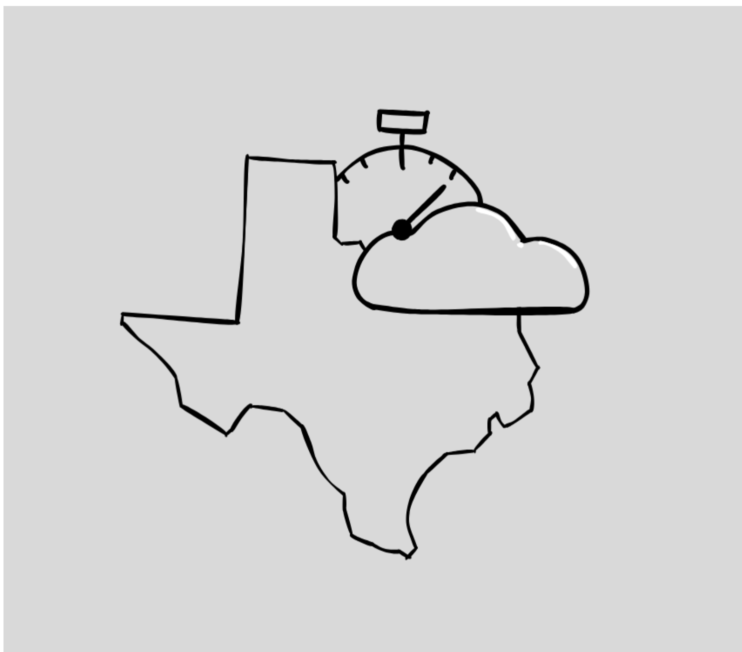
DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
THE UNIVERSITY OF TEXAS AT ARLINGTON

PROJECT CHARTER
CSE 4316: SENIOR DESIGN I
SPRING 2025



FORECAST TX
SEVERE WEATHER PREDICTION MODELING

NOE SANCHEZ
KEVIN SIMBAKWIRA
KRISTAL PHOMMALAY
KENIL PATEL
MASON BERRY

Revision History

Revision	Date	Author(s)	Description
0.1	2.17.2025	NS, KS, KP, KP, MB	Document creation
1.0	7.14.2025	NS, KS, KP, KP, MB	Updated for Closeout

Contents

List of Figures

1 Problem Statement

The increasing frequency and severity of extreme weather events in Texas, such as hailstorms, high winds, and flooding, pose significant threats to lives, property, and local economies. As climate patterns become more unpredictable, communities face greater challenges in preparing for and responding to these events. Despite advancements in technology, there remains a gap in accessible, region-specific predictive tools that could help individuals, businesses, and governments better anticipate and mitigate the impact of severe weather.

Currently, many insurance providers and emergency management agencies rely on historical data and broad weather forecasts that lack the precision needed for localized decision-making. This results in reactive measures rather than proactive strategies, contributing to higher economic losses and reduced response effectiveness. Furthermore, homeowners and businesses are often left unprepared for increasingly frequent and severe storms, leading to property damage, financial hardship, and, in some cases, injury or loss of life.

A predictive modeling tool capable of forecasting the frequency and severity of severe weather events in Texas over the next 5 to 10 years would address this gap by providing actionable insights to stakeholders. Such a tool could enable more effective disaster preparedness, support data-driven insurance policy adjustments, and help local governments implement early warning systems. Ultimately, improving the ability to predict and prepare for severe weather events could reduce financial losses, enhance public safety, and save lives.

2 Methodology

The team will develop a predictive weather model to forecast severe weather events in Texas, focusing on one peril (e.g., wind, hail, or rain) in a region with frequent occurrences. The project aims to leverage historical weather data from sources such as NOAA, which provides records from January 1950 to October 2024, ensuring a robust dataset for model training and validation.

The predictive model will utilize machine learning algorithms, including ARIMA, SARIMA, Random Forest, LSTM, or others to forecast both the frequency and severity of events over the next 5 to 10 years. The team will conduct feature engineering using historical weather patterns, geographical data, and other relevant variables to enhance model accuracy.

The solution will include a cloud-based computing environment providing the necessary resources for model training and deployment. Data storage will be managed using a relational or NoSQL database to house historical records, predictions, and user-generated queries. The data pipeline will include automated ETL (Extract, Transform, Load) processes to ingest, clean, and prepare data from NOAA and other sources.

