Exploring NOAA Precipitation Types and Temperature Profiles: An Analysis of 18 Years of Weather Data in the United States

CSI 2300 Final Project

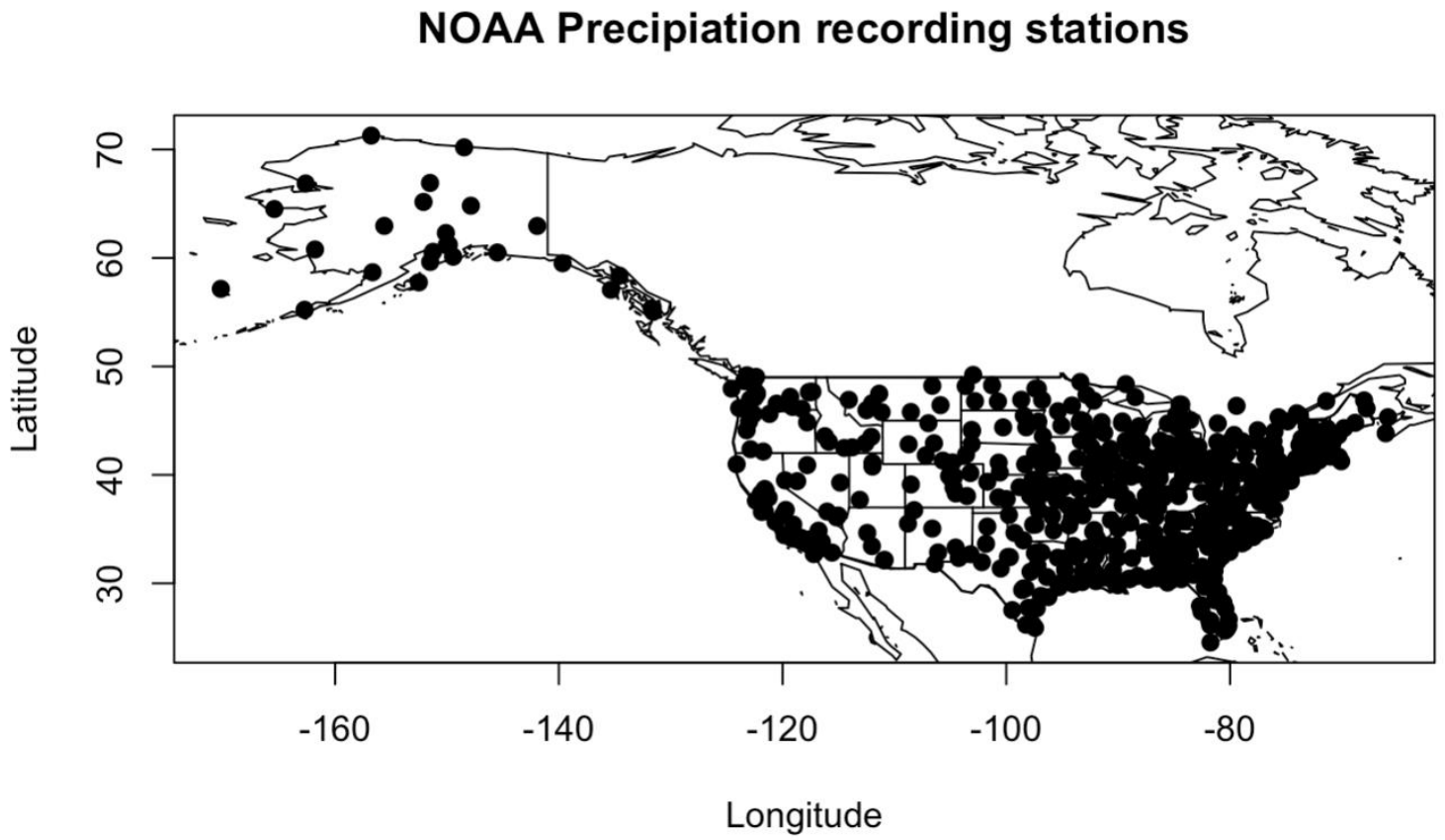
Conner Stallman, Ken Mitchell, and Landon Bragg

