Forecasting Frost Events For California Growers

Raymond Yan
Agricultural & Biorenewable Systems Management, Penn State University

Simon Yan
Computer Science, University of California at Santa Barbara

1 Introduction

Frost events are one of the most damaging weather risks to California Central Valley's agricultural producers. A frost event during bloom periods can significantly damage plant tissues, and ultimately cause lower yields and economic losses. Traditional frost forecasting methods depend heavily on heuristics and local knowledge, which often fails to capture unforeseen variability in microclimates, geography, and other atmospheric factors. Leveraging machine learning modelling with long-term weather measurements and meteorological datasets can help growers better forecast frost events based on historical trends.

This project aims to develop a framework to leverage on-the-ground CIMIS weather stations and ERA5 historical atmospheric data to predict frost events across multiple time horizons. The goal of the approach is to conduct Machine Learning predictions based on measurements at the ground level and reinforcing those predictions with macro-level frost dynamics from atmospheric data. The framework can then be expanded beyond CIMIS's 145 weather stations (run by California's Department of Water Resources) into weather stations owned by growers, leading to highly regional frost forecasting for their production operations.

2 Data Overview

2.1 Primary Dataset

We used one primary dataset for training our model: California Irrigation Management Information Systems (CIMIS) hourly weather data. Our CIMIS dataset consists of hourly measurements from 18 weather stations distributed across agricultural regions in California, and the dataset spans 15 years (2010-2015).

The raw dataset was from a *cimis_all_stations.csv* file, which once parsed, contained each weather station's near-surface measurements for these variables.

Dataset Variables

The following are core variables from every weather station measurement that we trained our model on:

- Air Temperature
- Dew Point
- Soil Temperature
- Relative Humidity
- Wind Speed
- Solar Radiation

Derived Variables

To improve our predictions, we also derived secondary features from our dataset:

- Temperature Drop
- Dew Point Depression
- Vapor Pressure Deficit
- Air-Soil Temperature Difference
- Cyclical Time Encodings

2.2 Dataset Processing

The raw CIMIS dataset required several transformations before it could be used for frost forecasting model development. Each record contained hourly measurements of temperature, humidity, wind, radiation, and soil conditions. We applied preprocessing and feature engineering steps to make the dataset more climatologically accurate and suitable for model development.

Temporal Features

Frost risk is highly tied to seasonal cycles, so we encoded the *hour-of-day* and *month-of-year* using sine and cosine transformations. This preserves the cyclical nature of days and months and allows our model to learn from those periodic patterns. For example, the temperature is always going to be colder on average in winter months than summer months.

Lagged Temperature Trends

The cooling dynamics leading up to frost events can be a strong predictor of future temperature changes. We calculated temperature changes over multiple time windows (3, 6, 12, and 24 hours) to represent these trends. The records that lacked sufficient historical data across the 15 years were excluded from the processed dataset to ensure consistency.

Atmospheric Indicators

Three variables were derived from the dataset features to represent common indicators of frost-related processes: Dew Point Depression, Vapor Pressure Deficit (VPD), and Air-Soil Temperature Difference.

- Dew Point Depression: Dew Point Depression is defined as the difference between air temperature and dew point, indicating a proximity to saturation.
- Vapor Pressure Deficit (VPD): Vapor Pressure Deficit is derived from saturation and actual measure vapor pressure - this represents how dry the atmosphere is. [science]: Lower values of VPD indicate a higher risk of impending frost.
- Air-Soil Temperature Difference: Air-Soil Temperature Difference highlights radiative cooling potential, meaning larger difference values indicate faster surface cooling in relation to subsurface conditions.

Frost Event Labelling

Binary frost indicators were defined as the number of hours where the air temperature is ≤ 0°C. The forecasting was conducted for four predictions horizons: 3 hours, 6 hours, 12 hours, and 24 hours into the future.

2.3 Model Training

We implemented a supervised learning framework using classifiers from XGBoost to help predict frost occurrence events across multiple horizons (3 hours, 6 hours, 12 hours, 24 hours). The XGBoost library is great for handling the nonlinear relational nature of weather data.

