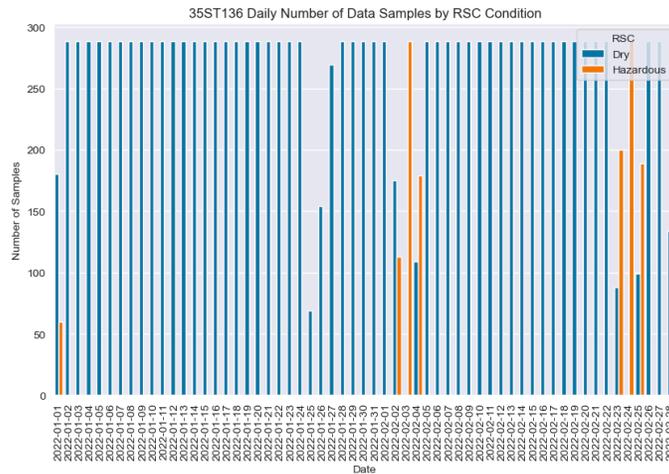


Figure 4-4 Daily Number of Data Samples by Surface Condition

The following figure displays the Pearson’s correlation coefficient for speed and weather features with road surface conditions – where 1 corresponds to hazardous and 0 to dry, per station. We notice a strong negative correlation with the speed mean, ranging from -0.68 to -0.80, which can be interpreted as follows: as the mean speed of the nearby segments increases, conditions tend to be dry. Whereas, when it decreases, that is when hazardous conditions arise.

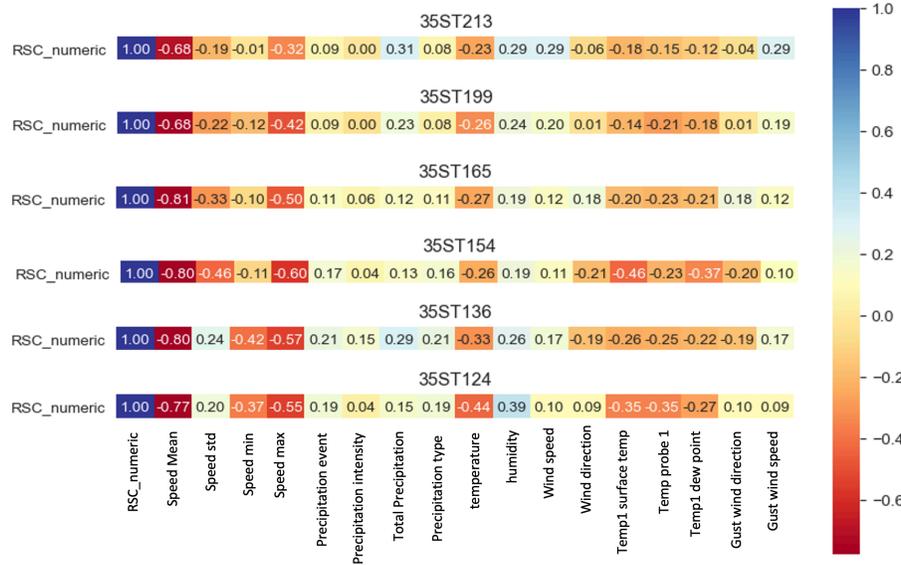


Figure 4-5 Correlation of Speed and Weather with RSC for Various Stations

The following graph presents the confusion matrix on the test set of three models trained to classify road surface conditions near station 136 in Oklahoma County. The class distribution, of road surface conditions, was approximately 92% for dry and 8% for hazardous. The models can almost perfectly the dry conditions while missing a minor group of the icy conditions. The misclassification error ranges up to approximately 10% - where 0.67% of hazardous conditions were misclassified as dry by the logistic regression while 7.64 were correctly classified as hazardous; since 0.67 by 7.64 is around 8.7% error.

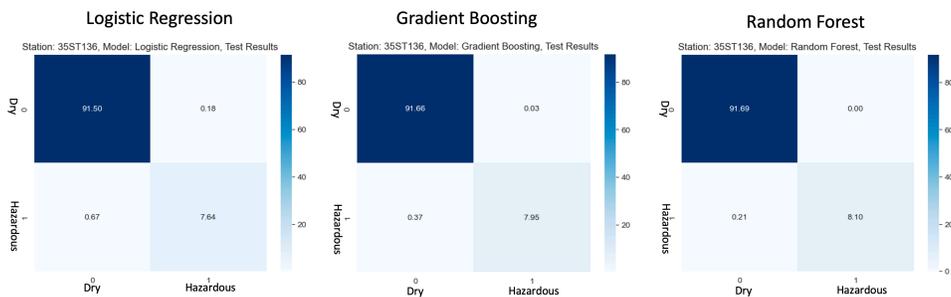


Figure 4-6 Results of Models for Station 35ST136

Next, [Figure 4-7] shows the aggregated results across stations 213, 199, 165, 154, 136, and 124.

A confusion matrix was calculated per station and model, and then matrices were averaged across

stations per model. We notice the same pattern, as in the results for station 136 in [Figure 4-6], some icy conditions were incorrectly classified as dry, corresponding to 0.78 out of 5.95 which is around 13% error for logistic regression as an example.

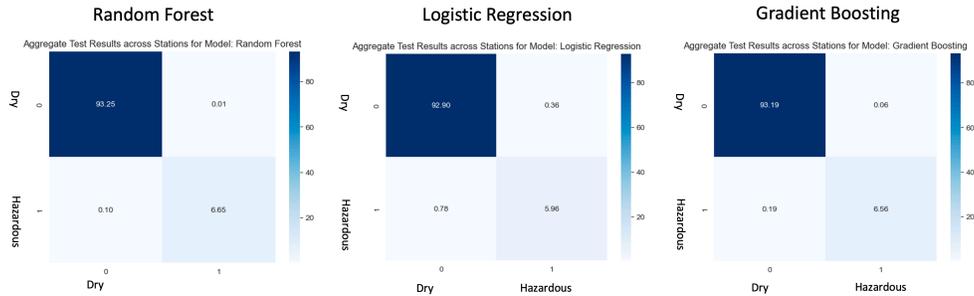


Figure 4-7 Aggregated Results of Models across Stations

## 4.4 Summary

In conclusion, this chapter delved into the development and experimentation of machine learning models aimed at classifying road surface conditions. The models provided predictions by leveraging weather and speed data – which makes our approach more comprehensive. Our experimentation included leveraging "ground truth" labels from ODOT, obtained during storm road inspections, and an independent approach that solely depended on RWIS weather sensor data to approximate surface conditions. The rationale behind creating a model not contingent on ground truth labels was twofold: initially, this was a preliminary experiment conducted before labels became available. Furthermore, obtaining these labels can be resource-intensive, given that it requires operators to traverse highways and manually record surface conditions. Therefore, an automated system or model that can approximate these conditions using RWIS data could offer significant benefits. Notably, our models demonstrated a commendable ability to correctly identify most of the hazardous conditions, albeit with a margin error of 13%. With these insights, we hope to pave the way for further advancements and experimentation in road safety measures and innovative solutions in the domain.

## **5 Investigating Data-driven Efficiency Measures of Road Clearance and its Effects on Speed During Storms**

In this section, we examine the relationship between road clearance time by The Oklahoma Department of Transportation (ODOT) snowplow trucks during snowstorms and its effect on traffic speed. The experiments discussed in this chapter achieve this by exploring temporal statistical indicators that carefully inspect speed progression by comparing traffic speed status before, during, and after the clearance ends. There seems to be no existing literature exploring efficiency or the relationship between clearance time and speed, which motivated us to carry out the analysis detailed in this chapter. However, the lack of previous research posed challenges, especially in the evaluation of the developed efficiency metrics. Additionally, unlike a physical, mechanical, or power system where calculating efficiency is straightforward by taking the ratio of the useful output to the input, coming up with efficiency metrics characterizing the impact of road clearance time on traffic speed is not as easy as we do not have a true label describing the actual efficiency. Therefore, our metrics relied on intuition and statistical analysis to derive various ratios that may be considered efficiency indicators (e.g., average speed improvement ratio before and after clearance). Regardless of the above challenges, our curiosity to attempt to quantify efficiency has made it possible to work through them.

To preface the experimental discussions, we summarize the data utilized: we used geospatial time series data collected from MATT deployed in ODOT snowplow trucks, capturing timestamps and GPS coordinates. Additionally, we employed HERE traffic speed data, which aggregates vehicular speeds over highway segments every 5 minutes. Lastly, hazardous road surface conditions by ODOT, complement our dataset to investigate impacts of clearance on speed.

We initiated experiments by gathering clearance data for January and February of 2022, a period that witnessed hazardous conditions, for division four with a total of 419 trips. Note that not all trips are for clearance or sand deployment, some of the trips can be for construction purposes. We selected trips within a certain vicinity (i.e., 1 mile) from weather stations that have road surface condition data as well as speed data for the segments on the road. We used speed data from multiple segments near a station to extract aggregate statistics describing the distribution at that moment or for the past temporal window (e.g., mean, standard deviation, min, max, kurtosis, skewness) and labeled weather data with it. However, we also used speed data from all these segments without only relying on these aggregates which will be described later.

## 5.1 Statistical Temporal Speed Change Measures

To begin, we defined five temporal speed efficiency measures that we used to compare system status (speed distribution) for different periods including before and after clearance. [Figure 5-1] provides a visual explanation of this.



*Figure 5-1 Illustration of Temporal Measures*

First, the average speed improvement ratio is calculated by taking the difference between the average speed for the periods preceding and following clearance and then dividing it by the average speed for the preceding period, in [5-1]. To calculate the average speed for the preceding period, we identify the trip start time and subtract the window size (e.g., 1h, 5h, etc.) to mark the boundaries of this period. Then we filter the data table, within this period, with labeled speed data which consists of speed aggregates from all the segments, and use it to calculate the average speed.

The idea is that if a storm happened and the road was cleared, we should notice some improvement in the overall average speed across all the segments.

$$\frac{\mu_{after} - \mu_{before}}{\mu_{before}} \quad 5-1$$

Second, the change ratio in speed variability is similarly calculated for before and after clearance periods by the aggregated standard deviation difference, as shown in [5-2]. The intuition is that if less variability is observed after clearance, this should be an indicator of a higher efficiency.

$$\frac{\sigma_{after} - \sigma_{before}}{\sigma_{before}} \quad 5-2$$

Third, the ratio of drops of speed below a threshold was calculated like previously for both periods. The number of data samples where speed is below a certain threshold in the before period over the overall number of samples within that period will yield a percentage representing the drops during this time window. Then this percentage is also calculated for the after period, resulting in another ratio. Next, taking the difference of these two values over the ratio for the before period will give us the overall change in this ratio from the period before clearance to the period after. The idea behind this is that a clear road will more likely have free-flow traffic which will lead to a lower percentage of values below the threshold and thus higher efficiency.

Next, the improvement ratio in peak speed is calculated by taking the difference between the maximum speed from both periods and then dividing it by the maximum speed from the before period. Higher peaks, after clearance, will presumably mean higher efficiency.

Lastly, the average time for the speed, after clearance, to converge the mean speed in equivalent dry (non-stormy) conditions. Once the speed converges after clearance, the time difference between that moment in time and the trip end is the time it took the speed conditions to recover after that trip. Subsequently, lower average time will assumably correspond to higher efficiency.

For example, let's take the trip 49558, shown in [Figure 5-1], as an example. The figure shows various statistics of the aggregated speed from the segments over time as well as the trip start and end times. We also show 5h before and after the trip representing the periods preceding and following clearance. The blue-shaded area represents hazardous road conditions showing that this trip happened amid a storm. The clearance time for this trip was 2h 20m 7s. The blue line shows the speed mean, and we notice a 13.03% improvement in the speed average. The standard deviation, representing speed variability, has shown a 49% reduction after clearance. There was a reduction of 50% in the drops of speed mean, blue line, below the threshold (i.e., 57 MPH). The threshold was chosen experimentally as the mean speed during hazardous conditions which is 57 MPH. However, the speed peak was reduced by 4% after clearance and did not improve (increase). In addition, the mean speed did not recover within the 5h window after clearance to the equivalent speed mean during dry conditions (i.e., 73 MPH). Clearly, in this example, the storm spanned almost two days and despite some improvement across some metrics, other metrics showed reduction instead of improvement.

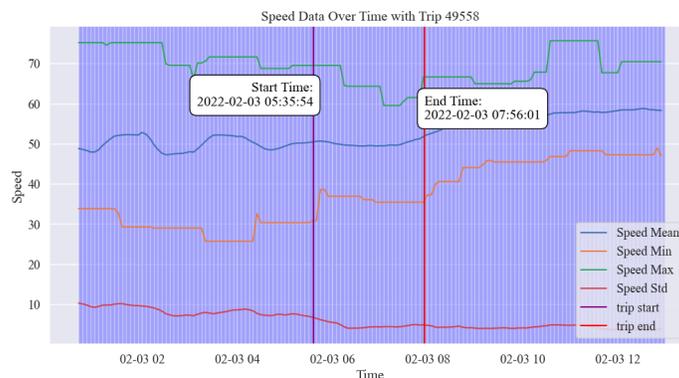


Figure 5-2 Aggregated Speed over Time for Trip 49558

However, after applying these metrics across the trips we noticed the limitations. These metrics were calculated per trip and then averaged across all 64 chosen trips. There was almost no tangible improvement in the first four metrics after clearance. However, looking at the last metric, the time

it took speed to recover dry speed conditions, we noticed that around 38% of trips recovered with an average recovery time of 50m while the others did not recover within this 5h window. In retrospect, we notice that these statistical metrics focus on the change ratio, and looking at the progression of speed in [Figure 5-2Figure 5-1], we do not notice such dramatic changes that can be reflected in these metrics. Additionally, comparing the previous 5h to the next 5h as blocks, will potentially hide some of the changes over time that will be averaged out and missed because of comparing the blocks. Besides, after examining multiple trips, we noticed the speed generally is not dropping greatly during storms which is clear in [Figure 5-2] as we notice the average speed before and after the clearance did not vary by a lot. This can mean several things, first, it means that drivers drive normally (around standard highway speed) while they can, second, it could mean that these observed weather surface conditions are not particularly slowing down the traffic. For example, temperatures could be around freezing degrees but if the road has low or no precipitation, ice pellets are less likely to develop which means vehicles can still drive around normal speeds on this road. Another example is that if the road is wet, due to rainy conditions, these conditions will not withhold traffic from its normal flow unless they are slick and dangerous. In summary, there are a lot of nuances that contribute to weather and surface conditions and their actual direct impact on traffic speed.

## **5.2 Efficiency with Sliding Window Recovery Detection**

Next, to account for the intricate changes in speed over time after clearance, we switched to determine convergence to normal speed conditions using a sliding window algorithm. Therefore, in this case, rather than comparing the speed distribution of a block of time before and after clearance, we instead compare it for each timestep to have a more accurate speed comparison. The following graph, [Figure 5-3] demonstrates this for a trip where we show the 30-minute sliding

speed mean, standard deviation, minimum, and maximum for both dry and current conditions. A vector of these speed metrics for a timestep is considered a distribution representative at an interval. Sliding dry speed metrics are calculated by sliding the window and filtering data in which conditions are not hazardous (dry), while current represents the actual observed speed conditions, hazardous and dry. You can see in the figure, the blue-shaded area showing the hazardous conditions, during which the speed metrics all reduce below their dry equivalents. However, when the storm clears and the conditions become dry, pink-shaded areas, we notice the current and the dry metrics are closing and the gap between the two distributions is becoming insignificant.