Abstract: Our dataset, NOAA Precipitation types, provides data that spans 18 years and focuses on the precipitation events of non-summer months. The data records precipitation events at various weather stations and categorizes them into either Rain, Snow, Ice Pellets, or Freezing rain. The observations that are recorded are correlated with altitude temperature profiles from corresponding grid boxes in a numerical weather prediction model. This model allows us to understand how temperature influences precipitation type. The project aims to identify outliers and improve classification of precipitation types using temperature profiles. A way we are finding outliers is by finding data points that we know should not exist but do exist. In this report, we are looking at visibly outlying temperature profiles of ice pellets and looking at precipitation classified as snow that is measured above the freezing point.

Introduction

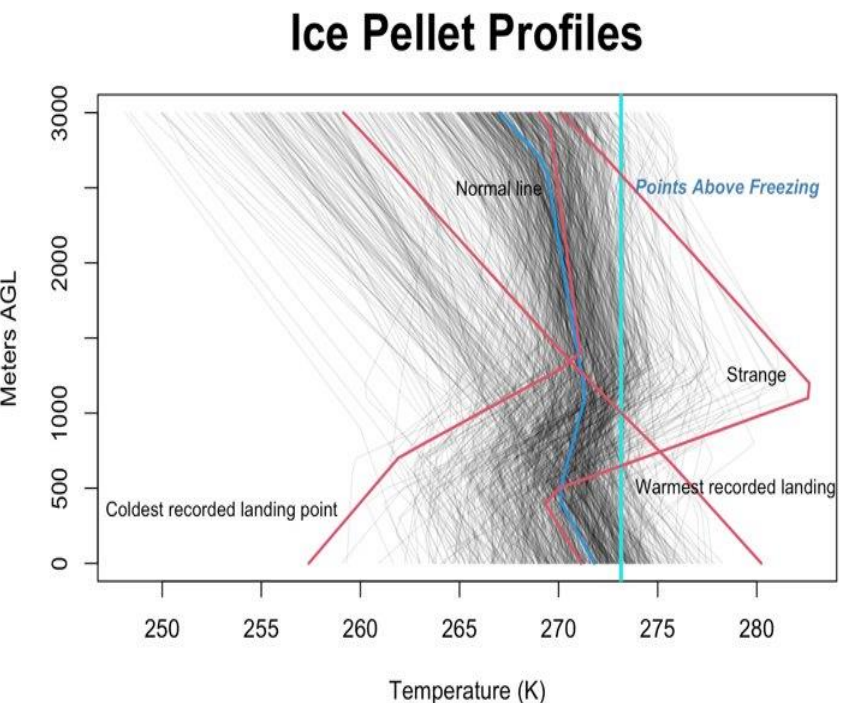
- Weather patterns, especially precipitation events, play a crucial role in various aspects of our daily lives, impacting agriculture, transportation, and public safety. Understanding and accurately predicting these patterns are essential for informed decision-making and risk mitigation. In this context, our project focuses on analyzing a comprehensive dataset provided by the National Oceanic and Atmospheric Administration (NOAA), specifically targeting precipitation events during non-summer months over 18 years
- The NOAA Precipitation Types dataset records precipitation events at different weather stations, categorizing them into Rain, Snow, Ice Pellets, or Freezing Rain. Our objective is to delve into the relationship between temperature profiles and precipitation types, seeking to identify outliers and enhance the classification accuracy of these events.
- The project primarily addresses two noteworthy anomalies within the dataset: discrepancies in temperature profiles of ice pellets and instances where snow is recorded at temperatures well above freezing. These anomalies raise concerns about the accuracy of existing weather prediction models and their real-world implications for sectors that rely heavily on precise weather forecasts.

Map of NOAA precipitation recording stations in the U.S.



First topic of interest: Ice Pellets

- To address the issue of ice pellet temperature discrepancies, we embarked on a systematic exploration of the dataset, creating visualizations and tables to highlight outliers. Despite extensive efforts, our findings indicate that the outliers in ice pellet temperature profiles appear to lack any discernible correlation with specific recording stations or days, posing challenges in predicting or addressing these anomalies effectively.