Training Dataset

For training data, we preprocessed the CIMIS dataset and split it into training and testing sub-datasets. We used 80% of our records for training and the remaining 20% for evaluation.

Feature Set

The models were trained on a combination of raw meteorological variables and derived features:

- **Core Features:** Air Temperature, Dew Point, Soil Temperature, Relative Humidity, Wind Speed, Solar Radiation.
- **Derived Features:** Dew Point Depression, Vapor Pressure Deficit, Air-Soil Temperature Difference, Lagged Temperature Drops (3 hours, 6 hours, 12 hours, 24 hours).

The combined feature set helps provide both immediate indicators and contextual signals for impending frost risk.

Model Configuration

Our horizon-specific model was trained with hyperparameters like the following:

- Number of Estimators: 500
- Max Depth: 5

- Learning Rate: 0.05
- Subsample Ratio: 0.8
- Column Sampling: 0.8
- Evaluation Metric: Log Loss

We selected these configuration values for empirical performance reasons rather than a fully-exhaustive grid search, which balances model accuracy with computational efficiency.

Evaluation Metrics

To assess model performance, we applied both probabilistic and classification metrics:

- **Brier Score:** Brier Score measures the accuracy of probability forecasts.
- **ROC-AUC:** ROC-AUC evaluates discrimination between frost and non-frost events.
- **PR-AUC:** PR-AUC highlights performance under an imbalance of the classes.
- **Expected Calibration Error (ECE):** ECE quantifies alignment between predicted probabilities and the observed frequencies.