A web-based interface will be developed to facilitate user interaction, allowing users to input parameters, view predictions, and filter results by year, peril, or location. The interface will feature interactive dashboards to visualize predictions and historical trends. Additionally, RESTful APIs will be implemented to enable users to retrieve data and predictions programmatically in formats such as CSV or JSON.

To ensure security and compliance, user authentication and authorization protocols will be implemented to protect sensitive data. The team will also conduct usability testing with stakeholders to refine the web interface and cross-validation techniques to assess model accuracy.

3 Value Proposition

As a leading provider of home and auto insurance, State Farm faces increasing risks and financial burdens due to the rising frequency and severity of severe weather events. Our predictive model will provide data-driven insights that can enhance their risk assessment, underwriting processes, and loss mitigation strategies.

Improved Risk Assessment and Policy Pricing
By forecasting future severe weather events in specific Texas regions, State Farm can adjust insurance policy pricing more accurately. The model helps assess which areas are more prone to hail, wind, or flooding, enabling better risk segmentation.

Claims Reduction through Proactive Measures
With access to predictive data, State Farm can notify policyholders about potential future risks, encouraging preventive actions such as reinforcing roofs, securing property, or purchasing additional coverage. This proactive approach can reduce the volume of claims and lower financial losses.

Enhanced Decision-Making with Data Analytics
The system will store historical and predictive weather data, offering insights into long-term climate trends. This enables State Farm's actuaries and data scientists to refine their catastrophe modeling and reinsurance strategies.

Competitive Advantage in the Insurance Market
By integrating machine learning-based forecasting into their operations, State Farm can position itself as an industry leader in leveraging AI for risk management, differentiating itself from competitors.

4 Development Milestones

- Project Charter first draft - 17 Feb 2025
- System Requirements Specification - 17 Mar 2025
- Architectural Design Specification - 07 Mar 2025
- Demonstration of Data Pipeline - March 2025
- Demonstration of Machine Learning Algorithm Output - April 2025
- Detailed Design Specification - June 2025
- Demonstration of User Interface - June 2025
- Demonstration of User interaction with model through UI - July 2025
- Final Project Demonstration - 1 Aug 2025

5 BACKGROUND

The insurance industry is experiencing unprecedented challenges due to the increasing frequency and severity of extreme weather events, particularly in high-risk regions like Texas. State Farm, one of the nations leading insurance providers, has identified a critical need for more accurate long-term weather prediction capabilities to better serve their customers and manage risk. While current meteorological services can provide accurate short-term forecast, there's a significant gap in the ability to predict weather patterns years into the future, a capability that could revolutionize how insurance companies approach risk assessment and policy management.

State Farm has partnered with our senior design team to develop a sophisticated weather prediction model that can forecast weather patterns up to 10 years into the future by analyzing 25 years of historical data. This partnership stems from State Farm's direct experience with the increasing frequency of severe weather events in Texas, including hail, wind damage, flooding, and other natural disasters. These events have led to rising insurance rates and, in some areas, questions about long-term insurability - a concerning trend for both the company and its policyholders.

The status quo of weather prediction primarily focuses on short-term forecasting, leaving insurance companies to rely on historical data and broad climate trends for long-term planning. This approach becomes increasingly inadequate as weather patterns shift and extreme events become more frequent. State Farm recognizes that better predictive capabilities could transform their ability to:

- Assess and price risk more accurately across different geographical areas.
- Develop proactive mitigation strategies for high-risk regions.
- Guide policy decisions regarding coverage in various areas of Texas.
- Help customers prepare for and potentially prevent weather-related losses.
- Most critically, contribute to saving lives through better preparation and planning.

The develop team, consisting of senior engineering students specializing in computer science, is working directly with State Farm's Teams. This collaboration provides access to extensive historical data that create a unique opportunity to correlate weather events wither their actual impact on communities and infrastructure.

The project's ultimate goal aligns with State Farm mission to protect and support their customers: by better predicting where and when sever weather events might occur, the company can work proactively to protect both lives and property. This could include developing new insurance products, adjusting coverage strategies, and working with communities on preparedness initiatives. The potential impact extends beyond just insurance rates and coverage - it represents a fundamental shift towards using predictive technology to create more resilient communities in the face of increasingly sever weather patterns.