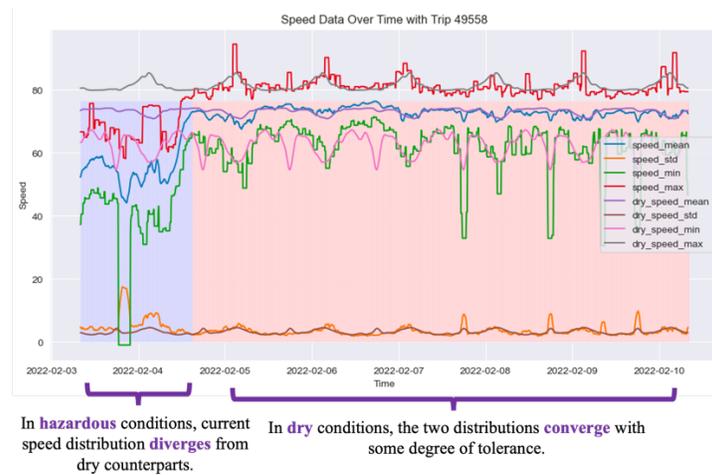


Figure 5-3 Illustration of Sliding Aggregated Speed Variation During Hazardous and Dry Conditions

## 5.2.1 Efficiency Metric

In this experiment, we define the efficiency  $E$  of a trip as the ratio of the clearance time  $C$  to the recovery time  $R$ , where  $C$  and  $R$  are measured in hours.

$$E = \left(\frac{C}{R}\right) \times 100\% \quad 5-3$$

The following figure provides a visual explanation of the intuition behind this efficiency measure. A highly efficient trip will improve the traffic overall conditions rapidly enough so that they converge back to their normal standards close to the trip's end, as shown in the upper part of the

figure. Conversely, lower efficiency is associated with examples where the recovery was a long time after the trip ended.

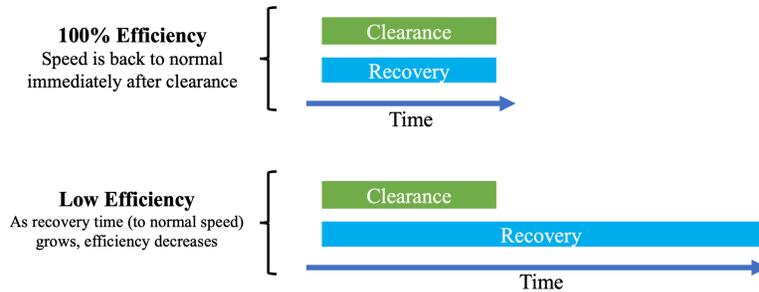


Figure 5-4 Illustration of Efficiency Measure using Clearance and Recovery Times

### 5.2.2 Recovery Detection Algorithm

Recovery time  $R$  is temporal difference, in hours, between the moment at which speed recovers to normal conditions post-clearance  $t_R$  and the start of clearance  $t_S$ , as shown in in [5-4]. The figure [Figure 5-5] provides a visual explanation of this, the intuition is that the speed distribution will eventually converge to the dry equivalent at  $t_R$ .

$$R = t_R - t_S \tag{5-4}$$

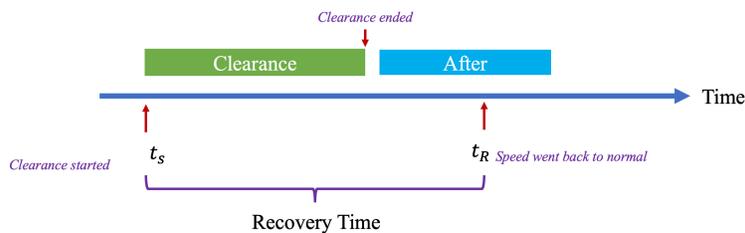


Figure 5-5 Illustration of Recovery Time Detection

Subsequently, to determine  $t_R$  – the convergence interval, we analyze the convergence of current and baseline (dry) speed distributions, represented by a vector  $S(t) = (\mu, \sigma, min, max, \kappa, \gamma)$  – which includes measures of centrality, variation, and shape of distribution over a selected previous time window (i.e., 15min). Adding the kurtosis  $\kappa$  and skewness  $\gamma$  enhances the representation by capturing the distribution shape, as shown in [Figure 5-6].

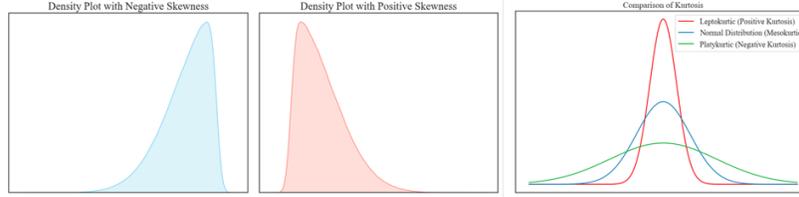


Figure 5-6 Illustration of Skewness and Kurtosis Roles in Capturing the Distribution Shape

First, speed metrics are normalized using MinMax scaling to account for scale variation across metrics and avoid affecting the dissimilarity measure. Scale variation is presented in [Figure 5-3], for example, the standard deviation is on a much lower scale compared to the speed mean. Unlike standardization (i.e., Z-score normalization), MinMax scaling does impose a Gaussian distribution on scaled data. Scaling is shown in [5-5], where  $X$  is the original value in each row of the vector for the feature (e.g., speed mean), and  $\tilde{X}$  is the scaled version:

$$\tilde{X} = \frac{X - \min(X)}{\max(X) - \min(X)} \quad 5-5$$

Subsequently, we apply a sliding window algorithm to calculate the dissimilarity between current and dry speed distributions at each time interval to find the convergence moment. As shown in [Figure 5-7], each time interval  $t$  will have a corresponding current speed distribution as well as a distribution for the dry conditions.

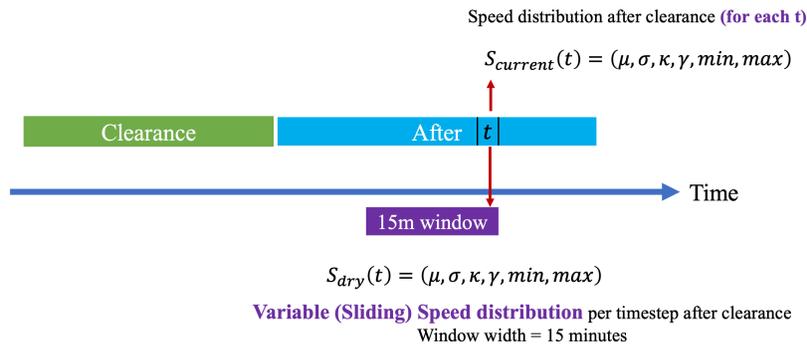


Figure 5-7 Illustration of Variable Sliding Speed Distribution

The dissimilarity (distance), in [5-6], between speed distributions representing current and dry conditions, is the weighted Manhattan distance (city blocks) between distribution vectors; that

include  $(\mu, \sigma, min, max, \kappa, \gamma)$ . The distance is calculated for each time interval  $t$  after the clearance ends to find recovery time. Once the distance  $D(S_1, S_2)$  drops below a certain, experimentally chosen, (tolerance) threshold  $\varepsilon = 0.3$ , then these distributions are considered to have converged to normal conditions, as shown in [5-6]. It should be emphasized that uniform weights were assigned to  $\mu, \sigma, \kappa$ , and  $\gamma$ , each set to a value of 1. In contrast, the weights for  $min$  and  $max$  were assigned a value of 0.25, as our analysis indicated higher variability or noise associated with these parameters, as evident in [Figure 5-3]. For consistency and interpretation, all weights have been normalized to ensure their summation equals one.

$$D(S_1, S_2) = \mathbf{w} \cdot \|S_1 - S_2\|$$

$$D(S_1, S_2) \leq \varepsilon \tag{5-6}$$

The following figure provides a distribution as a histogram of the measured efficiency across the trips. It is imperative to note that trips were subjected to specific filtering criteria. Specifically, any trip without hazardous conditions as defined by ODOT labels, lacking snowy conditions as determined by the precipitation type from the RWIS station, or with a clearance time of 30 minutes or less was excluded. After applying these criteria, the dataset was narrowed down to a total of 40 trips.

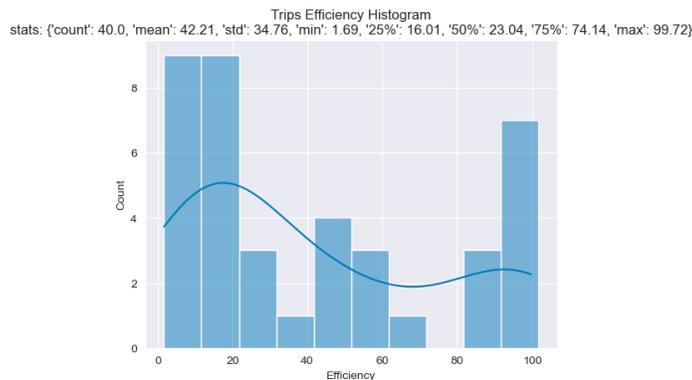


Figure 5-8 Distribution of Efficiency

Observing the distribution in [Figure 5-8], we notice that efficiency appears to have a bimodal distribution with two primary peaks. This suggests the existence of two groups of trips: those that have relatively low efficiency, in the 10-30% efficiency range, and those that are highly efficient, in the 80-100% range. The distribution also has a sparse mid-range recognized by the noticeable dip around the 40% to 70% efficiency range. Indicating a fewer number of trips with a moderate level of efficiency. The mean efficiency was 42.21% indicating that on average, trips are below the 50% mark efficiency. The variation in efficiency measured by the standard deviation was 37.76% suggesting a wide variation in the efficiency of the trips. The predominance of lower efficiency manifested by having more than half of the trips with an efficiency lower than 50%, since the median is 23.03%, might suggest that clearance has a subsidiary impact on traffic speed, existence of external confounding variables or factors affecting the speed, or that achieving higher efficiency is more challenging than one might assume. Despite this prevalence of lower-efficiency trips, a considerable portion of the trips with high efficiency, in the 80%-100% range, might suggest that in some scenarios or conditions, snow clearance is highly impactful on traffic speed.

### **5.3 Segment-level Speed Efficiency Evaluation**

Finally, we present the analysis on a segment level by analyzing the observed changes in speed average and standard deviation per segment over a selected window of one hour pre- and post-clearance. The change ratios for the average and standard deviation are calculated as shown previously in [5-1] and [5-2]. [Figure 5-9] shows the change ratio in both segment-level average speed (left) and standard deviation (right) for example trip 48741 near station 213. Data from segments traveling in both directions, north and south, are shown in the graphs. [Figure 5-10] displays weather data from RWIS station 213, indicating snowy conditions before and during

clearance, ceasing afterward, and confirming clearance operations were conducted in response to snow.

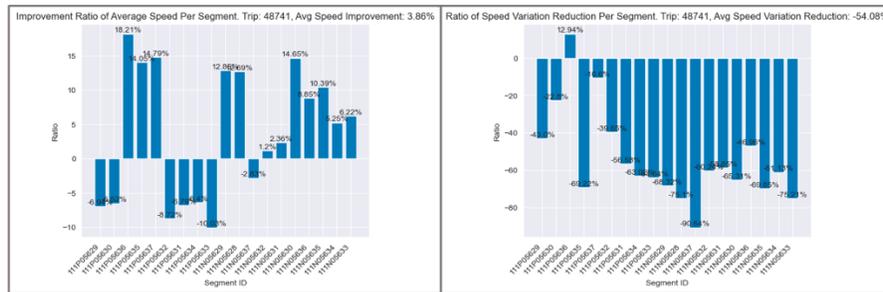


Figure 5-9 Change Ratio Per Segment in Speed Average and Variation for Trip 48741

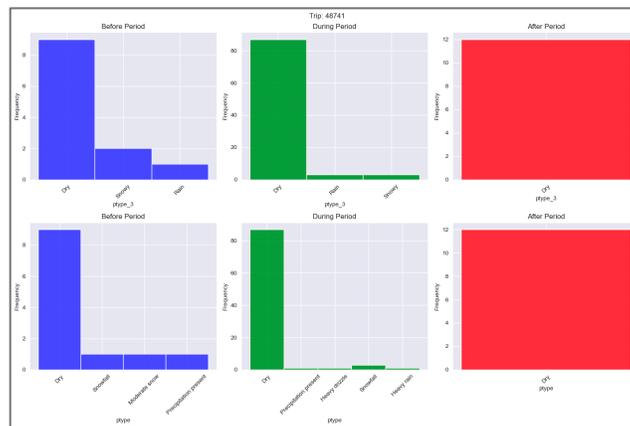


Figure 5-10 Observed Weather Condition by Nearby RWIS Station 213 Pre, During, and Post Clearance

Observing the change in average speed in [Figure 5-9] on the left, we notice that while some segments show a positive change in the average speed ratio, indicating a slight increase in speed, the extent of this improvement is modest, with all improvements being below 20%. This modest change can be attributed to the similarity in average speeds both before and after clearance. Essentially, since the speed doesn't drop significantly before clearance, post-clearance improvement is also limited. However, we notice that variation (right) dropped noticeably for most segments with a more pronounced range with a mean reduction of 54.08% - suggesting that clearance significantly helps in reducing variation in speed, thus, resulting in free-flow traffic.

The following figure shows the distribution for both average speed (left) and standard deviation (right) change ratios, as histograms with bin widths of 5, and 10, respectively. We can see that the same analogy of the results shown for the previous trip generalizes across the rest of the trips near station 213. The improvement in speed average, on a segment level, is limited to below 20% and with a central tendency in the 0-5% range. Speed variation distribution appears to be left skewed, note that a negative change ratio corresponds to a reduction in speed variation. A significant number of segments experienced a reduction in speed variability post-clearance. Fewer segments saw an increase in speed variability, evident from the dip in the distribution in the positive range.

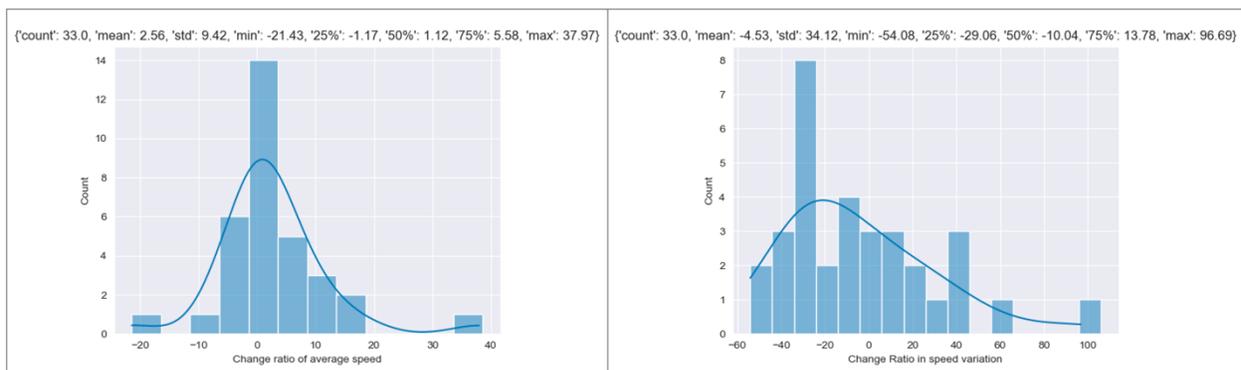


Figure 5-11 Distributions of Change Ratio in Speed Average and Variation

Therefore, to summarize. For average speed, snow clearance has a varied effect on the average speed of various segments. While some segments experience a mild increase in average speed, others observe a slight decrease. However, for most segments, the impact is minimal. For speed variation, clearance tends to reduce speed variability in a significant number of segments. This could mean that post-clearance, traffic flow becomes more consistent and less erratic. However, a few segments see increased variability, which may be due to other influencing factors or conditions. In conclusion, while snow clearance generally stabilizes traffic flow by reducing speed variability, its effect on the average speed of segments is more mixed and subtle.

## **5.4 Summary**

To summarize, the investigation of traffic speeds post-snow clearance showed limited improvement in speed metrics across 64 trips, with only 38% of trips showing notable recovery times within 5 hours post-clearance. Average speeds during storms were not significantly reduced, indicating that certain weather conditions had minimal impact on traffic flow.

Efficiency analysis revealed a bimodal distribution, with most trips falling below 50% efficiency, suggesting that factors other than snow clearance significantly influence traffic speed. While snow clearance generally reduced speed variability, indicating a stabilization of traffic flow, its impact on average speed was minimal and inconsistent across different segments.