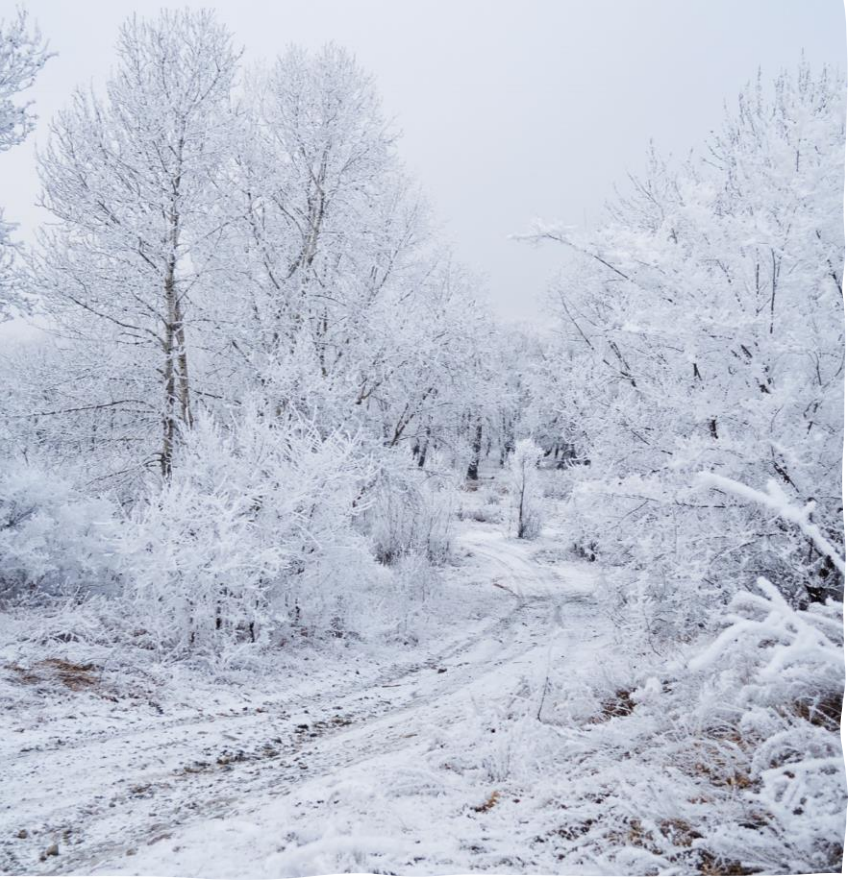
Warmest 10 Pellets by station

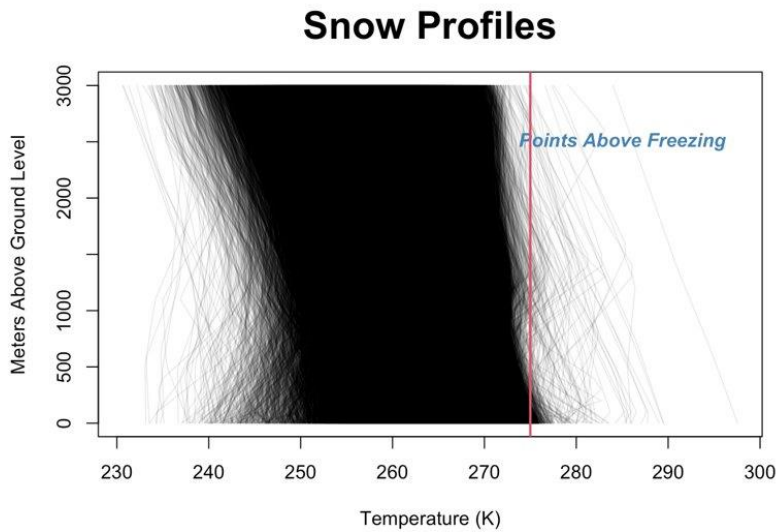
Twb.prof	station.ind
280.230246284025	536
278.296556774507	546
277.856475964943	271
277.749196500468	369
277.624964699544	537
277.451665320035	369
277.449372969745	541
277.175008996423	541
277.085710295741	369
277.073260463577	8