3 Results

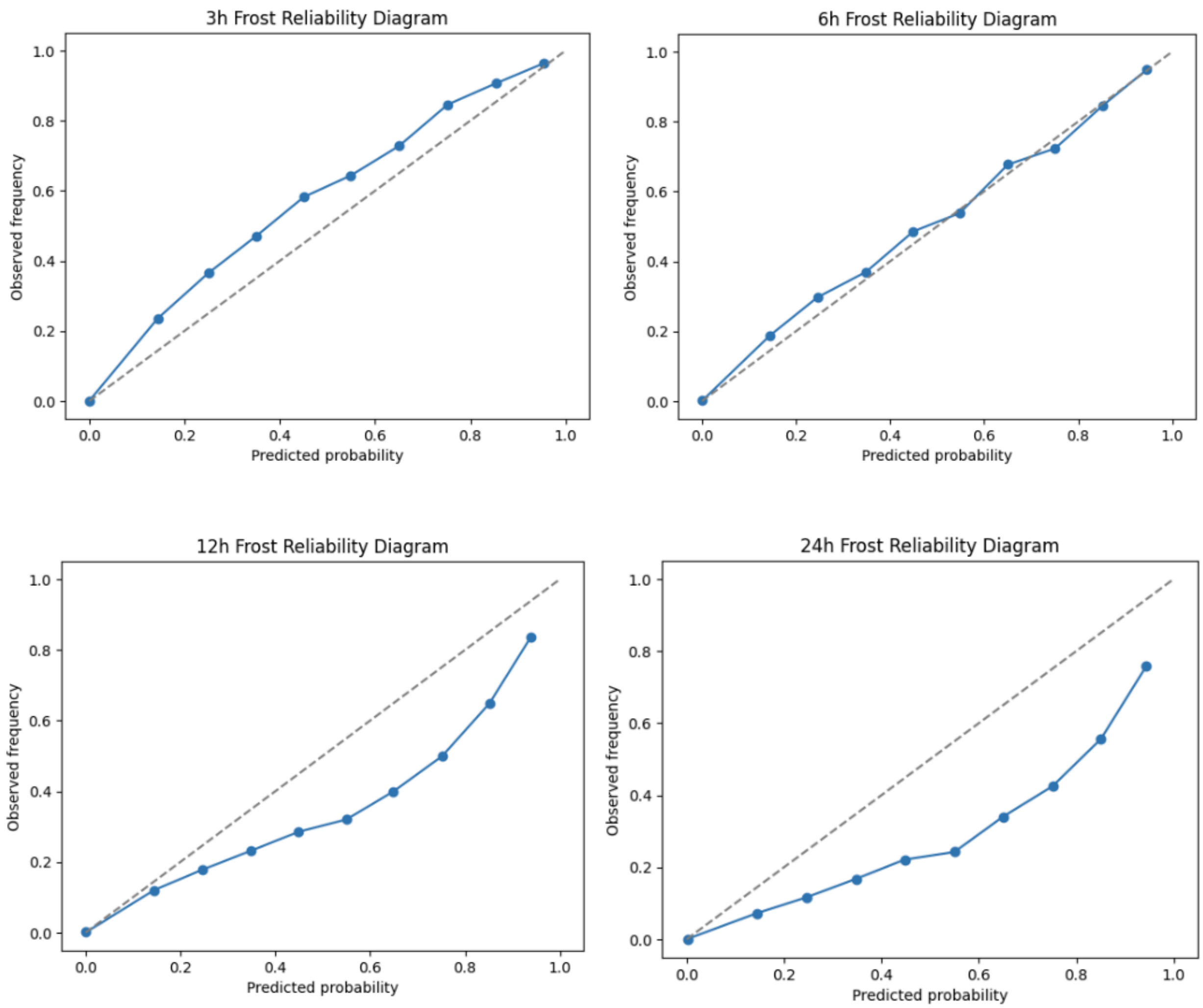
3.1 Model Evaluations

We evaluated our trained XGBoost models across four forecast horizons: 3 hours, 6 hours, 12 hours, 24 hours. The performance was assessed using probabilistic accuracy and calibration quality. Below are the results for the time horizons:

Time Horizon	Brier Score	ROC-AUC	PR-AUC	ECE
3 hours	0.003585	0.997513	0.867440	0.001782
6 hours	0.004728	0.994769	0.775625	0.001059
12 hours	0.007266	0.990368	0.606882	0.004192
24 hours	0.008928	0.987563	0.537952	0.008475

Table 1: The metrics of our XGBoost model across the four time horizons.

The following diagrams depict the reliability of predictions across the four time horizons:



Summary of Reliability

- **Short Time Horizon (<6 hours):** Extremely low brier scores indicate a highly accurate probability forecast. High ROC-AUC also indicates near perfect discrimination between frost and non-frost events. High PR-AUC shows strong performance for detecting rarer frost events.
- **Medium Time Horizon (12 hours):** At a 12-hour horizon, the Brier score starts to increase, indicating growing uncertainty. ROC-AUC still remains strong, but PR-AUC significantly weakens - rarer frost events are more difficult to detect at this time horizon.

- **Long Time Horizon (24 hours):** At a 24-hour time horizon, the Brier score continues to increase. While ROC-AUC remains high, PR-AUC continues to significantly weaken. It becomes difficult to distinguish between frost and non-frost events at this time horizon.

3.2 Feature Importance

Analysis of the importance of features also revealed patterns across the four time horizons:

- Dew Point Depression: Analysis found that Dew Point Depression is the most influential predictor of frost events.
- Lagged Temperature Drop: Lagged Temperature Drops provide strong signals in shorter time horizons, revealing trends in cooling.
- Air-Soil Temperature Difference: This secondary predictor can help highlight radiative cooling potential, making it an important feature to consider.
- Cyclical Encoding: Winter months tend to be cooler than Summer months. Night time tends to be cooler than daytime. Our calibrations were improved by encoding these diurnal and seasonal cycles into our modelling.

3.3 Summary of Performance

Our CIMIS-only models demonstrated that we can reach reliable frost event forecasts in short time horizons by encoding cyclical indicators and common indicators of frost. The accuracy of predictions becomes less useful for longer time horizons. External datasets can be combined with the CIMIS datasets to provide more accurate forecasts and drivers unforeseen by weather stations at the ground level.

4 External Datasets (ERA5)

To strengthen the predictive abilities of our baseline CIMIS-only modelling, we incorporate several external datasets that help capture a broader view on atmospheric factors, radiative processes, and terrain-based factors. This additional data provides context beyond the ground-level weather station measurements, enabling our model to generalize to other diverse climates and topographical regions. It also helps with strengthening the forecasting at longer time horizons.