## **6 Importance of Subsurface Probes in RWIS**

In this section, we investigate the importance of subsurface temperature probes in Road Weather Information Systems (RWIS). This research study achieves this objective by analyzing and quantifying the relationship and potential impact, during winters, on each of the following: ambient/surface temperatures, weather events (i.e., icy, snowy, rainy, dry), road surface conditions (RSC) (i.e., hazardous, dry), and road clearance time.

The motivation and hypothesis behind this are that utilizing subsurface temperature probes in RWIS may lead to enhanced prediction of hazardous surface conditions and adverse weather events and achieve more efficient snow removal operations by focusing clearance efforts of snowplow trucks on roads that need it the most during extreme weather. Another potential use case of such models involves traffic management by adjusting speed limits during hazardous surface conditions which may prevent accidents. Such findings can hopefully improve the overall road safety and prompt incorporation of subsurface temperature probes into existing RWIS systems that do not have them.

Existing research literature appears to have overlooked the significance of subsurface temperature probes in RWIS, underscoring the need for this study to prove or disprove their importance and provide oriented insights for the public. Literature studies have focused on exploring different methods for forecasting as well as instantaneous prediction of road surface temperature (RST) as highlighted in section 2.

This study utilized data from RWIS stations deployed on I-35 with a focus on data corresponding to the 6 northern stations as this region usually exhibits the coldest climate. The stations provide a variety of weather data, every 5 minutes, including ambient, surface, and subsurface temperatures. The data also includes the precipitation type, measured as a number, corresponding to various

weather events covering dry, rainy, snowy, and icy weather, as shown in [Figure 6-1]. These values are then mapped to a coarser representation including dry, rainy, snowy, and icy representing weather events in the analysis.

**Precipitation Type Table**

Synop Code w <sub>s</sub> w <sub>s</sub>	Meaning
0	No precipitation.
40	Precipitation present.
51	Light drizzle.
52	Moderate drizzle.
53	Heavy drizzle.
61	Light rain.
62	Moderate rain.
63	Heavy rain.
67	Light rain and/or drizzle with snow.
68	Moderate rain and/or drizzle with snow.
70	Snowfall.
71	Light snow.
72	Moderate snow.
73	Heavy snow.
74	Ice pellets.
89	Heavy hail.

*Figure 6-1 Precipitation Type Table Mappings to Events of RWIS Weather Sensor [10]*

Road surface conditions (RSC) data, which are manually collected by visual inspection by ODOT operators, were used to label RWIS weather data as hazardous and dry. Hazardous conditions include slick spots highway, slick hazardous highway, light snow, moderate snow, heavy snow, snow packed, blowing snow, and ice. Given this, we can immediately imagine the potential benefits of an accurate model predicting extreme surface conditions since the process of manual road weather inspection might be prone to human error manifested in subjective or inaccurate evaluation. Lastly, clearance time data was acquired from the MATT website which receives time and location data, along with other data, from the real-time mobile application.

## **6.1 Exploratory Data Analysis (EDA)**

Before jumping into the analysis, let us first inspect the progression of temperatures over time, shown in [Figure 6-2], for 5-minute data collected from station 35ST213, for February 2022, as an example, which witnessed two storms at the beginning and end of the month. We notice that

subsurface temperatures expectedly follow the same pattern depicted by ambient and surface temperature. However, we observe that ambient and surface temperatures dip below subsurface temperatures during storms. Also, subsurface temperatures exhibit less variation compared to ambient and surface especially during storms. Therefore, improving our understanding of the relationship between subsurface temperature and correlation with ambient and surface temperature can potentially uncover useful insights, particularly, during storms; supporting the need for this study.

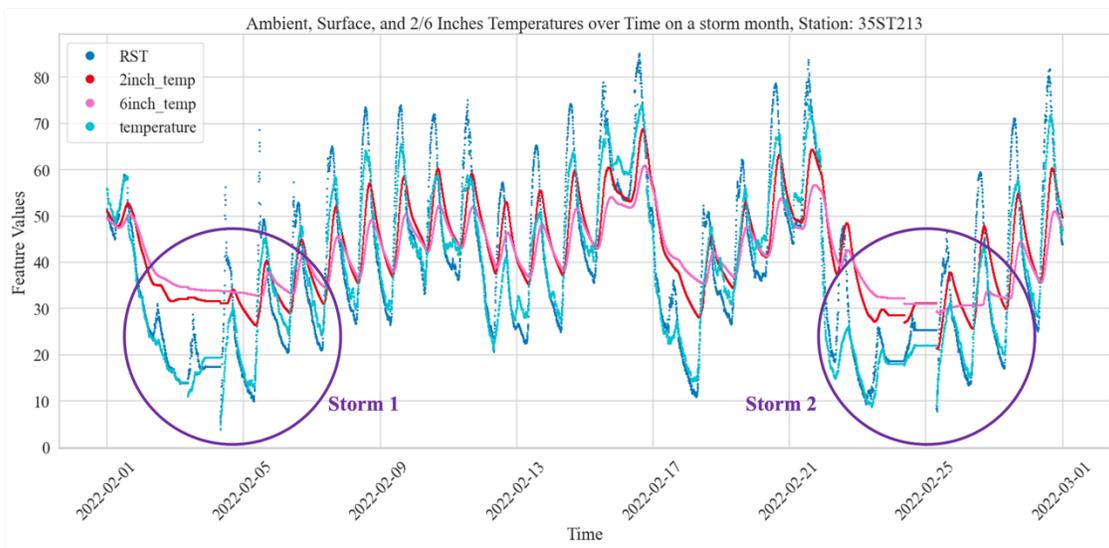


Figure 6-2 Illustration of Temperature Data Over Time and Observed Variation During Storms

If we observe the monthly distributions of surface and subsurface temperatures for winter months shown in the following figures, we notice that the variation in temperature decreases as move deeper, from surface to subsurface probe 1 and deeper to probe 2. Also, months that experienced storms such as February 2021 and 2022 show a larger variation across all three temperatures compared to other months, these high variations are coming from the drops in temperature from the norms such as the ones shown in [Figure 6-2].

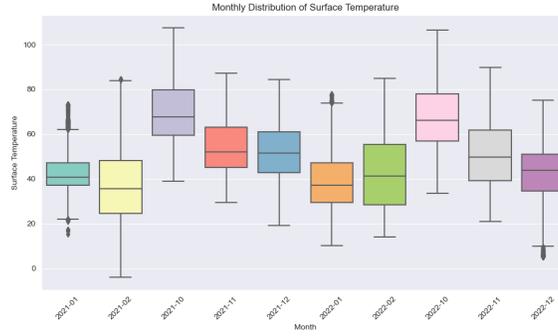


Figure 6-3 Monthly Distribution of Surface Temperature

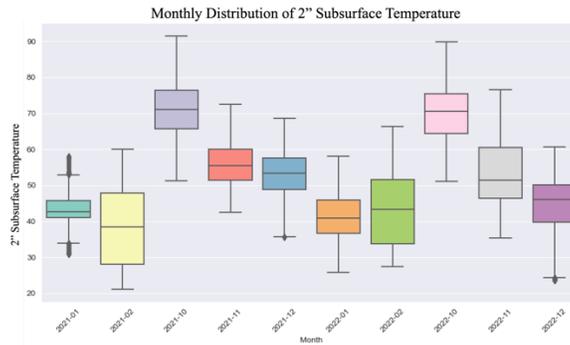


Figure 6-4 Monthly Distribution of 2\" Subsurface Temperature

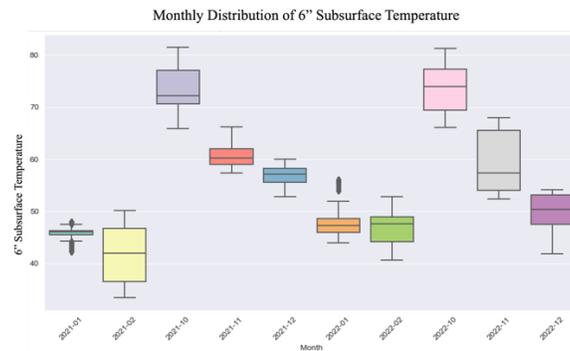


Figure 6-5 Monthly Distribution of 6\" Subsurface Temperature

## 6.2 Hypothesis Testing

Hypothesis testing serves as a foundational step in exploring the importance of subsurface probes in RWIS by establishing a preliminary understanding of the key relationships or lack thereof in the data before diving into intricate analyses like regression or machine learning models. This methodical approach ensures that subsequent analyses are grounded in statistically verified

relationships, adding robustness and credibility to our findings. In simple terms, hypothesis testing validates whether the observed data aligns with or deviates from what is expected under certain assumptions of a null hypothesis. Statistical significance tests are used to validate whether the observed effect or relationship between data attributes is genuine or might occur by random chance, quantifying the confidence levels in the results. These tests assess the p-values associated with the correlation tests. If the p-value is below a predefined significance level (e.g., 0.05), we can reject the null hypothesis and conclude that there is a statistically significant relationship.

### 6.2.1 Correlation Analysis

Pearson’s correlation was used to conduct hypothesis testing for the relationship between subsurface temperatures and other measured temperatures including ambient, surface, and bridge. The null hypothesis (H0), in this case, states that there is no significant relationship between these variables. The alternative hypothesis on the other hand suggests the existence of such a relationship. Using Pearson’s correlation, [Figure 6-6] shows a heatmap of correlation coefficients for temperature data for station 35ST213 for Sep, Oct, Nov, Dec, Jan, Feb, and March representing the winters of 2021 and 2022. The data displays relatively high correlation coefficients among subsurface temperatures and other variables. Note that the p-value is less than 0.005. Therefore, we can refute the null hypothesis and accept the alternative one.

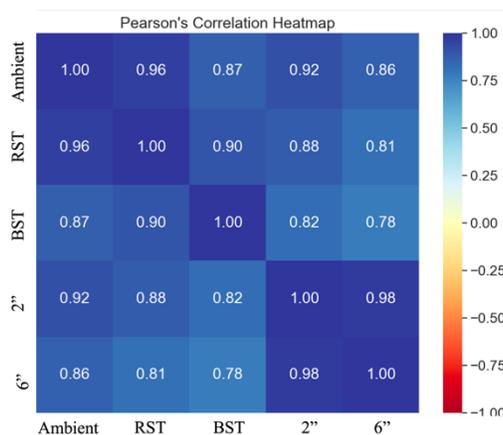


Figure 6-6 Correlation Heatmap of Ambient, Surface, and Subsurface Temperatures

### 6.2.2 Analysis of Variance (ANOVA)

To conduct hypothesis testing for subsurface temperature with non-continuous variables such as road surface conditions (hazardous or dry), we use Analysis of Variance (ANOVA) which if significant (high F-statistic) tells us that the mean of at least one group deviates from the rest; proving heterogeneity in temperature between groups. [Figure 6-7] shows a violin plot representing the distribution of subsurface temperature data for both probes during cooler months with the corresponding F-statistic for station 213. Similarly, we can accept the alternative hypothesis and say that subsurface temperature has a significant correlation with road surface conditions.

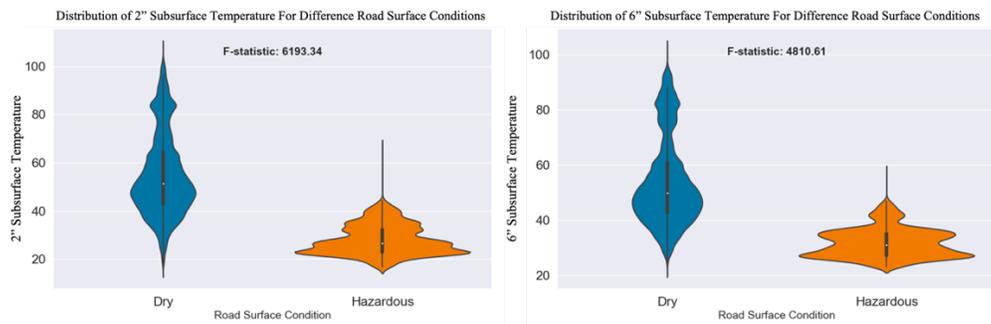


Figure 6-7 Distribution of Temperature per Surface Condition at 2" and 6" levels

For weather events, [Figure 6-8] shows the distribution of subsurface temperature for both probes, and [Figure 6-9] shows surface, and ambient temperature for various events during winter months.

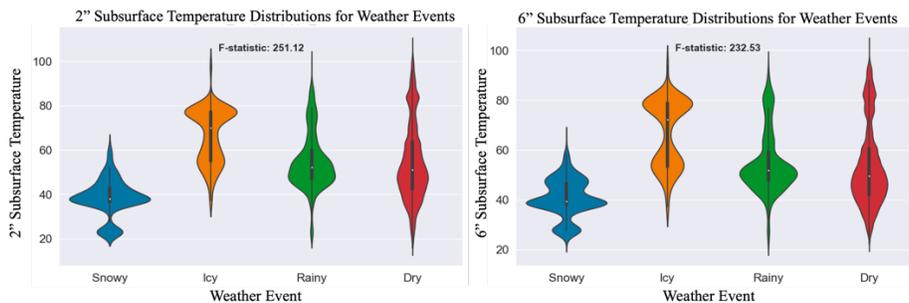


Figure 6-8 Subsurface Temperature Distribution at 2" and 6" for Various Weather Events

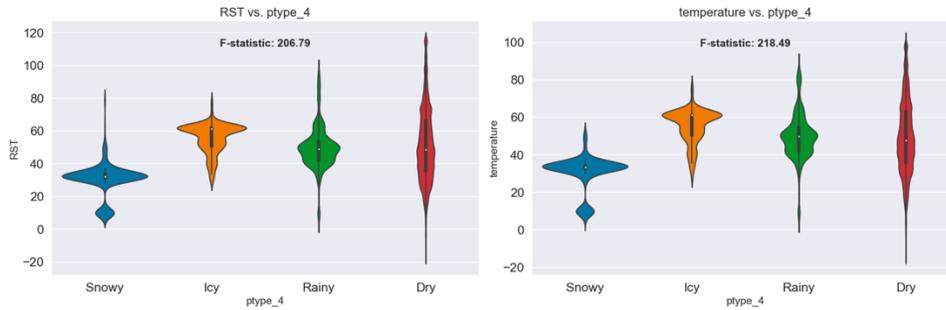


Figure 6-9 Surface (left) and Ambient (right) Temperature Distribution for Various Weather Events

The ANOVA tests prove, with F-statistic scores of 251 for probe 1 and 232 for probe 2, that the subsurface temperature groups, for different events, are not homogenous. It is worth mentioning that due to thermal inertia subsurface temperatures during winter are warmer than ambient surface temperatures [16], which explains why the temperature distributions for snowy, icy, and rainy conditions are warmer for subsurface probes compared to ambient and surface measurements. However, if we look at the distribution of the icy events, representing ‘ice pallets’ and ‘heavy hail’ in the precipitation type table [Figure 6-1], we notice that it is higher than expected. We expect to see a distribution like snowy events, shifted down to lower freezing degrees. This might suggest possible sensor measurement inaccuracy or noise which could mean that it will be trickier to develop models that can identify weather events accurately. But it could also mean that other unknown confounding variables are affecting the distribution.

### 6.3 Regression Analysis

Regression analysis allows us to model the rate of change between subsurface temperature and other variables such as ambient temperature by drawing a regression line and calculating the slope. The slope also provides the orientation of the relationship as positive representing direct linearity or negative representing reciprocal relationships. The correlation coefficient gives us the strength of the relationship. Statistical significance tests were used to make sure confidence levels were high. By comparing regression lines across various weather and surface conditions, we can

improve our understanding of the relationships over these conditions. Given that temperature is usually colder at night compared to daylight, this gives a finer understanding of the relationship by inspecting both rather than grouping all temperature values through the day and looking at these regression lines. Brightness, measured in Lux provided by the weather sensor [10], was used to classify data into two categories: day and night. If the brightness level is less than 100 Lux, then it is considered as day, otherwise, it is labeled as night. A rolling average window of 30 minutes was used to smooth out the brightness over time for 5-minute data since raw data can be noisy. Sun radiation affects all temperatures, which is why we are using brightness to label the data compared to only using time periods to separate day and night because time will not be as accurate as actual sensor measurements because of overcast days, temporal (daily) variations in the sunset and sunrise times, etc.