Coldest 10 Pellets by station

Twb.prof	station.ind
257.395679907458	488
259.061543642577	55
259.311968982747	489
260.926573838878	20
261.647267866152	473
262.230975446935	363
262.394640426063	363
262.6362322661	444
262.945971347851	495
263.00632956661	494

- We start by noticing some strange data points in the ice pellet profile graph such as 2 data points that have a 23 degree kelvin difference in temperature. Because of these strange outliers we need to take a closer look. We find that by observing the warmest recorded ice pellets and the coldest ice pellets that there does not seem to be a correlation between the station or day that the recording was made and the outliers that exist. This leads us to believe that outliers when it comes to ice pellets are simply inevitable and near impossible to predict or address because there is no evident correlation visible. Going forward it would be helpful to have more data points so we can hopefully observe a more obvious trend.



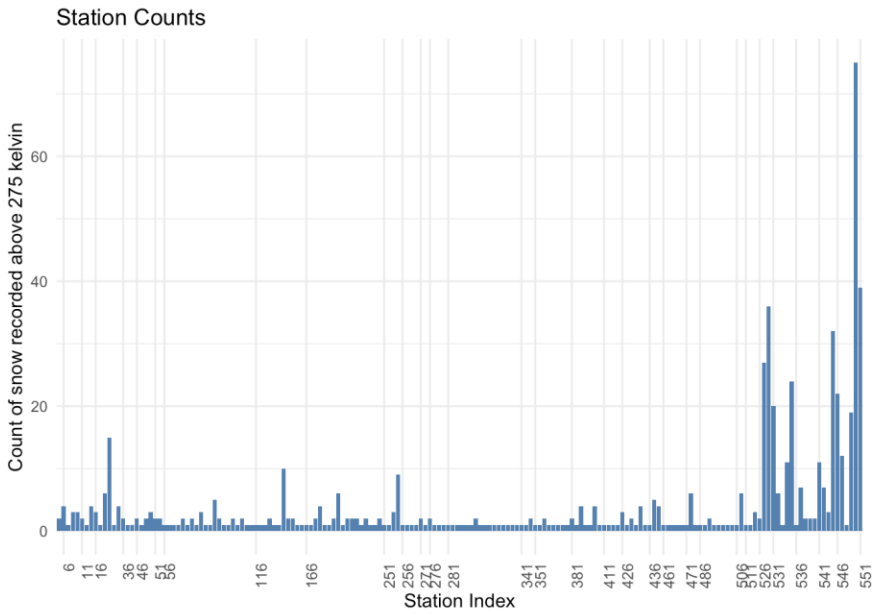


Second topic of interest: Snow profiles

- Next, we observed instances where snow was recorded at temperatures above the freezing point, prompting an investigation into the recording stations contributing to these inaccuracies. Our analysis revealed geographical clusters of problematic stations, suggesting potential regional influences on data accuracy and emphasizing the need for localized investigations into climatic variations and data collection protocols.