6 Related Work

GenCast: GenCast is a Google DeepMind developed AI model that forecasts weather up to 15 days ahead. "GenCast marks a critical advance in AI-based weather prediction that builds on our previous weather model. A GenCast forecast comprises an ensemble of 50 or more predictions, each representing a possible weather trajectory. (GenCast) is adapted to the spherical geometry of the Earth, and learns to accurately generate the complex probability distribution of future weather scenarios when given the most recent state of the weather as input" [3]. Google has made the source code and weights for GenCast open source in order to promote growth in the industry of AI based weather forecasting.

FourCastNet: A collaborative project built by Nvidia, CalTech and other contributors. "FourCastNet (short for "Fourier ForeCasting Neural Network") is capable of delivering an accurate 1 week weather forecast in under 2 seconds. This is significantly faster than ECMWF's Integrated Forecasting System with better or comparable accuracy" [2]. FourCastNet is open source and documentation for reproduction of FourCastNet outputs is made available in NVIDIA's Modulus framework site.

National Service Storm Laboratory MRMS: The NSSL develops a variety of weather forecasting, predicting and modeling technologies to enhance fields of weather related research. One of their systems, MRMS (Multi-Radar/Multi-Sensor) "was developed to produce severe weather, transportation, and precipitation products for improved decision-making capability to improve hazardous weather forecasts and warnings, along with hydrology, aviation, and numerical weather prediction" [5]. Current works among this project include "Evaluating and advancing machine learning and artificial intelligence techniques along with product development within a cloud computing environment" [5] among other strategies to enhance storm forecasting products.

ENSO forecasting with ML: Ankur Mahesh (UC Berkeley) provides a research/lecture notebook along with accompanying source code for adapting ML models to forecast changes in the ENSO (El Nino/ Southern Oscillation) climate pattern [4]. El Nino is a climate pattern characterized by unusually warm ocean surface temperatures in the central and eastern tropical Pacific Ocean. This climate pattern has a significant impact on resulting weather in Texas. Being able to use ML to identify changes and trends in this climate pattern could be key in recognizing changes in severe weather activity here in Texas.

ClimateLearn: A research/lecture notebook on "Machine Learning for Predicting Weather and Climate Extremes" which provides code and examples for setting up models as well as instructions for accessing data through API calls to ERA5 data. "Climate change has led to a rapid increase in the occurrence of extreme weather events globally, including floods, droughts, and wildfires. In the longer term, some regions will experience aridification while others will risk sinking due to rising sea levels. In this tutorial, we aim to introduce the participants to machine learning approaches for addressing two fundamental challenges: Predicting climate variables into the future and spatial down scaling" [1].

GenCast and FourCastNet, while powerful, may not be the best fit for our sponsor due to being short/mid range models for weather. MRMS is heavily research based and pursues many alternatives for weather forecasting. Additionally, the project is part of a national organization which may not be a great fit for our sponsor. The ENSO and ClimateLearn forecasting research notebooks are promising because they seek to provide long term climate change trends which may be useful in predicting long term severe weather as requested by our sponsor. These research notebooks are educational and would require company resources to pursue research and development of their own tools.

7 System Overview

- **Historical Weather Data:** Sources include ERA5 reanalysis single-level datasets and NOAA storm event archives. These datasets provide high-quality, long-term weather information used to train and validate machine learning models.
- **Data Ingestion and ETL Layer:** Responsible for collecting, cleaning, and transforming raw weather data into a usable format for modeling and storage.
- **Data Storage:** Serves as the central repository for preprocessed data, model outputs, and historical trends.
- **Predictive Modeling:** Uses machine learning algorithms to forecast the frequency and severity of severe weather events based on historical patterns.
- **API and Integration Layer:** Provides a secure RESTful interface to access historical and forecast data for visualization and external use.
- **User Interface:** A web-based dashboard that allows users to interact with the model, visualize forecasts, and download data.
- **Security:** Implements user authentication for the web interface and access control within the GCP cloud environment. Security also covers backend infrastructure used by the project team for managing data and services.