Looking at winter data Oct-March for 2021 and 2022. Regression lines for surface conditions (RSC) for night and day are shown in [Figure 6-10]. The regression lines during dry conditions for both day and night have linear with slopes  $m = 1.04$  and  $m = 1.05$ , respectively. However, the slopes become steeper during hazardous conditions with values  $m = 1.67$  and  $m = 1.21$  for night and day, respectively. This corresponds to 60% increase from 1.04 to 1.67 in rate of change from dry to hazardous conditions during the night. During the day, the increase is less prominent with a percentage increase of 15% accounting for the change from  $m = 1.05$  to  $m = 1.21$ . To better understand the interpretation of the slope, for instance, for  $m = 1.67$ , this means for every one-degree of change in subsurface temperature, the ambient temperature will change by a magnitude of 1.67, the steeper the slope, the faster the change rate. In summary, observing subsurface temperature during hazardous conditions, especially during the night, is critical. This is because the rate of change is 60% higher than in dry conditions, which could potentially mean it is more

dangerous since strong drops in ambient temperature will help in forming ice on the road or causing snowfall.

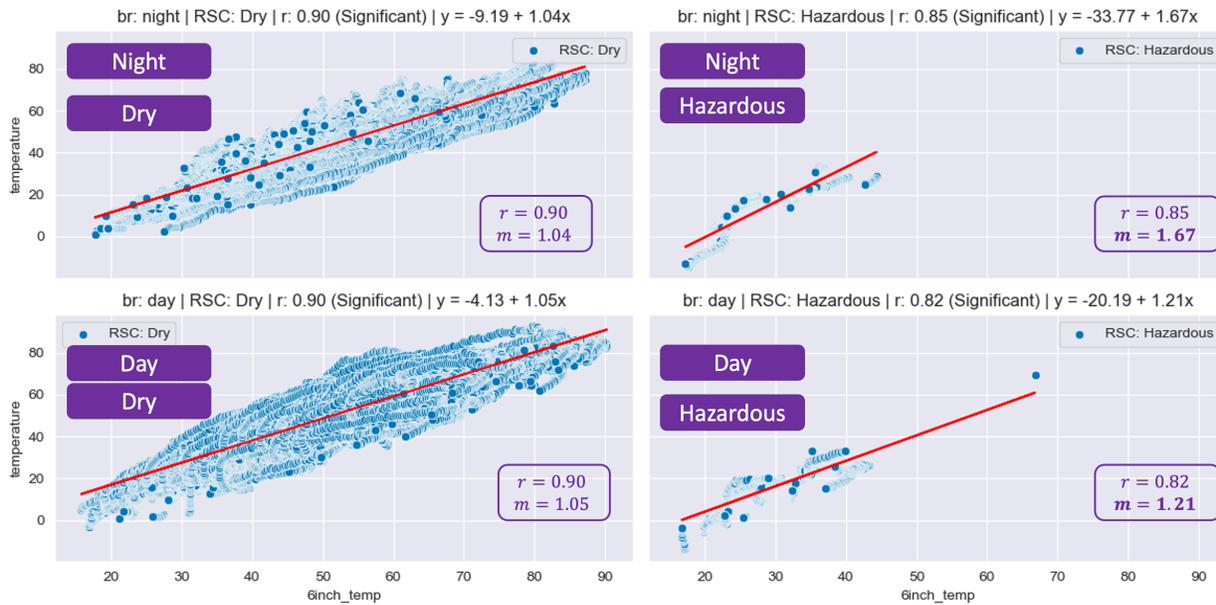


Figure 6-10 Regression Slope per RSC for Daytime and Nighttime

Next, let's explore the relationship between subsurface and ambient temperature for various weather events, shown next in [Figure 6-11]. Similar to the results observed for surface conditions, we observe an almost perfect linear relationship during dry weather for both day and night, when precipitation is not present night and day do not seem to affect the rate of change or correlation coefficient. Rainy weather displays similar patterns in relationship to dry. Unlike surface conditions, we do not notice steeper slopes when comparing dry to icy or snowy weather. However, we do notice stronger correlation coefficients and steeper slopes for both icy and snowy when comparing day to night. For snowy weather, a coefficient of 0.74 and a slope of 0.68 during the day change to 0.88 and 1.02 during the night, respectively, corresponding to percentage increases of 19% for correlation and 50% for slope. For icy weather, the coefficient and slope increase from 0.54, and 0.36 to 0.86, and 0.65, respectively corresponding to percentage increases of 59% for correlation and 81% for slope. In summary, like surface conditions, observing the relationship

between subsurface and ambient temperatures is useful, especially for icy and snowy weather as the correlation and rate of change increase from daylight to nighttime.

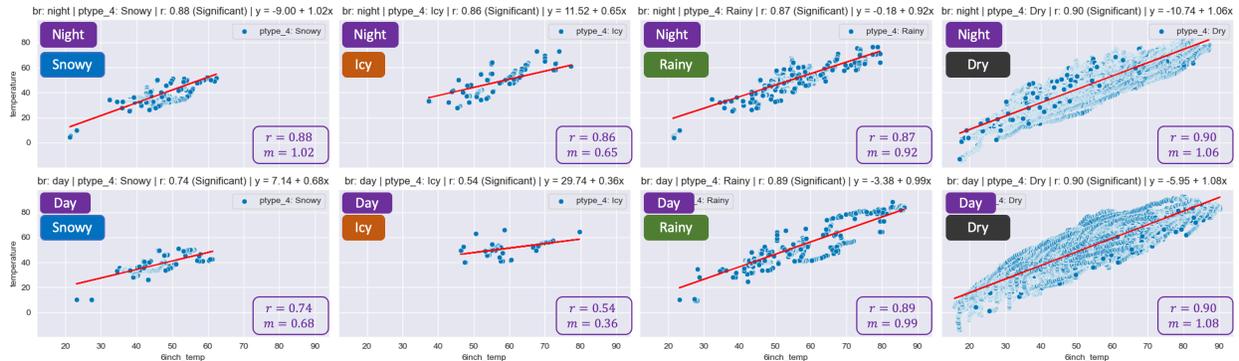


Figure 6-11 Regression Slope per Weather Event for Daytime and Nighttime

Another idea is to observe and analyze the relationships during crossings where the ambient temperature drops below subsurface temperature, as shown during storms earlier in [Figure 6-2]. After dividing the dataset into two groups, the first represents when the ambient is below the subsurface and the other group represents the other case. We first notice that all hazardous road surface conditions fall in the first group and none in the second group where the ambient is above the subsurface temperature. This makes the regression relationships the same as observed in [Figure 6-10]. Second, and similarly, most icy, and snowy events lie in the first group which makes the relationships quite like those in [Figure 6-11]. Consequently, the relationship between subsurface and ambient temperature is quite important for the detection of hazardous surfaces and adverse weather events (icy and snowy) because, once the levels of ambient temperature drop below subsurface counterparts, these conditions become more likely to develop.

## 6.4 Relationship with Clearance Time

To find out the relationship between subsurface temperatures and clearance time, we selected trips that pass within a 3-mile vicinity of the weather station, thus, falling within the weather scope of the station. Trips that do not have snowy conditions are very short (less than an hour), or whose

road surface conditions correspond to less than 20% of the total clearance time of the trip were filtered out. This results in 12 trips for January and February of 2022 near station 213, the distribution of the clearance time for the trips is shown in [Figure 6-12]. [Figure 6-13] shows the distribution as a boxplot of the subsurface temperature of the first-level probe during the clearance per trip and the distribution for the preceding hour of the clearance. Essentially, we are capturing both the state of the subsurface temperature during clearance in addition to a window of 1 hour before clearance starts; which serves as a baseline, control, or a reference. It is useful to look at a window preceding the clearance for a few reasons, first, clearance usually happens during hazardous surface conditions while the preceding hour will describe the normal, or usual, distribution of subsurface temperature. Additionally, comparing subsurface temperature during and before clearance could provide insights into how the subsurface temperature evolves during a snow event and clearance efforts. Now, by analyzing the results shown in [Figure 6-13] showing distributions of subsurface temperature over clearance time, we notice that as clearance time increases, the centrality (e.g., mean) of the subsurface temperature distributions shifts to lower degrees. Note that the increased variation in clearance time is explained by the longer duration (hours) of the trip, to validate this, we reproduced the figure where the preceding window was equal to the trip duration, and the variation changed in the same manner for both before and during clearance.

To relate subsurface temperature to clearance time using regression analysis, we scatter plot the means of the subsurface temperature distributions over clearance time for both periods, during and before the clearance, and calculate and analyze the corresponding regression lines, as shown in [Figure 6-14]. We notice that clearance time increases as the mean subsurface temperature decreases, with a negative correlation of  $r = -0.68$  and a slope of  $m = -0.64$ . The slope is

steeper during clearance with a value of  $m = -0.64$ , compared to  $m = -0.37$  for the period before clearance.

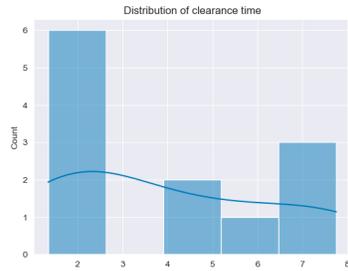


Figure 6-12 Distribution of Clearance Times

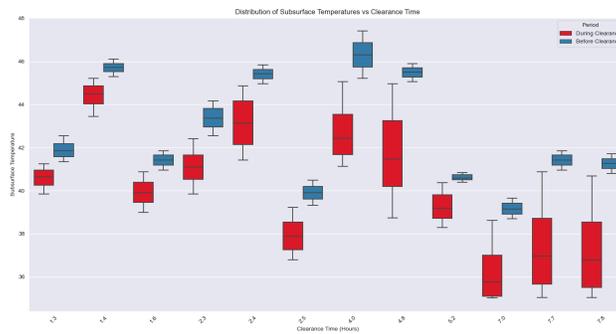


Figure 6-13 Distribution of Subsurface Temperature over Clearance Time Before and During Clearance

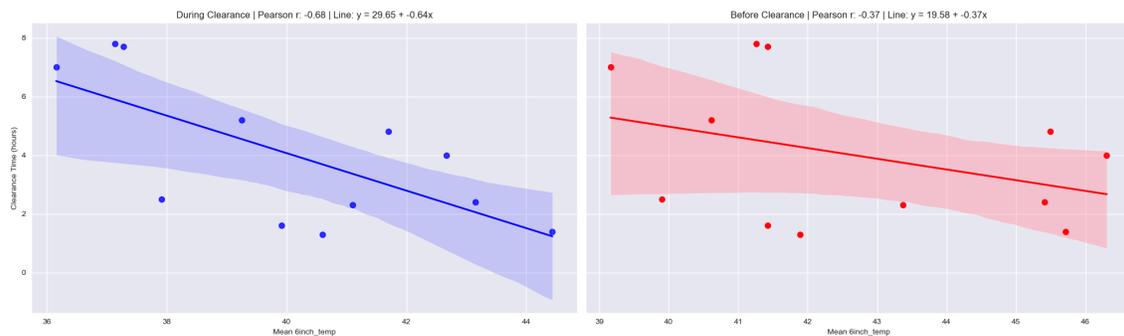


Figure 6-14 Regression Slope for Clearance Time over Mean Subsurface Temperature, Before and During Clearance

## 6.5 Machine Learning Model Development

To show the impact of the inclusion of subsurface temperature given by RWIS probes on machine learning models, we developed three models explained in [Figure 6-15]. The first model uses ambient, surface, and subsurface temperatures as features. The second model uses only ambient and surface temperatures as features while the third model uses only probes (subsurface

temperatures) as features. Using only these temperatures as features for the experiment will ensure that we observe the effect of incorporating probes in such models without having other weather features, once their effect is established, we can add other features as predictors. The models are trained to classify road surface conditions (RSC) in addition to weather events. The data was prepared for modeling by handling missing data using imputation using a window for the previous 15 minutes and eliminating outliers by thresholding. The dataset was split using an 80/20 percent split for training and testing, respectively. Data features were standardized to ensure that none of the features dominated the others because of having a different scale. This is because some machine learning models might give higher weights to the features with larger scales, thus, scaling all features will prevent this.

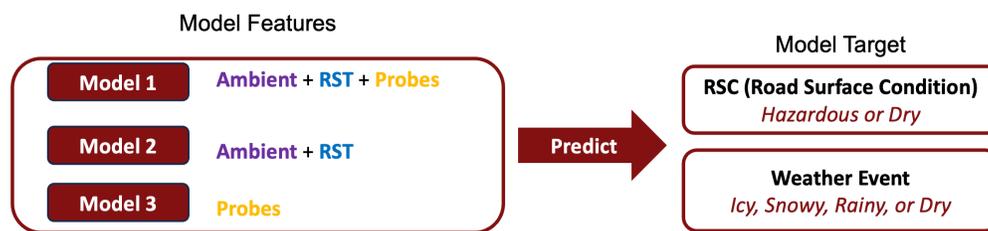


Figure 6-15 Illustration of the Machine Learning Models Experiment

The results of the experiment were conducted using a decision tree model where the Gini importance was used to calculate feature importance; the more a feature reduces Gini impurity (criterion), the higher the assigned weight. [Figure 6-16] shows the results for the classification of road surface conditions into hazardous and dry. Given the imbalanced nature of the dataset shown in [Figure 6-16], using appropriate metrics was necessary to avoid misleading results. Therefore, the confusion matrix and a precision-recall graph with an F1 score were used as opposed to using a measure like accuracy which inherently favors the major class. The first model appears to have shown the best results on the unseen test set with minimal false positives (0.12%) and false negatives (5.52%), and the highest F1 score (0.94). However, what is interesting is that model 3

(using only probes) managed to achieve better results with an F1 score of 0.74 compared to model 2 (using ambient and surface) with an F1 score of 0.64 corresponding to a 17% F1 score improvement which shows that subsurface temperatures have strong predictive power for road surface conditions. But models 2 and 3 are not perfect and they struggle in differentiating surface conditions with 38.45% of all hazardous falsely reported as dry by model 2 and 24.34% false negatives by model 3, whereas all three models seem to have minimal false positives. The imperfect percentage of false negatives is expected within this controlled test we are doing since our aim is to show the effect of using probes in models rather than perfecting the performance of the models. Plus, road surface conditions have many confounding factors affecting them which won't be all captured by the given features. To improve the performance of such models, multiple possible approaches can be taken such as incorporating more weather features since we noticed from the results that by using only subsurface temperatures along with ambient and surface temperatures, we get an F1 score of 0.94 (a high precision and recall model), this suggests that adding more weather features can potentially enhance these metrics even further. Another approach is implementing SMOTE (Synthetic Minority Oversampling Technique) to oversample the minority classes in the training set while keeping the test set imbalanced reflecting the real world. Inspecting the feature importance for the models, we see that model 1 favors ambient, followed by subsurface temperatures, and then surface. Suggesting that subsurface temperature is of higher importance for prediction compared to surface temperature. Model 2 also favors ambient over surface temperature. Model 3 only slightly favors the deeper probe of the two probes.

In retrospect, comparing the results of the three models, we conclude that incorporating subsurface temperatures significantly improves results, and model 1 which utilizes all these temperatures as features has only around 5% misdetections of hazardous conditions on the test set.