4.1 Atmospheric Reanalysis (ERA5)

We utilized the *open meteo public API* to query data from ERA5. ERA5 is a comprehensive reanalysis dataset that incorporates data from satellites, ground stations, and ships for advanced weather prediction modelling. The ERA5 dataset covers the entire earth and provides continuous coverage of data across time. ERA5 helps bolster our observations from CIMIS weather stations. We utilized a wide range of variables relevant to frost prediction including:

- 2 Meter Temperature (min, max, and mean): Standardized temperature near the surface.
- Apparent Temperature: Temperature accounting for wind, humidity effects, etc.

- Pressure at Mean Sea Level: Considers ridges and other features that can cause radiative cooling effects.
- Cloud Cover: Can help distinguish frost events when there is radiative shielding.
- Weather Codes: Includes weather conditions like fog, precipitation, etc.

We integrated ERA5 variables with CIMIS observations so that the XGBoost model can learn from both local microclimates as well as macro-level atmospheric effects.

4.2 Cloud and Radiation Variables

Cloud and radiation variables describe how energy is exchanged between the surface and the atmosphere, which explains frost formation. In short, if skies are clear and it is night time, the earth's surface loses heat to space through radiative cooling - this increases frost risk. On the other hand, clouds act like an insulating blanket, trapping in radiation and lowering frost risk.

Our external dataset includes such variables:

- Total Cloud Coverage: % of sky obscured by clouds. This is a generalization of overall radiative shielding effects.
- Low-Level Cloud Cover: Focused on clouds closer to the surface, which are most effective at radiative shielding.
- Daylight Duration: # of hours of sunlight per day, which affects how long the surface is heated during the day.
- Sunshine Duration: Actual # of hours of solar radiation
- Shortwave Radiation Sum: Sum of solar energy entering.

In short, cloud coverage variables help predict whether the clouds are trapping in radiation or not. It can help explain frost events that occur from advective cooling. Sunshine and shortwave radiation help to quantify how much heat is stored in the earth's surface during the day, which also affects nighttime cooling. Finally, daylight duration incorporates seasonal factors into the modelling. We merged these variables with baseline CIMIS measurements to create a hybrid set of features. The integration better helps explain frost dynamics like radiation and insulation, while anchoring to ground-level measurements.

4.3 Moisture and Energy Variables

Moisture and energy indicators help explain how water and heat are exchanged between surface and atmosphere. These are processes that influence how quickly the ground can cool at night.

We incorporated these variables from the external dataset:

- Soil Moisture: Measurement of water content near the surface.
- Precipitation Sum: Total rainfall.
- FAO Evapotranspiration: Standardized measure of the atmosphere's demand for water, which combines radiation, temperature, humidity, and wind.

Soil moisture affects how fast heat is released - dry soil can cool more rapidly than wet soil. High evapotranspiration rates could indicate stronger atmospheric demand, which accelerates cooling under clear skies. Combined with cloud and radiation variables, the evapotranspiration variable can help the model distinguish between radiative frost events and latent heat flux effects.

4.4 Topography

Terrain features can exert a strong influence on microclimates in ways that singular weather stations cannot capture. For example, higher elevations can cool more rapidly since they are in a thinner atmosphere. We incorporated elevation data from the external dataset for each weather station. The next section summarizes the results after incorporating external datasets together with our CIMIS dataset.

5.1 Model Evaluation (using External Dataset)

The integration of external variables significantly improved performance across all four forecast horizons. The following table shows the performance across the four time horizons:

Time Horizon	Brier Score	ROC-AUC	PR-AUC	ECE
3 hours	0.003472	0.997882	0.873682	0.001584
6 hours	0.004623	0.995800	0.788443	0.001140
12 hours	0.006756	0.991543	0.631130	0.002949
24 hours	0.007314	0.991261	0.649790	0.005716

Table 2: The metrics of our XGBoost model across the four time horizons, now with external data from ERA5 and topography included.