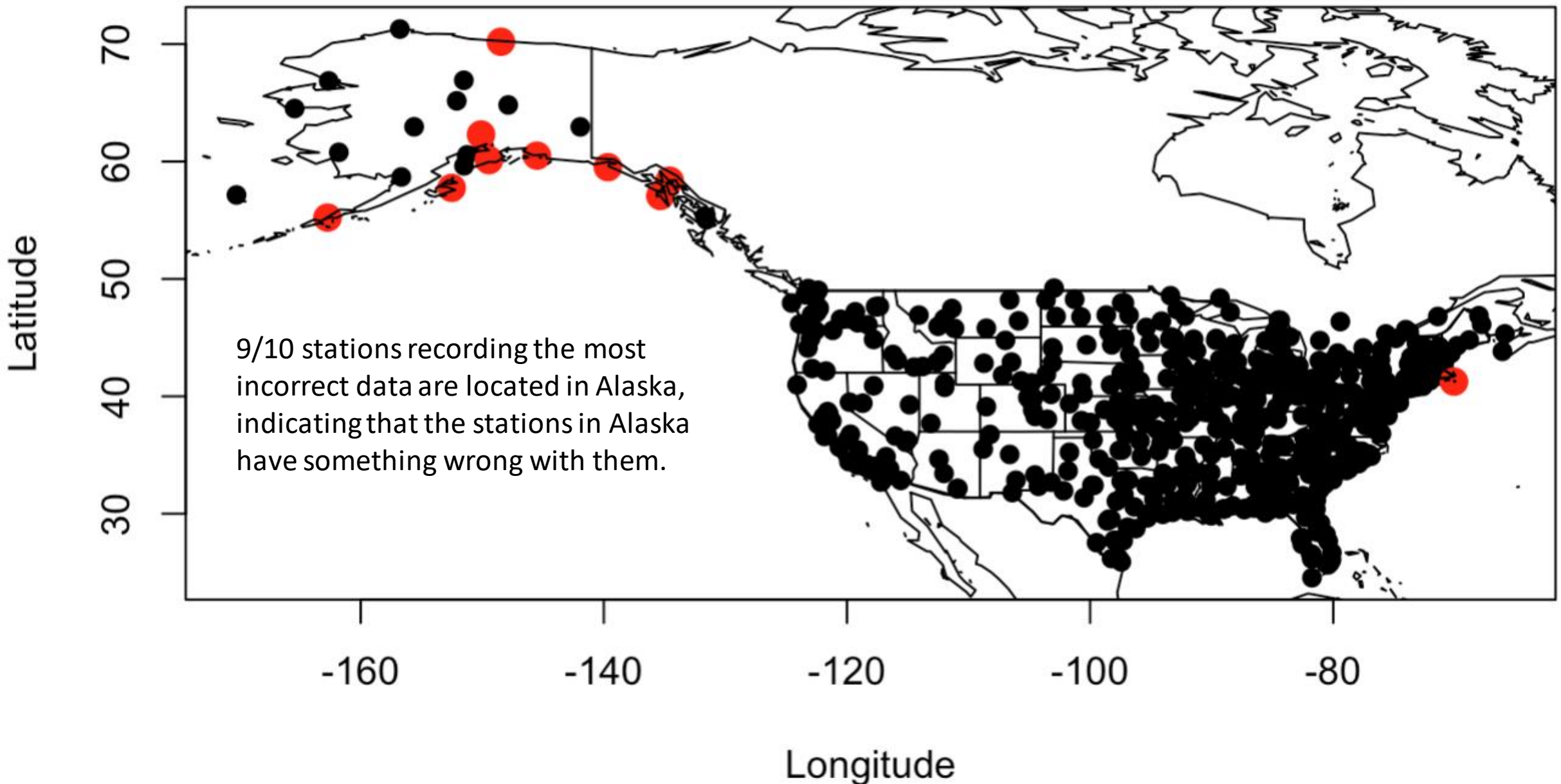
Top Station Indexes

Top Stations		
station.ind	count	index
550	75	1
551	39	2
530	36	3
545	32	4
529	27	5
535	24	6
546	22	7
531	20	8
549	19	9
25	15	10



•By summing each data point where the condition of snow being measured above 32 degrees occurred and contributing it to the station in which the incorrect data point occurred, we can observe which recording stations may be flawed or broken. As opposed to the previous data, we were able to find the roots of our outliers/problematic data points. We found that the majority of our outliers were being recorded at 10 stations. Because of that, we should investigate why there are so many of these impossible outliers occurring at those stations.

NOAA Precipiation recording stations



Conclusion

The findings of our project hold significant implications for weather forecasting models and the understanding of precipitation patterns during non-summer months. The identification of outliers in ice pellets and snow data raises questions about the accuracy of existing models and prompts considerations for improving public safety and efficiency in sectors like agriculture and transportation. The geographical disparities in outlier occurrences suggest potential regional influences on data accuracy, urging further investigation into local climatic variations. The impact on safety measures, particularly in regions prone to mistaken outlier recordings of snow, underscores the practical applications of our analysis. Moreover, the existence of outliers prompts a critical examination of data quality and collection protocols at recording stations, emphasizing the need for standardized practices.