Figure 6-16 RSC Class Distribution for Training and Testing Datasets

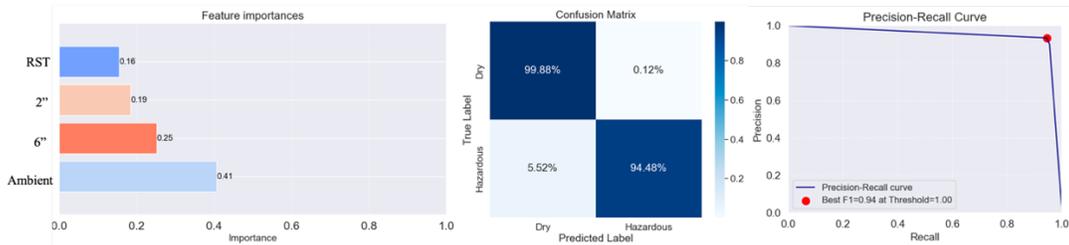


Figure 6-17 Model 1, RSC Results: Feature Importance, Confusion Matrix, PR Curve

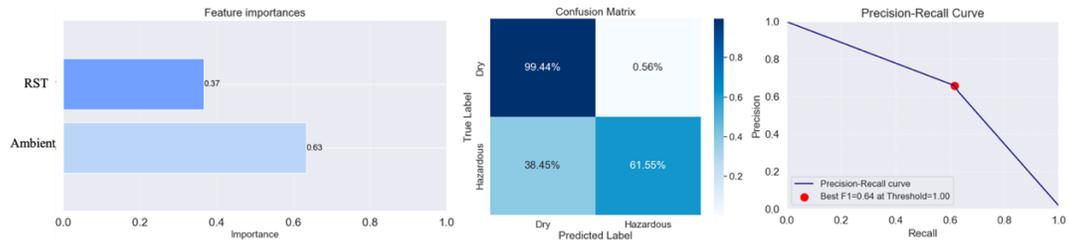


Figure 6-18 Model 2, RSC Results: Feature Importance, Confusion Matrix, PR Curve

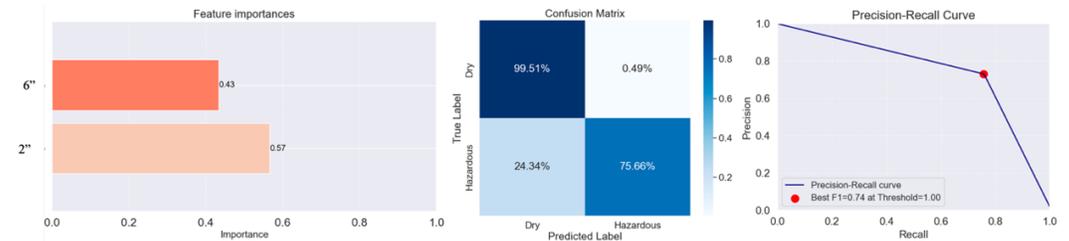


Figure 6-19 Model 3, RSC Results: Feature Importance, Confusion Matrix, PR Curve

Next, let's inspect the model's results for the classification of weather events into icy, snowy, rainy, and dry. By observing the confusion matrix of model 1 and comparing it to models 2 and 3, we notice that model 1 outperforms other models in identifying minority classes (i.e., icy, snowy, and rainy) but still has some imperfections. Also, once again model 3 results are better than model 2

but only slightly for snowy and rainy conditions, [Table 6-1] summarizes F1 scores for minority conditions. In summary, the models show the same performance pattern where using all temperatures as features yields the best results, followed by the model of the probe and then ambient with the surface temperature model. Additionally, the models seem to be less efficient in classifying weather events compared to road surface conditions, this is obvious by inspecting the misdetections in the corresponding confusion matrices and the lower F1 scores. We have seen earlier in this chapter that the distributions of subsurface temperature for different weather events are not as distinct as the observed distributions for various surface conditions which makes differentiating weather events harder. However, the classification of weather events shows that by incorporating probes in models, we get better results.

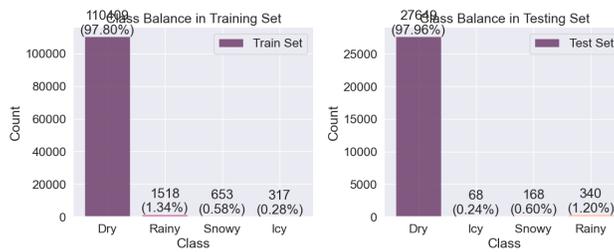


Figure 6-20 Weather-Event Class Distribution for Training and Testing Datasets

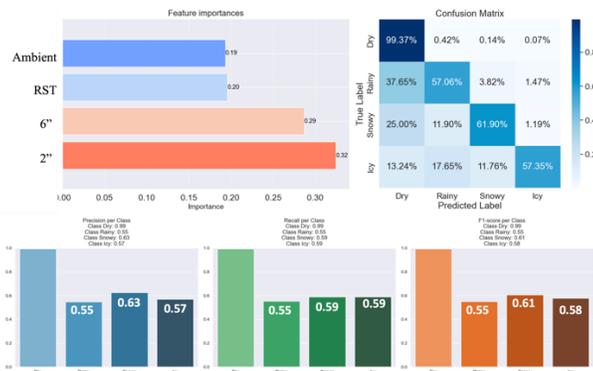


Figure 6-21 Model 1, Weather-Event Results: Feature Importance, Confusion Matrix, Per class: Precision, Recall, F1 Score

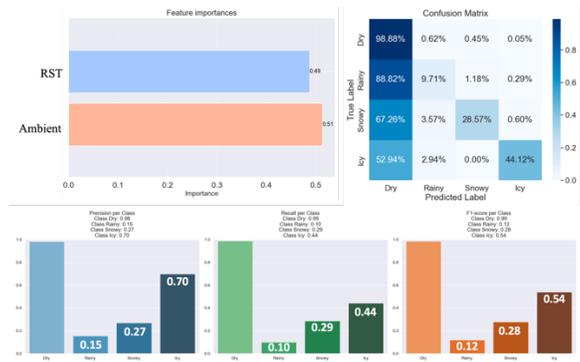


Figure 6-22 Model 2, Weather-Event Results: Feature Importance, Confusion Matrix, Per class: Precision, Recall, F1 Score

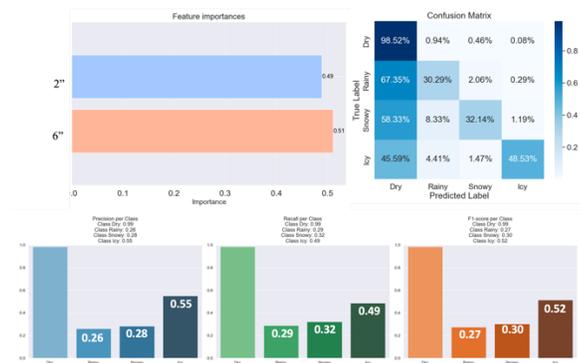


Figure 6-23 Model 3, Weather-Event Results: Feature Importance, Confusion Matrix, Per class: Precision, Recall, F1 Score

Event	Model 1	Model 2	Model 3
Icy	<b>0.58</b>	0.54	0.52
Snowy	<b>0.61</b>	0.28	0.30
Rainy	<b>0.55</b>	0.12	0.27

Table 6-1 Summary of Models, Weather-Event F1 Scores on Minority Classes

## 6.6 Seasonal Variation and Trend Analysis through Parametric Modeling

In this section of the chapter, we focus our analysis on understanding the relationship between subsurface temperature and ambient temperature as the weather transitions into the colder months. The goal is to develop a parametric model as a function  $f$  mapping the relationships over time between ambient and subsurface temperature. Investigating the temporal change of this parametric relationship for both daylight and nighttime will help us enhance our understanding of the dynamics of these signals and, more importantly, highlight the role and effect of subsurface temperature on ambient temperature – that directly affects road surface conditions.

### 6.6.1 Regression Parametric Modeling

This relationship is modeled by the slope  $m$  of a linear regression model, as in [6-1], whose independent variable (or explanatory variable)  $X$  is the ambient temperature belonging to the x-axis and the dependent variable (or response variable)  $Y$  is the subsurface temperature of probe 1 or 2 belonging to the y-axis. Where  $\beta$  is the intercept and  $t$  is time.

$$\hat{Y}_t = f_{m,\beta}(X_t) = mX_t + \beta \quad 6-1$$

The choice of linear regression as a parametric model for this task is due to its simplicity, interpretability of slope and intercept, and its ability to capture the trend in data, thus, uncovering strong associations even if they do not strictly imply causation. Ordinary Least Squares Regression (OLS) was used to estimate coefficients of the best-fitted regression line by minimizing the sum of square errors (SSE). The error captures the square difference between the predicted subsurface temperature by the model and the actual measured temperature.

Data collected from various weather stations at 5-minute intervals for a period spanning 27 months from January 2021 up to March 2023 was used in this analysis. Outliers were removed by thresholding as some data points were outside the possible range of temperature values. The data

cleaning and processing stage included inspecting temperature data over time and eliminating data that had issues. For example, some stations observed time periods where temperature is a constant value. Another example is when there is a loss in data for some months which might be due to multiple potential reasons such as sensor error or station downtime. [Figure 6-24], is an illustrative example of a typical month of temperature data, for January 2021 from Station 35ST213 – the number of samples is 8928 corresponding to 31 days of 5-minute data. Ambient temperature is shown in dark blue, subsurface at 2” in brown, and at 6” in light blue. As expected, we notice that ambient is observing the most variation with time, while as we go deeper underground this variation becomes less pronounced and has some lag in it.

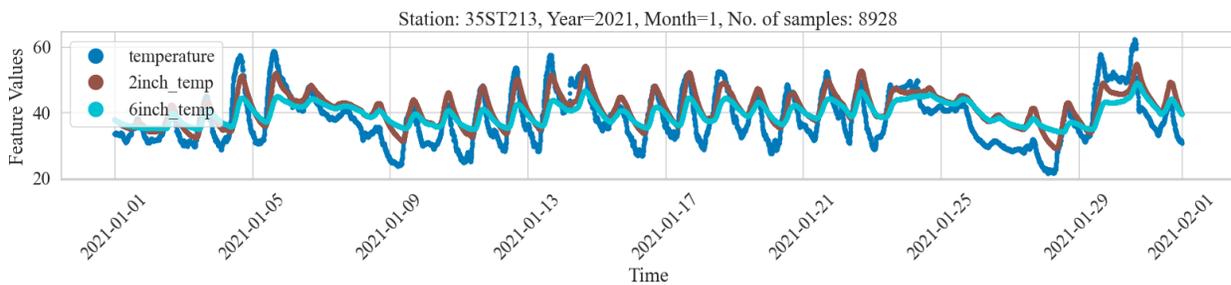


Figure 6-24 Signal Dynamics of Ambient and Probes Temperature over Time

## 6.6.2 Temporal Segmentation

Having prepared the data for the analysis, we subsequently segment the data points into daylight and nighttime where 6 AM to 6 PM is considered daylight and the complimentary is nighttime. Segmenting the data into these two classes is to capture the characteristic changes in temperature where daylight data will contain increasing patterns due to sunlight radiation, whereas nighttime data will contain decreasing trends in temperature. The relationship must be different in these two circumstances, and it is important to capture these nuances in the relationship over time. [Figure 6-25] shows the segmented ambient temperature, highlighted with different opacity levels, of

January 2021 for station 35ST2123. We can see those increasing and decreasing patterns, mentioned earlier, in this figure.

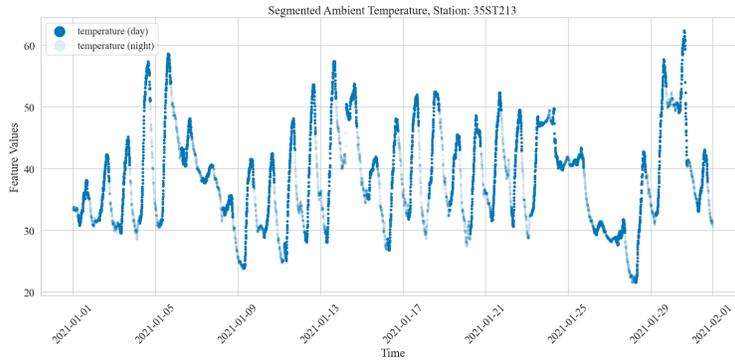


Figure 6-25 Increasing and Decreasing Temperature Patterns for Daytime and Nighttime

[Figure 6-26] shows an example scatter plot of the ambient and 2” subsurface temperature for November 2022 for station 35ST213. The left plot shows all temperature data, whereas the two plots on the right show segmented daylight and nighttime data points. First, in all three cases, we notice that the data has a linear relationship with an oval-like shape. The graphs also show the regression line equation, Pearson’s correlation coefficient  $r$ , and descriptive statistics of both variables including mean, standard deviation, kurtosis, and skewness. We also notice that nighttime (middle) data points have a thinner shape compared to daylight (right) or overall distribution (left). We model the relationship using a regression model as shown below per month, and consequently, we can observe the changes in the rate of change over time.

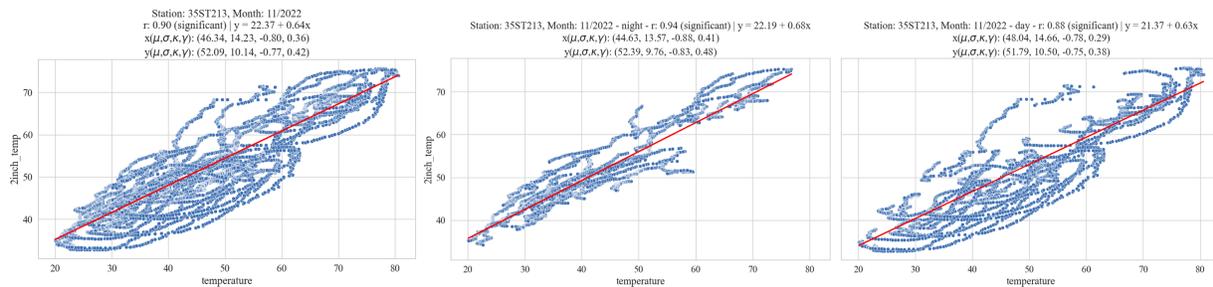


Figure 6-26 Regression Lines for 2” and Ambient: Unsegmented, Nighttime, and Daytime

The following figure shows scatterplots with regression lines for night (left) and day (right) from September 2022 to March 2023. Providing a visual presentation of the shape changes over time, especially in winter months. Notice that winter months (November, December, and January) have thinner distributions, especially at night, compared to other months.