Comparison to CIMIS-only Model

- Very Short Time Horizon (3 hours): When including the external dataset, ROC-AUC is nearly identical to the CIMIS-only model. PR-AUC and calibration metrics are slightly improved.
- Short Time Horizon (6 hours): The ROC-AUC metric starts to show improvement when compared to the CIMIS-only model. The improvement between PR-AUC also continues to show.
- Medium Time Horizon (12 hours): At the 12-hour horizon, the external dataset model starts to show a clear improvement. The ROC-AUC and PR-AUC are both higher. ECE also continues to reduce.
- Long Time Horizon (24 hours): In the 24-hour horizon, the external dataset model substantially improves. PR-AUC increases by around 20% in comparison to the CIMIS-based model.

5.2 Example Predictions

We displayed sample forecasts at individual CIMIS weather stations across various years and seasons to examine how external datasets help capture both microclimate and atmospheric effects. Below are examples:

```
Panoche at 2021-03-24 12:00:00

There is a 0.0% chance of frost in the next 3h, predicted temperat
ure: 21.19 °C
There is a 0.0% chance of frost in the next 6h, predicted temperat
ure: 19.24 °C
There is a 0.0% chance of frost in the next 12h, predicted tempera
ture: 9.36 °C
There is a 0.0% chance of frost in the next 24h, predicted tempera
ture: 18.74 °C
```

Prediction 1: CIMIS Weather Station at Panoche, CA

```
Oakdale at 2011-12-07 08:00:00

There is a 0.1% chance of frost in the next 3h, predicted temperat
ure: 8.84 °C
There is a 0.1% chance of frost in the next 6h, predicted temperat
ure: 13.02 °C
There is a 0.4% chance of frost in the next 12h, predicted tempera
ture: 4.84 °C
There is a 97.0% chance of frost in the next 24h, predicted temper
ature: -0.20 °C
```

Prediction 2: CIMIS Weather Station at Oakdale, CA. The output predicts a high chance of a frost within the next 24 hours, with a predicted temperature of below the freezing point.

5.3 Leave-One-Station-Out Validation (LOSO)

To make sure that our models are generalizable across a diversity of microclimates, we applied the Leave-One-Station-Out Validation (LOSO) approach. In this approach, models are trained on 17 stations and tested on the 18th station. This is essentially a simulation of training on 17 weather stations and deploying in an unseen 18th station.

LOSO Validation Results

Time Horizon	Brier Score	ROC-AUC	PR-AUC	ECE
3 hours	0.002254	0.998569	0.842734	0.000326

6 hours	0.002983	0.997714	0.775673	0.002021
12 hours	0.004613	0.996020	0.662762	0.006009
24 hours	0.006896	0.993872	0.609978	0.010604

Table 3: The metrics of our XGBoost model across the four time horizons, with LOSO validation.

Findings

- **Generalization:** The models maintained high ROC-AUC when tested on unseen stations. The external variables provide indicators that transfer across diverse regions.
- **ECE Challenge:** ECE increased for longer time horizons, which reflects some difficulty in probability reliability across certain regions.
- **Temperature Forecasting:** The regression performance remained strong with MAE values consistent across different time horizons.

LOSO Validation helps models generalize across regions where the frost history is unknown.

For practical purposes, California's Department of Water Resources cannot deploy and manage CIMIS weather stations in every microclimate variation across the state. The generalizability of this type of model allows for integration with personal weather stations where CIMIS stations are not present.

6 Conclusion

Our reliability diagrams provided a clear view of how well our predicted probabilities aligned with observed frost events. At short horizons (3 hours and 6 hours), forecasts were highly reliable (predictions tracking close to diagonal dotted line). At a 12-hour and 24-hour time horizon, calibration starts to weaken.

Overall Findings

We found that while training a XGBoost model on CIMIS stations located in diverse regions of California can help forecast frost events in unforeseen regions, there are certain limitations. By incorporating data from ERA5 and information on topography, we were able to service features that tracked common dynamics of frost events. For example, we derived radiation and cloud coverage variables to account for adjective cooling. It is a perfect example of an event that on-the-ground-only datasets may not be able to detect. The usefulness of a model that combines macro-level atmospheric dynamics with ground-level measurements is that it can be more generalizable to regions where frost event history is unknown. The model can be integrated with weather stations in regions that don't yet have a CIMIS station to help make predictions.