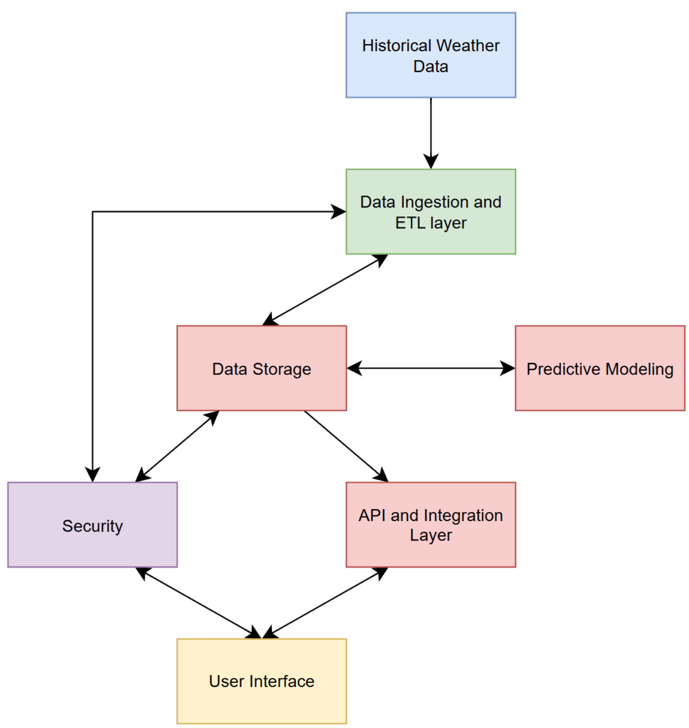


Figure 1: System Overview Diagram

8 Roles & Responsibilities

Stakeholders: State Farm

- **Amy and Dawsen** - Stakeholders from State Farm who are involved in the project and provide key insights and feedback to guide development.

Primary Sponsor Communication Leads:
Noe Sanchez and Mason Berry share responsibility for managing communication with project sponsors. They serve as liaisons between the team and stakeholders, ensuring alignment on goals, requirements, and deliverables.

Team Members:

- **Kenil Arvind Patel** - Full Stack Engineer
- **Kristal Phommalay** - Full Stack Engineer
- **Kevin Simbakwira** - Full Stack Engineer
- **Mason Berry** - Full Stack Engineer
- **Noe Sanchez** - Full Stack Engineer

Scrum Master: Rotating Role The Scrum Master role will rotate among team members to ensure everyone gains experience in this position. The Scrum Master will facilitate the agile process, manage meetings such as sprint planning, daily stand-ups, and sprint reviews, and help maintain workflow. The focus will be on ensuring the team follows agile practices, resolving impediments, and fostering collaboration.

9 Cost Proposal

This section details the cloud resources actually used and a realistic cost estimate based on **Google Cloud Platform (GCP)** pricing. The project relied on:

- **Compute Engine VMs**: e2–medium instances (2 vCPU, 4 GiB) hosting the FastAPI backend.
- **Cloud Storage Buckets**: 400–500 GiB of ERA5/NOAA data and model artefacts, accessed frequently over roughly four months.
- **Cloud SQL (PostgreSQL)** for processed data and metadata.
- GCP networking and REST API services for data delivery to the front-end.

Free-tier credits and sustained-use discounts were applied where possible.

9.1 Revised Budget

Resource	Usage Estimate	Cost (USD)
Compute Engine VMs	e2–medium, ∼300–400 instanceâhours	$24–30
Cloud Storage	400–500 GiB stored for ∼4 months; frequent read/write and region egress	$80–100
Cloud SQL (PostgreSQL)	Light usage, within free-tier thresholds	$0–5
Networking / API calls	Front-end data delivery and minor egress	$5–10
External Storage for Project Closeout	Prices range from ∼30-90	$60
	Estimated Total	**$170–205**

Table 1: Revised GCP cost breakdown.

9.2 Budget Overview

The total infrastructure expenditure remained well below the overall project budget of **$800**. No additional funding is anticipated.

10 Facilities & Equipment

The State Farm Project, "Predictive Modeling for Severe Weather Events," is primarily software-based; therefore, facilities and equipment must center around computing, data access/storage, and collaborative tools.

The Forecast TX project relied on a combination of cloud and on-campus computing infrastructure to support model training, data processing, and system deployment.

Cloud Computing Infrastructure

The primary cloud platform used was **Google Cloud Platform (GCP)** due to its ease of integration with machine learning workflows and support for student projects. Services utilized included:

- **Compute Engine VMs**: Deployed backend services and APIs for frontend communication.
- **Cloud Storage**: Used to store and serve 400–500 GB of historical and predictive weather data.
- **Cloud SQL (PostgreSQL)**: Managed structured datasets for querying and visualization.