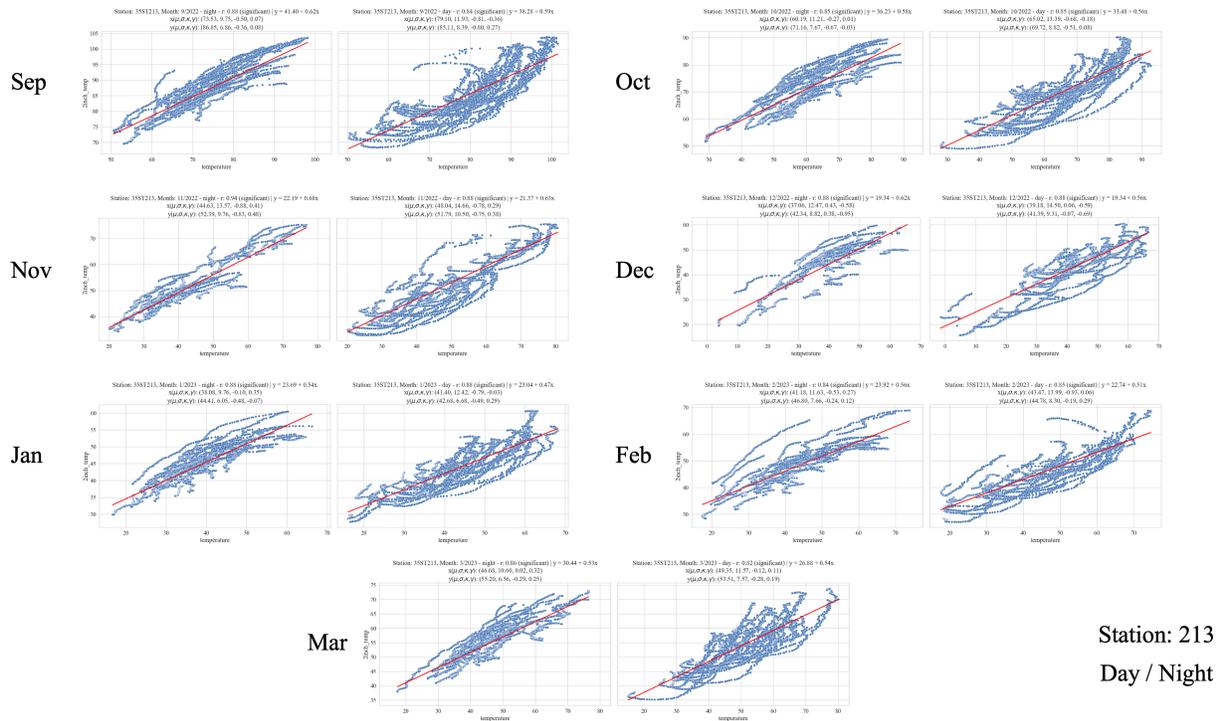


Figure 6-27 Monthly Regression Lines for 2" and Ambient: Nighttime, and Daytime

### 6.6.3 Seasonal Trend of the Rate of Change

Now that we understand the methodology for modeling the relationship, let's inspect [Figure 6-28] which shows the slope for day and night over time for station 213 from Sep 2022 to Mar 2023 – winter data, using ambient and 2" subsurface temperatures. Understanding the pattern at this finer level, for one winter, will help us comprehend the bigger picture when presenting the data for more stations for the whole 27-month period. First, we notice that the rate of change for both day and night ranges between 0.45 and 0.65. Second, the regression slope, representing the rate of change, begins at a relatively high value in September ( $y = 41.40 + 0.62x$ ), reaches its peak in November ( $y = 22.19 + 0.64x$ ), undergoes a decline

until January, and subsequently starts an upward trajectory once more. Let us explain this behavior, we have three phases: an increase, a decrease, and a final increase. First, the slope peaks in November because that is when the colder weather hits, as the ambient temperature drops at a faster rate than the subsurface temperature – making the rate of change steeper. However, as winter progresses and both ambient and subsurface temperatures become more stable in their change, this will cause a decrease in the rate of change. The third phase exhibits another increase in the rate of change due to the weather warming up again and the ambient increasing at a faster rate than the subsurface temperature, thus, changing the rate of change (or relationship) again. Finally, we notice that the rate of change is higher at nighttime, this is because nights are usually colder causing rapid drops in ambient resulting in a more pronounced rate of change compared to daytime. In addition, we can see this in [Figure 6-29], where the monthly temperature distribution, for day and night, is shown as a violin plot for ambient, and subsurface at 2” and 6”. Looking at the ambient (left), we notice that in November, the cold weather hits as the distributions observe a noticeable shift down compared to October. Also, most data points are concentrated at lower degrees – the wider lower part of the violin. However, subsurface temperatures at both levels are catching up but at a slower rate, this results in a higher gap between ambient and underground temperatures and, therefore, a change in the rate of change. Moreover, we can see that the distributions stabilized for the next three months which is reflected as a decrease in the slope because the gap between ambient and subsurface temperatures is closing. And this pattern will repeat itself subsequently.

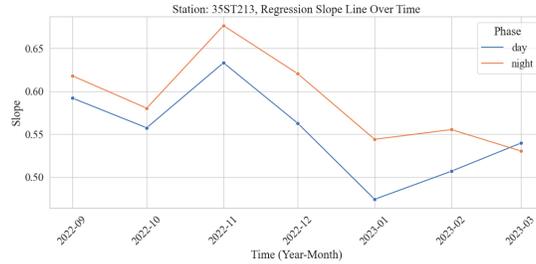


Figure 6-28 Monthly Rate of Change (Slope) for Daytime and Nighttime

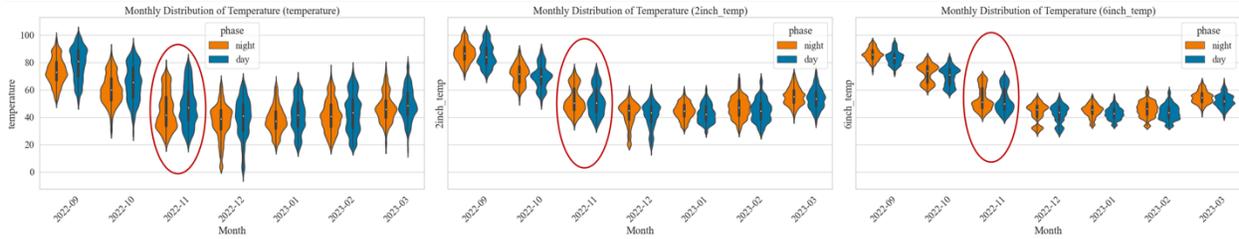


Figure 6-29 Monthly Temperature Distribution for Daytime and Nighttime, for Ambient, 2", and 6" Subsurface Probes

Having established an understanding of the winter rate of change characteristics, we can extend the analysis to additional stations and over extended periods. [Figure 6-30] and [Figure 6-31], show 27 months, from January 2021 to March 2023, of the rate of change (slope) for various RWIS weather stations on I-35 modeling the relationship between ambient and subsurface temperatures at both levels. Stations are in order from north to south as follows: 213, 199, 187, 165, 136, 124, 107, 092, 058, 051, and 032. It is important to note that these stations might be miles away from each other, for example, station 213 is near the Kansas border up north in Kay County while 032 is down south in Carter County. Therefore, the climate is expected to change as we observe stations from north to south gradually – northern stations will consequently observe colder weather.

For ambient and 2" subsurface temperature in [Figure 6-30], we notice that all stations follow the same yearly behavior. The rate of change begins relatively low in January, peaks in June, and then decreases gradually again. The rate of change ranges between 0.4 and 1.4 for all stations. Variations in the rate of change over time can be attributed to the pace at which ambient temperature fluctuates relative to subsurface temperature. Increases happen in hot summer months when the sun's

radiation causes the ambient temperature to increase at a much faster rate than the subsurface temperature which is colder since it preserves its temperature as this weather will take time to affect underground temperatures. The same is true for colder months when the ambient is reducing more quickly while the subsurface temperature is warmer. This is due to thermal inertia.

For ambient and 6” subsurface temperature in [Figure 6-31], we also notice that all stations follow a certain behavior that is different than the one observed at the 2” level. The pattern begins at a low rate in January, peaks 3 times in June, August, and October, and then finally decreases again. The reason for this difference is that first, the range is from 0.1 to 0.8 which is much smaller compared to 2” – which makes these peaks less influential. Notably the rate of change for the deeper probe is always, for all stations and all months, less than that of the shallower probe. Moreover, an important distinction, is at 6” the heat or cold holds over time, therefore, changes in ambient temperature do not catch up as much as at the 2” level.

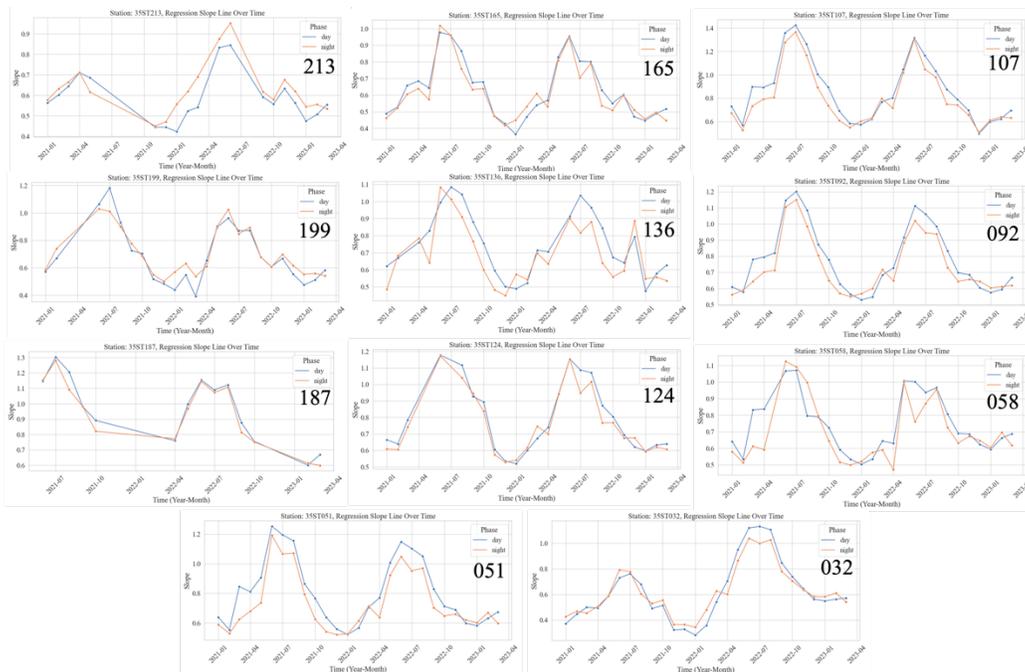


Figure 6-30 Monthly Behavioral Pattern of Rate of Change for Day/Night for Ambient and 2” Subsurface Probe for Various Stations

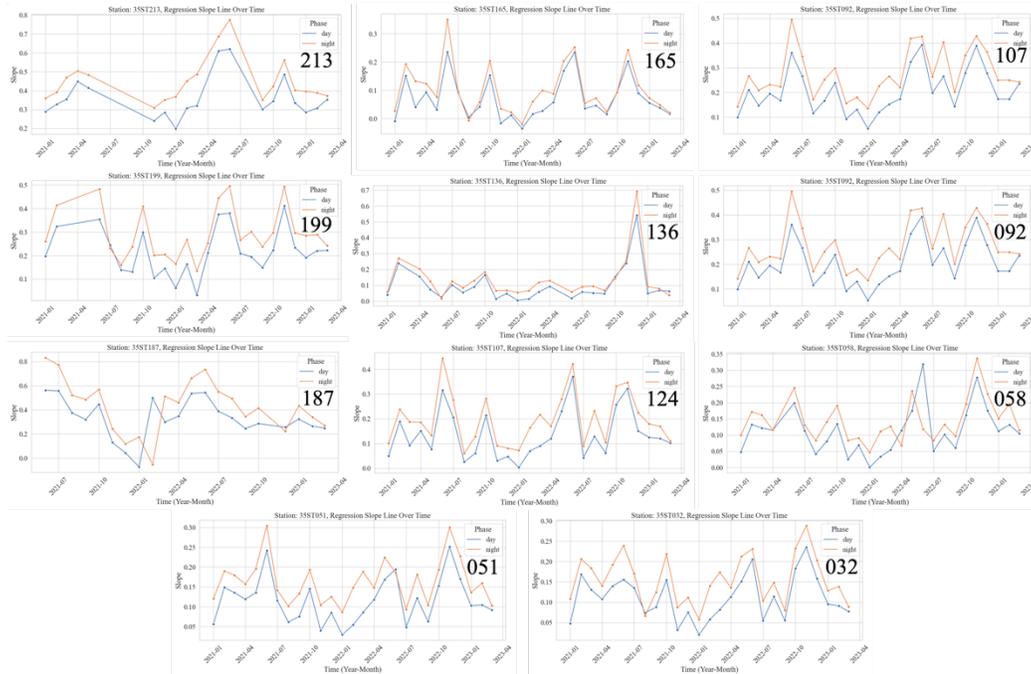


Figure 6-31 Monthly Behavioral Pattern of Rate of Change for Day/Night for Ambient and 6" Subsurface Probe for Various Stations

Further nuances can be also discussed. First, for winter, the rate of change between night and day seems to flip over time – the slope of daytime becomes higher. [Figure 6-32] shows an example of two stations 213 and 136, one station 213 had the slope for nighttime above daytime while the other station 136 exhibited a flipping behavior over time. To understand this, we examined various parameters over time including distribution statistics (mean, standard deviation, skewness, kurtosis, and the interquartile range – difference between third and first quartiles), these variables were not able to explain this flipping behavior. However, we found that the correlation coefficient  $r$  had a similar behavior, thus explaining the flipping. Therefore, the observed oscillations in the rate of change (slope) between daytime and nighttime are explained by the distribution shape. Another way to interpret this is it is affected by the distortion in distribution (how wide or thin the oval shape is, previously shown in [Figure 6-27]) measured by the correlation coefficient. [Figure 6-33] shows the monthly correlation coefficients for both stations. We notice that when the day

and night coefficients are close, the rate of change for nighttime is higher than that of the daytime. Whereas when we observe significant drops in correlation for the night, this causes the nighttime slopes to become lower than the day. Furthermore, lagged correlations at 1h, 2h, and 3h showed similar behavior as observed in [Figure 6-33] but with larger effects (wider gaps).

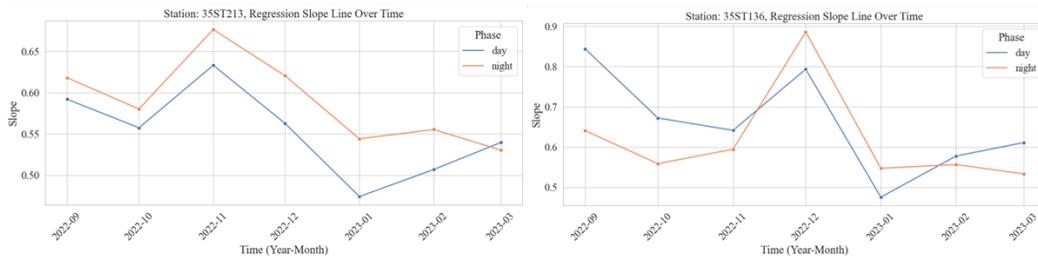


Figure 6-32 Monthly Rate of Change for Day/Night for 213 and 136

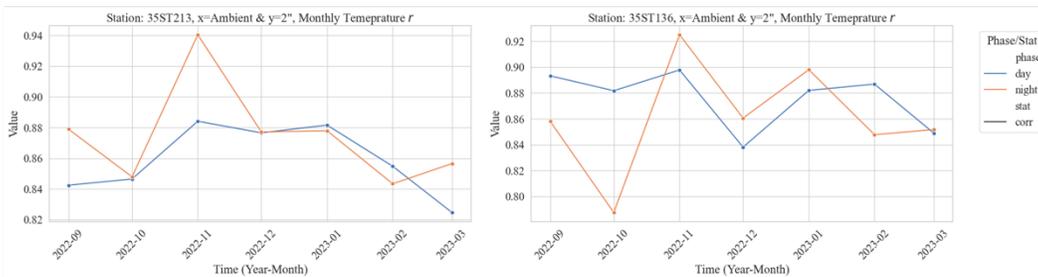


Figure 6-33 Monthly Correlation Coefficient for Day/Night for 213 and 136

The second nuance to mention is the pattern in the rate of change (slope) changes for southern stations compared to northern ones for the partial period from September 2022 to March 2023. In [Figure 6-30] showing the slope over time for ambient and 2” subsurface temperature, if we inspect the period from September 2022 to March, we notice that all northern stations down to 136 have the same behavior but this behavior changes at 124 as we go toward the south. For northern stations, the slope peaks around November for 213, 199, 187, 165, and December for 136 – the delay in 136 is explained by the delay in the cold weather reaching that region. Whereas the southern stations do not exhibit this increase and then decrease, instead, the rate of change decreases and then stabilizes. This can be attributed to the climate variations we mentioned earlier. Since these stations are several miles apart, they fall within distinct climatic zones.

## 6.6.4 Daily Patterns of the Rate of Change

Thus far, we have shown monthly analysis of the rate of change, however, it might be useful to look at a finer level, for example, the daily slope. [Figure 6-34] shows the daily rate of change (slope) for some stations from January 2021 to March 2023. The left two figures are for the slope mapping ambient to 2” subsurface temperature, where the above is during the daytime and the bottom plot is for the nighttime. Similarly, the right two figures are for the 6” subsurface probe. Expectedly, the daily rate of change will be noisier than the monthly equivalent, but we can still get useful information from it. The range in which the slope changes for probe 1 is from -0.5 to 2.5 while for probe 2 is from 0 to 0.6 which is a considerably smaller range. Additionally, if we examine the overall trend in which the slope is moving over time, we notice it is quite reminiscent of the monthly pattern that we showed earlier in [Figure 6-28] – that is there is an increase until June/July and then a gradual decrease.