Free-tier offerings and GCP credits were leveraged to minimize costs.

On-Campus Computing Resources

The team also made use of GPU-enabled desktop systems available in the **CSE department labs** for training large machine learning models. These workstations provided faster experimentation and model iteration than available cloud-based GPU options within budget constraints.

This hybrid infrastructure enabled efficient and cost-effective development of the predictive weather system.

11 Assumptions

- Selected peril (Hail and thunderstorm) will not change throughout the project.
- Compute Platform will be decided by the start of the next sprint cycle (AWS or GCP).
- Data Gathering, feature engineering, and model selection will take one and a half sprints.
- The First working model will take two sprint cycles.
- The Model will not predict different results each time and will be consistant

12 Constraints

- Total development costs must not exceed $800
- Non-localized hail events, we are making a localized prediction of peril
- Need to control project complexity to allow implementation within the given time frame and budget.
- State Farm Changing Requirements
- CSE 4316 related Time constraints
- Other Team related events, such as people getting sick
- Technical: AWS or GCP removing features we are counting on.

13 Risks

The following high-level risk census contains identified project risks with the highest exposure. Mitigation strategies will be discussed in future planning sessions.

Risk description	Probability	Loss (days)	Exposure (days)
Insufficient historical weather data quality or gaps in NOAA records	0.40	15	6.0
Delays in cloud platform setup (e.g., AWS or GCP)	0.30	12	3.6
Model training requires longer time due to limited computing resources	0.50	20	10.0
Web interface development delays due to front-end integration issues	0.25	10	2.5
Difficulty obtaining stakeholder feedback for usability testing	0.20	8	1.6

Table 2: Overview of highest exposure project risks

14 Documentation & Reporting

14.1 Major Documentation Deliverables

14.1.1 Project Charter

The Project Charter will be updated and maintained by each team member during each sprint as the team makes decisions and adjustments to the project. Any decision or adjustment concerning the project's operation including but not limited to its architecture, timeline, or requirements. The initial version of the document will be available on February 17, 2025, and the final version of the document will be available on April 28, 2025.

14.1.2 System Requirements Specification

The System Requirements Specification (SRS) document will reflect the requirements specified by our sponsors from State Farm. It will be updated as the sponsors communicate their goals and desires for the project. The initial version of the document will be released on March 17, 2025. The final version of the document will be released on April 28, 2025.

14.1.3 Architectural Design Specification

The Architectural Design Specification (ADS) will describe the weather predicting system's layers and their interactions. It will be updated during each sprint to reflect any changes to the SRS. The initial version will be completed on April 7, 2025. The final version will be completed on April 28, 2025.

14.1.4 Detailed Design Specification

The Detailed Design Specification (DDS) will be maintained by each team member. Any updates will be made to reflect the design choices made by the team and requirements of the sponsors. The initial version of the document is planned to be released on April 14, 2025. The final version will be completed on April 28, 2025.

14.2 Recurring Sprint Items

14.2.1 Product Backlog

Items will be added to the product backlog based on the SRS and our sponsors' needs. These items will be prioritized by categorizing them as critical with a large impact to the overall project, moderate with a moderate impact to the project, or non-essential with a low impact on the project. The group along with the sponsors will decide on the priority of tasks through discussions in meetings. Microsoft Teams is being used to maintain and share the product backlog between team members and stakeholders. Each task is assigned to a team member using the planner application in Microsoft Teams.

14.2.2 Sprint Planning

Each sprint will be planned by the team through a meeting online or in person. The tasks that are to be done during the sprint will be added to the Microsoft Teams planner. There will be 8 sprints in total for this project spanning from the spring and summer of 2025.

14.2.3 Sprint Goal

The sprint goal is decided by the team. Our team will have weekly meetings with our sponsors to incorporate their feedback and requirements for the sprint.

14.2.4 Sprint Backlog

The team will decide what items will be added to the sprint backlog. The backlog will be maintained through the Microsoft Teams planner. They will then be put into a chart to display the hours put into each item.

14.2.5 Task Breakdown

Each individual task will be assigned to a team member through the Microsoft Teams planner. Each team member will choose what task they prefer to complete during each sprint. The time spent on the tasks will be documented on a Google sheet where each member inputs how many hours they have worked on a task.