Figure 6-34 Daily Rate of Change for Day/Night, for both Subsurface Probes for Various Stations

### 6.6.4.1 Rate of Change Distribution Analysis

To better understand and make use of the daily relationship, we created a distribution of the rate of change for day and night with bin width equal to 0.05, shown in [Figure 6-35] for winter data from September 2022 to March 2023 from station 213. The left graph contains histograms of the slope between ambient and 2” subsurface temperature while the right graph is for the 6” level. Comparing day and night in both plots, we notice that they are more distinguishable at 6” – where the night distribution has a larger standard deviation while the day distribution has higher kurtosis –more stable. A high kurtosis distribution (Leptokurtic) suggests that a great number of points are condensed in the middle, meaning that the rate of change remains relatively stable. This becomes clearer when examining the daily slopes in [Figure 6-36] – the left plot is for ambient and 2” whereas the right is for 6”. Notice that in both cases, the daytime daily slope (in blue) is significantly more consistent (or stable) than the nighttime slope (in orange).

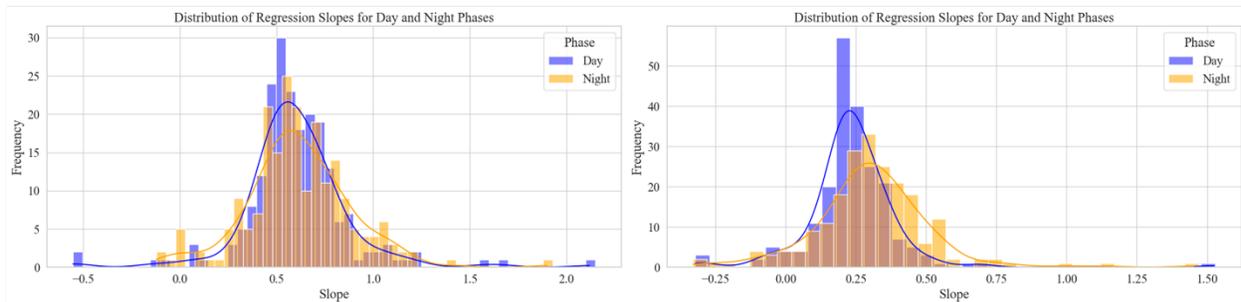


Figure 6-35 Distribution of Daily Rate of Change for Day/Night for 2” and 6” Subsurface Probes

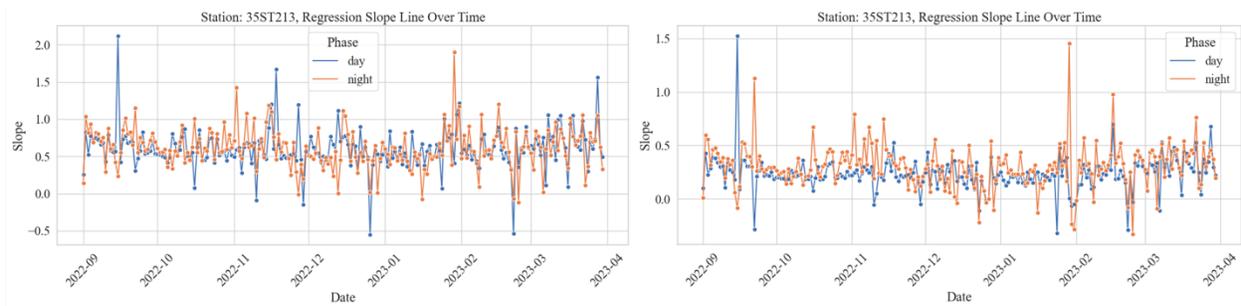


Figure 6-36 Daily Rate of Change for Day/Night for 2” and 6” Subsurface Probes

Another interesting thing to look at is the distribution of the rate of change for various months. The next two graphs show, for station 213 in winter from September 2022 to March 2023, monthly slope distributions for probe 1 at 2” in [Figure 6-37] and probe 2 and 6” in [Figure 6-37]. We notice that colder months have more variation and lower peaks but with higher centrality of the distribution. Whereas warmer months (e.g., 9 and 10) have lower distribution centrality, less variation, and a leptokurtic nature – suggesting more stability in the rate of change.

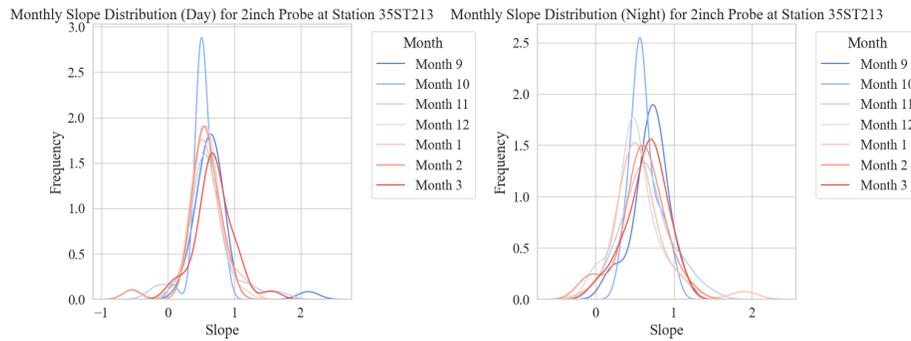


Figure 6-37 Monthly Distribution of Rate of Change for Day/Night for 2” Subsurface Probe

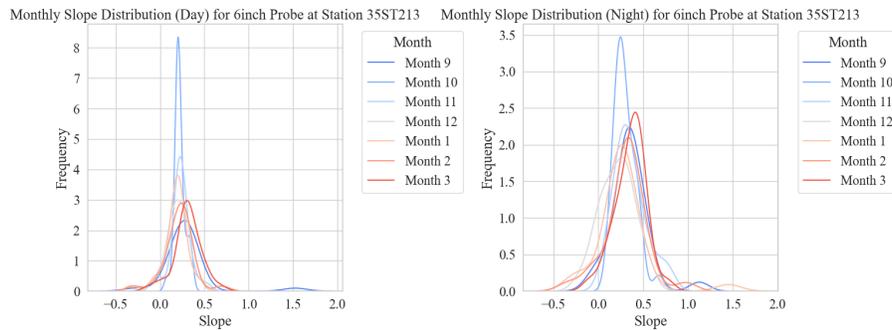


Figure 6-38 Monthly Distribution of Rate of Change for Day/Night for 6” Subsurface Probe

The following two graphs show the KDE estimate of the distribution of the daily slopes for the stations for both the first and the second probe. We notice that for the first probe, at 2”, the rate of change for the different stations is similar. However, slope distributions of southern stations (32, 51, 58, 92) tend to be more similar to each other than northern stations. This is expected as they are in areas experiencing different climates. Looking at the distributions for the 6” level, we notice that there exists a somewhat more notable disparity in the distributions. Northern stations seem to

have larger variation and more distributed overall distributions such as in 213, 199, and 165. Southern stations seem to have a more stable rate of change resulting in distributions with high peaks and less variation such as in 32, 51, and 58.

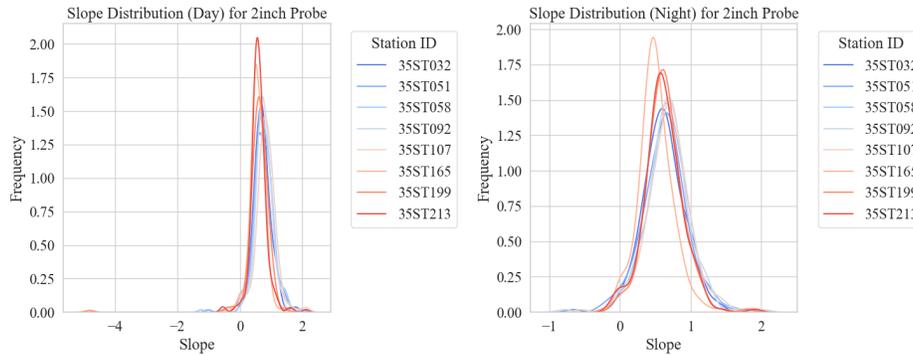


Figure 6-39 Spatial Distribution of Rate of Change for Day/Night for 2" Subsurface Probe

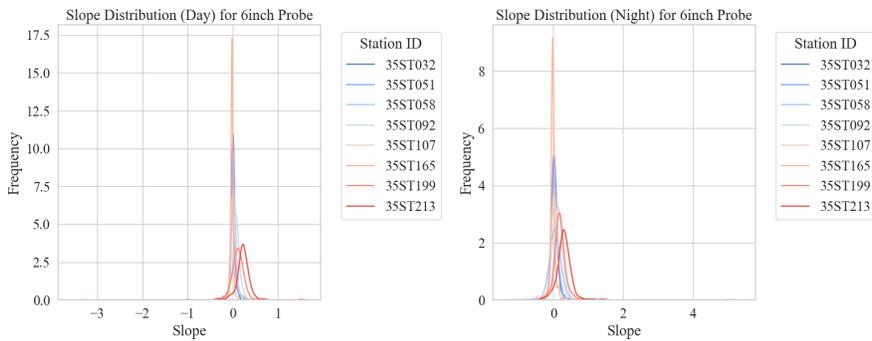


Figure 6-40 Spatial Distribution of Rate of Change for Day/Night for 6" Subsurface Probe

### 6.6.4.2 Wavelet Transform of the Daily Rate of Change

In our pursuit to study the rate of change over time to understand transitions and discern patterns in the relationship, we turn to apply the Wavelet Transform aiming to capture the dynamics of the daily rate of change. In [Figure 6-41], we show the wavelet scalogram for probe 1 (left) and probe 2 (right) at nighttime (upper) and daytime (lower) for station 213 for the daily slope from September 2022 to March 2023. This time-frequency graph is a result of applying the Continuous Wavelet Transform (CWT) using the Complex Morlet Wavelet (CMOR) to extract the coefficients and frequencies at 30 scales –to capture both high-frequency events of short duration and low-

frequency events extending over longer periods. Therefore, we are inspecting the different frequencies that the daily rate of change exhibits over time. The interpretation of frequency in this case is that higher frequency corresponds to frequent changes in the daily rate of change while lower frequency corresponds to more stable periods where the rate of change is either not changing or changing only slightly. We have shown the monthly rate of change for this station and period previously in [Figure 6-28], we noticed that the rate of change experienced high changes in November where peaked, and in January when it started increasing again. Now, looking at the scalogram for nighttime at 2'' (left upper plot) we notice that the wavelet transform can pick up on these changes in November and January from the daily signal. The daytime scalogram (left lower plot) is less pronounced as changes in the nighttime are stronger. For the second probe (right upper and lower plot), although changes in frequencies were picked up, but they were much less significant, as the deeper surface temperature takes more time to pick up on these changes. Additionally, we notice that the frequencies range between 0.1 to 0.5 Hz for both probes at day and night. In retrospect, we noticed that the wavelet transform was able to pick up on changes in the rate of change especially at night (stronger changes), however, we also applied the Short-Term Fourier Transform (STFT) but were not able to pick on these localized changes as the Wavelet Transform did.

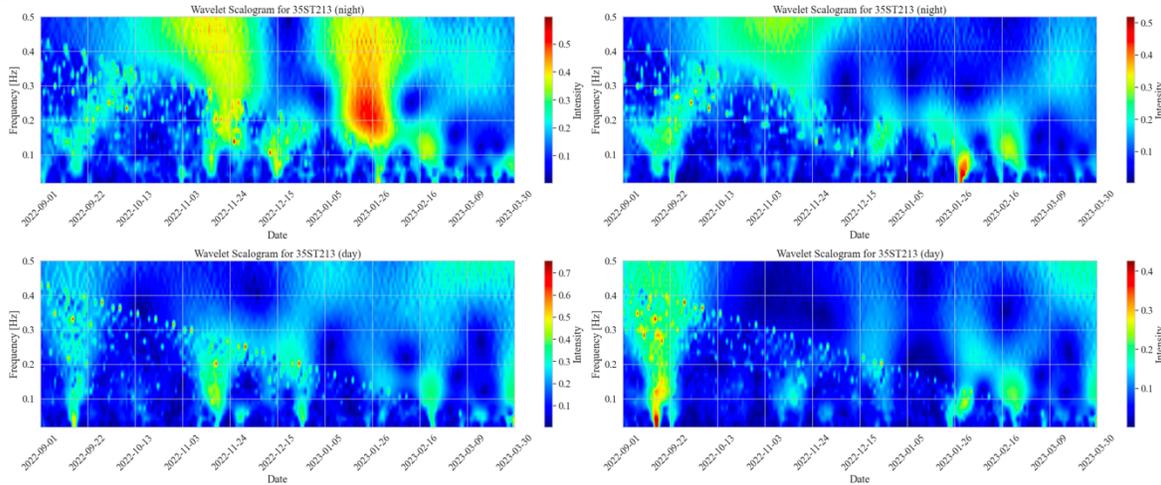


Figure 6-41 Wavelet Scalogram for Night/Day for 2", and 6" Subsurface Probes

## 6.7 Event-Based Analysis of Rate of Change

The next figure shows an example, from station 213, of a month that observed two storms, spanning the days 2 to 6 and 22 to 26. Due to the noticeable gap between ambient and subsurface temperature probes, the rate of change is expected to have different characteristics during such events compared to normal conditions.

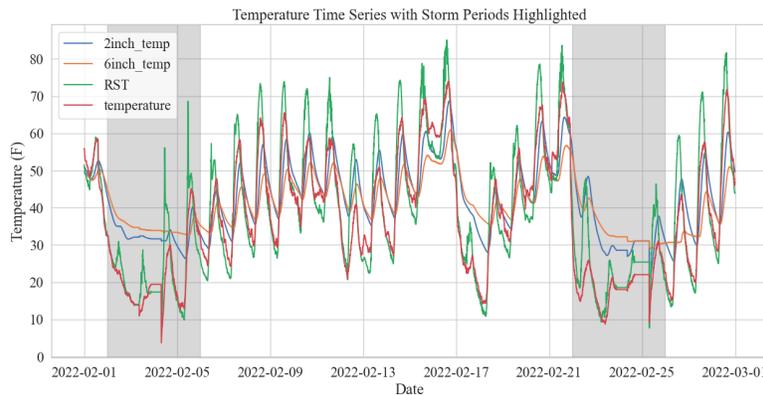


Figure 6-42 Temperature Progression Through Storms (shaded)

The following figure shows the unsegmented (left), and the segmented (right) for daytime and nighttime, slope distribution for 2", 6", and surface. We notice that storms generally observe higher variation in the distribution of the rate of change. Furthermore, the 6" storm slope distribution seems to be different for daytime, with similar to normal variation, and nighttime with a large

variation compared to normal – the variation at night for 6” accounting for the observed larger variation shown in the unsegmented graph.

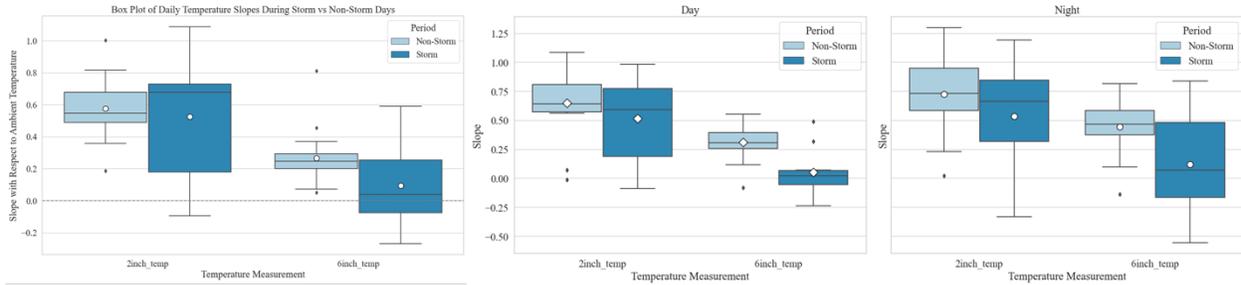


Figure 6-43 Slope Distribution for Event (Storm/Non-Storm): Unsegmented (left), and Segmented (right)

### 6.7.1 Spatial Analysis

The following graph shows, for 2” and 6”, the slope distribution per station for events – storms and normal. We notice the same behavior across stations, with higher variation during storms.