14.2.6 Sprint Burn Down Charts

Kevin Simbakwira will be the main person generating burn down charts but will be rotated if needed. The number of hours worked are logged in a CSV file in the Microsoft Teams app.

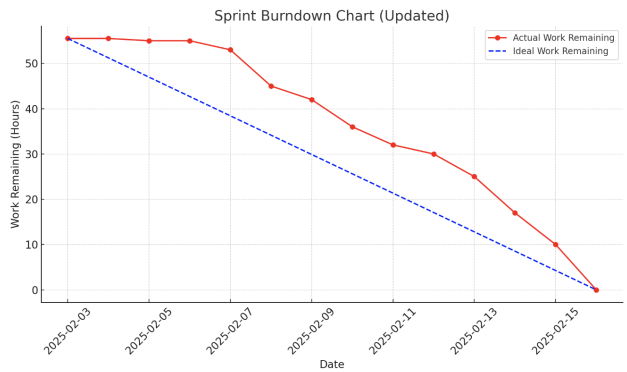


Figure 2: Example sprint burn down chart

14.2.7 Sprint Retrospective

Sprint retrospective will be done by the team by creating the sprint review together. The team will go through our objectives and how we can improve for the next sprint. Any questions, concerns, or comments will be discussed a the end of each sprint. Each sprint review will be due every 2 weeks on the Monday the sprint is ending. Each team member will also create an individual status report to give peer reviews and detail other tasks they completed during the sprint. Each individual status report will be due a day after the sprint ends.

14.2.8 Individual Status Reports

Each team member report on the sprint goal, sprint backlog, individual time expenditures, team burn down chart, individual retrospective, and peer reviews. This will be documented after every sprint.

14.2.9 Engineering Notebooks

There will not be an engineering notebook maintained for this project. All documentation will be done online and edited as needed during each sprint.

14.3 Closeout Materials

14.3.1 System Prototype

The final system prototype will be the web application that displays the predictions of our weather prediction model. The prediction of severe weather will be displayed in a consumable manner that clearly states how many years in the future the prediction takes place. A prototype acceptance test with our customer will be discussed to determine a deadline for a PAT. There will be nothing demonstrated off site.

14.3.2 Project Poster

The poster dimensions will be discussed later as we continue to work on the project more. The poster contents will include a description of the weather prediction model and the applications used to develop it.

14.3.3 Web Page

The project web page will include the system overview and a video demo of the web application. The web page will be updated throughout the project.

14.3.4 Demo Video

The demo video will showcase the results of the weather prediction model and how to navigate the web application. The length of the demo video has not been set as of yet.

14.3.5 Source Code

The source code will be maintained by each team member using Git and GitHub. Source code will only be provided to team members and sponsors.

14.3.6 Source Code Documentation

Currently, documentation standards have not been discussed by the team. The team will decide on a standard once we have further discussed the source code.

14.3.7 Installation Scripts

There has not been any discussion on how the customer will deploy software to new installations currently. The team will discuss how the application will be deployed by the customer in future meetings.

14.3.8 User Manual

A user manual has not been discussed currently. The team will continue to work on the application and create features, and if the team decides the features are not intuitive and user friendly, we will create a user manual.

References

[1] Hritik Bansal, Shashank Goel, Tung Nguyen, and Aditya Grover. Climatelearn: Machine learning for predicting weather and climate extremes. Google Colab Notebook, 2021. Accessed: 2025-02-16.

[2] NVIDIA Corporation. *FourCastNet: Training the Model*, 2024. Accessed: 2025-02-16.

[3] DeepMind. GenCast Predicts Weather and the Risks of Extreme Conditions with SOTA Accuracy. https://deepmind.google/discover/blog/gencast-predicts-weather-and-the-risks-of-extreme-conditions-with-sota-accuracy/, 2024. Accessed: 2025-02-16.

[4] Ankur Mahesh. Climate change ai summer school: Day 3 climate science tutorial: Seasonal forecasting. Google Colab Notebook, 2023. Accessed: 2025-02-16.

[5] National Severe Storms Laboratory (NSSL). Multi-Radar Multi-Sensor (MRMS) Project. Technical report, NOAA, 2024. Accessed: 2025-02-16.