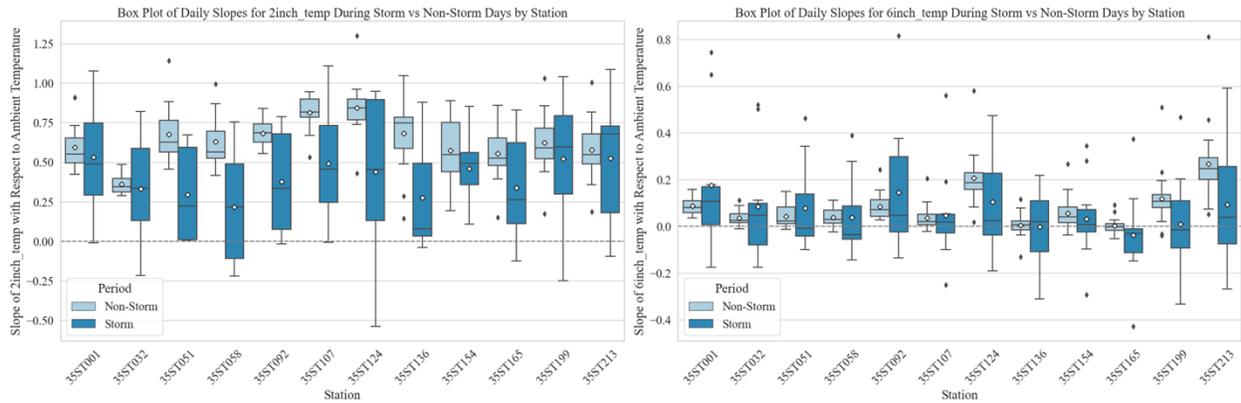


Figure 6-44 Slope Distribution per Station for 2” (left) and 6” (right)

Next, we show the segmented, for daytime and nighttime, slope distributions for 2” in [Figure 6-45], and 6” in [Figure 6-46], during storms. The results align with previous findings, nighttime storm distributions have larger variation compared to daytime for 6”, shown [Figure 6-43]. Whereas, for 2” shown in [Figure 6-45], both daytime and nighttime stormy distributions have larger variation compared to non-stormy. This is an important observation because we can then monitor the rate of change in real-time using a window (e.g., 5 days), and track the rate of change

for probes at day and night at both levels if we notice that variation in the rate of change is starting to increase especially for the second probe at night, this means that a storm is likely to come.

Box Plot of 2inch\_temp Slopes by Day/Night and Storm Period

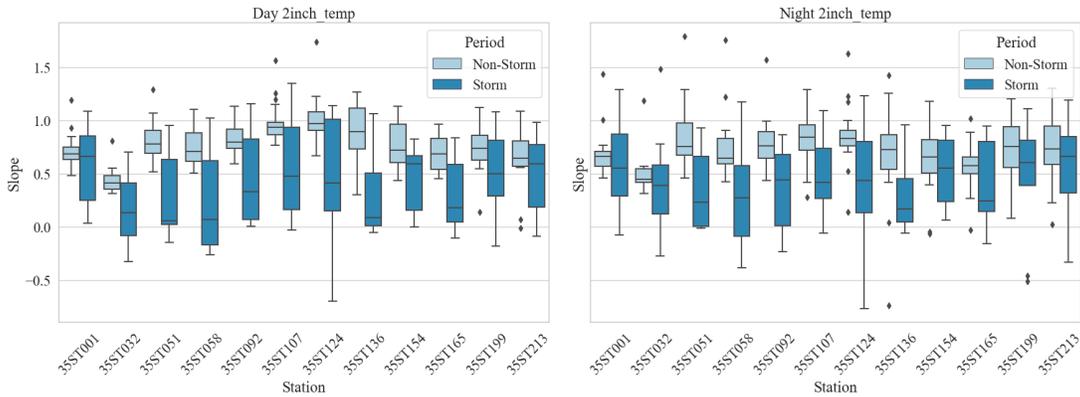


Figure 6-45 Segmented Slope Distribution per Station for 2" Probe

Box Plot of 6inch\_temp Slopes by Day/Night and Storm Period

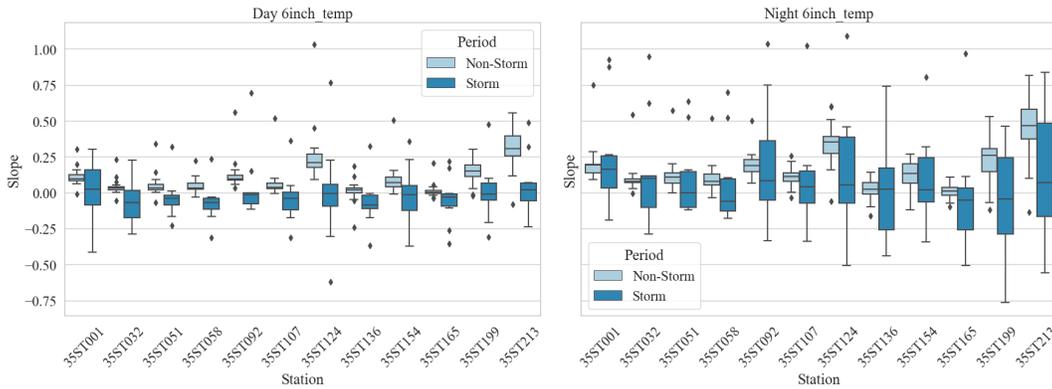


Figure 6-46 Segmented Slope Distribution per Station for 6" Probe

## 6.8 Summary

To summarize, this chapter provided an in-depth investigation of the importance of the subsurface temperature probes in RWIS. Additionally, we explored by parametric modeling the seasonal progression of the relationship between ambient and subsurface temperatures at both levels 2" and 6". Changes in the relationship were more significant at the 2" level compared to 6" due to thermal inertia and the propagation of colder or warmer weather through the air and into the deeper layers

under the surface. [Figure 6-30] and [Figure 6-30] characterized this parametric relationship using a linear regression slope over time for both daytime (i.e., 6 AM to 6 PM) and nighttime at both surface levels for several stations along I-35 exhibiting various climates. We noticed a distinctive temporal pattern in this relationship across the examined stations at both underground levels with ambient temperature. We have conducted an in-depth investigative analysis to understand the relationship over time by explaining the observed behaviors and patterns. We characterized the distribution of the rate of change for daytime and nighttime as well as for various months showing when the rate of change is stable and when it undergoes larger variations. Moreover, we applied the Wavelet Transform on the daily rate of change which was able to pick up on the distinctive changes in the relationship which aligned with our previous graph showing the monthly slopes for day and night over time. Finally, event-based analysis showed that during storms, the rate of change tends to vary more by exhibiting higher variation in the slope distribution for both probes. Notably, segmented distributions for daytime and nighttime showed that at 6", the rate of change variation is observed only during the night and not the day – this is not surprising since, during storms, most of the significant ambient temperature drops happen at night which affects the relationship with 6" subsurface temperature.

## 7 Conclusion and Future Work

The research presented in this thesis detailed a variety of analyses aimed at characterizing the dynamics of weather data to enhance our understanding and utilization of RWIS systems. We analyzed the rate of change, parametrically modeled by a regression slope, between ambient and subsurface temperature probes at two levels, two- and six-inches. The behavioral patterns were explored through seasonal progression into the colder months utilizing data spanning 27 months, from January 2021 to March 2023. We have seen the rate of change increasing during months that exhibit ambient temperature variation that is transitioned through subsurface layers at a slower rate – creating a gap between ambient and subsurface temperatures accounting for the steeper rate of change. Moreover, we observed that the rate of change for the deeper probe, at the six-inch level, always has a considerably lower magnitude compared to the two-inch subsurface probe. This is explained by heat preservation at the deeper level, where heat and cold take more time to propagate through layers starting from the atmosphere and transitioning into underground layers. The observed rate of change for the nighttime was mostly more pronounced than daytime. Moreover, storms were shown to significantly increase variation in the rate of change for day and night for the first probe whereas the second probe exhibited increased variation only at night. Machine learning model results were enhanced when subsurface temperature probes were incorporated – emphasizing their role in the prediction of surface conditions. Clearance time amidst snowstorms appears to stabilize traffic by reducing speed variability while having a lower subtle effect on the average speed, on a segment level. The experiments showed that utilizing weather data collected from RWIS stations and speed data collected from nearby interstate segments for machine learning models capable of correctly classifying road weather surface conditions with only up to 13% margin error. In conclusion, the analyses and experimentation presented in this thesis attempt to

pave the way for further advancement and better utilization of road and weather information system data and safer roads.

## **7.1 Future Work Remarks**

RWIS stations existed for years and making incremental advancements is critical to better utilize such systems to create safer roads, reduce traffic collisions, and reduce the effects of inclement weather. The work presented in this thesis builds on top of previous research and work exploring the potential of utilizing RWIS data in various applications. For example, [20] presented a comprehensive study focusing on building forecasting models of temperature data. Another study is [21], where RWIS data were utilized to enhance predictive models of catastrophic flash floods.

In retrospect, these systems can be highly useful for a spectrum of applications and research venues. In this section, we will try to highlight the benefits of better utilization of RWIS cameras capturing roadside and bridge side views by implementing computer vision and deep learning techniques for better detection of road surface conditions. Utilizing imagery to detect road weather conditions, as intriguing as it seems, is presented with many challenges. Challenges include images where multiple weather conditions occur together – weather can be wet while it is sunny as shown in [24]. Moreover, nighttime is particularly challenging as it is harder for these cameras to pick up on road details. Transitioning states between conditions (i.e., dry to wet) can be confusing. Icy conditions are hard to visually distinguish from dry conditions. The list goes on including but not limited to class imbalance, illumination, poor visibility, etc. Researchers have tried utilizing images to detect road conditions in a variety of ways. In [22], the authors framed road surface condition classification as an object detection problem. An object detector, Mask R-CNN, was used to extract the parts of the image that represent the road. The network architecture was adjusted

to include weather parameters along with images to make predictions. This can be implemented in RWIS in Oklahoma by utilizing subsurface probes and surface temperature sensors with images to develop better detectors. Another approach, shown in [23], is by utilizing a residual Convolutional Neural Network (CNN), where a Residual Network (ResNet) [25] backbone is used with pre-trained weights, and a few additional fully connected layers are fine-tuned on the specific road surface conditions of the training dataset. Furthermore, it is possible to utilize a vision transformer architecture with a self-attention mechanism as shown in [24] for road surface condition classification. As shown, there is a variety of approaches where imagery can be used for this task, however, what makes this task tricky to practically implement is the lack of a labeled image dataset. These studies mostly rely on manual labeling to build a labeled dataset suitable for supervised learning techniques.

Finally, let's provide some actionable suggestions on how these techniques can be utilized in ODOT RWIS. The first step is to build a dataset of images from various stations encompassing various weather conditions (i.e., snowy, icy, dry, etc.). Next, images should be labeled with the corresponding surface conditions either manually or by utilizing volunteering platforms where a human can choose the correct label. Once this dataset is ready, a variety of tasks can be explored such as real-time classifying road surface conditions or forecasting conditions for the next few hours. Potential experiments can include incorporating weather, speed, and image data into predictions for surface conditions, possible stormy events and inclement weather, traffic congestions, and more. Other possible applications of RWIS images can include the detection of potential roadwork, potholes, pavement issues, and so on.

## 8 References

- [1] Sharma, D., Wu, M., & Kwon, T. J. (2021). Safety Effects of Road Weather Information System (RWIS)-A Cost-Benefit Analysis (No. TRBAM-21-01521).
- [2] Rutz, J. J., & Gibson, C. V. (2013). Integration of a road surface model into NWS operations. *Bulletin of the American Meteorological Society*, 94(10), 1495-1500.
- [3] Jonsson, P. (2011, September). Classification of road conditions: From camera images and weather data. In *2011 IEEE International Conference on Computational Intelligence for Measurement Systems and Applications (CIMSA) Proceedings* (pp. 1-6). IEEE.
- [4] Kršmanc, R., Slak, A. Š., & Demšar, J. (2013). Statistical approach for forecasting road surface temperature. *Meteorological applications*, 20(4), 439-446.
- [5] Andreescu, M. P., & Frost, D. B. (1998). Weather and traffic accidents in Montreal, Canada. *Climate research*, 9(3), 225-230.
- [6] Dey, K. C., Mishra, A., & Chowdhury, M. (2014). Potential of intelligent transportation systems in mitigating adverse weather impacts on road mobility: A review. *IEEE Transactions on Intelligent Transportation Systems*, 16(3), 1107-1119.
- [7] Veneziano, D., Ye, Z., & Turnbull, I. (2014). Speed impacts of an icy curve warning system. *IET Intelligent Transport Systems*, 8(2), 93-101.
- [8] Linton, M. A., & Fu, L. (2016). Connected vehicle solution for winter road surface condition monitoring. *Transportation Research Record*, 2551(1), 62-72.
- [9] Eom, T., Kwon, Y., & Ko, H. (2021). Compact weather sensor node for predicting road surface temperature. *Sensors and Materials*, 33(8), 2763-2771.

- [10] Thies CLIMA radar weather station.  
<https://www.thiesclima.com/en/Products/Miscellaneous-Devices-Weather-stations/?art=197>
- [11] Surface Sentinel Temperature Sensor – Model 5439-00.  
<https://hsierra.com/product/surface-temperature-sensor-model-5439-00/>
- [12] Campbell Scientific temperature probe.  
<https://www.campbellsci.com/109#:~:text=The%20109%20is%20a%20rugged,or%20submerge%20it%20in%20water>
- [13] Roadside Embedded Extensible Computing Equipment (REECE).  
<https://itrafficsystems.com/index.html>
- [14] Sierra RV50X industrial LTE Router.  
<https://www.sierrawireless.com/router-solutions/rv50/>
- [15] Weather Impacts on Mobility by the Federal Highway Administration (FHWA):  
[https://ops.fhwa.dot.gov/weather/q1\\_roadimpact.htm#:~:text=Weather%20Impacts%20on%20Mobility&text=On%20freeways%2C%20light%20rain%20or,by%203%20to%2016%20percent.](https://ops.fhwa.dot.gov/weather/q1_roadimpact.htm#:~:text=Weather%20Impacts%20on%20Mobility&text=On%20freeways%2C%20light%20rain%20or,by%203%20to%2016%20percent.)
- [16] The Temperature of The Earth’s Interior  
<https://www.scientificamerican.com/article/the-temperature-of-the-earths-inter/#:~:text=The%20ground%20is%20cooler%20in,from%20rapid%20changes%20of%20temperature>
- [17] The Oklahoma Department of Transportation, Field Divisions with County Names  
<https://www.odot.org/stateco.htm>
- [18] How Ice Forms On Roads  
<https://snowicesalt.com/how-ice-forms-on->

[roads/#:~:text=Freezing%20rain%20is%20particularly%20dangerous,rapidly%2C%20leading%20to%20ice%20formation.](#)

- [19] What is a Road Weather Information System (RWIS) by the FHWA:  
<https://ops.fhwa.dot.gov/weather/faq.htm>
  
- [20] Afify, M. (2021). RWIS based road condition prediction using machine learning algorithms.
  
- [21] Mousselli, A. (2022). Investigating Technologies and Techniques for Flood Monitoring and Detecting.
  
- [22] You, J. (2019, December). Weather data integrated mask R-CNN for automatic road surface condition monitoring. In 2019 IEEE Visual Communications and Image Processing (VCIP) (pp. 1-4). IEEE.
  
- [23] Pan, G., Muresan, M., Yu, R., & Fu, L. (2021). Real-time winter road surface condition monitoring using an improved residual CNN. *Canadian Journal of Civil Engineering*, 48(9), 1215-1222.
  
- [24] Samo, M., Mafeni Mase, J. M., & Figueredo, G. (2023). Deep learning with attention mechanisms for road weather detection. *Sensors*, 23(2), 798.
  
- [25] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).