Modeling Work Order Responses to Precipitation: A Comparative Analysis of Regression and Machine Learning Techniques

Kevin Bentovoja

Department of Atmospheric and Oceanic Sciences, University of California Los Angles

AOS C111: Introduction to Machine Learning for Physical Sciences

Dr. Alexander Lozinski

1. Introduction

Extreme precipitation in the Los Angeles region is becoming increasingly episodic and intense, amplifying flood risk even during periods of prolonged drought. Yet campus-scale flood vulnerability, particularly how storm intensity translates into operational impacts, remains underexamined in both research and applied climate resilience planning. UCLA routinely experiences storm-related disruptions, including flooded buildings, blocked storm drains, and surges in maintenance requests, placing substantial strain on Facilities Management resources. Although UCLA maintains detailed weather and work order records, there is currently no quantitative framework capable of predicting how maintenance demand scales with rainfall intensity. This project addresses that gap by developing a comparative analysis of regression-based machine learning techniques to model the relationship between precipitation and work order volume. Using historical NOAA rainfall observations and UCLA's Facilities Management Maximo Raincall database, the study evaluates the predictive performance of multiple regression approaches, including linear, polynomial, ridge, and support vector regression (SVR), to determine which methods best capture the nonlinear dynamics between extreme precipitation events and work order responses.

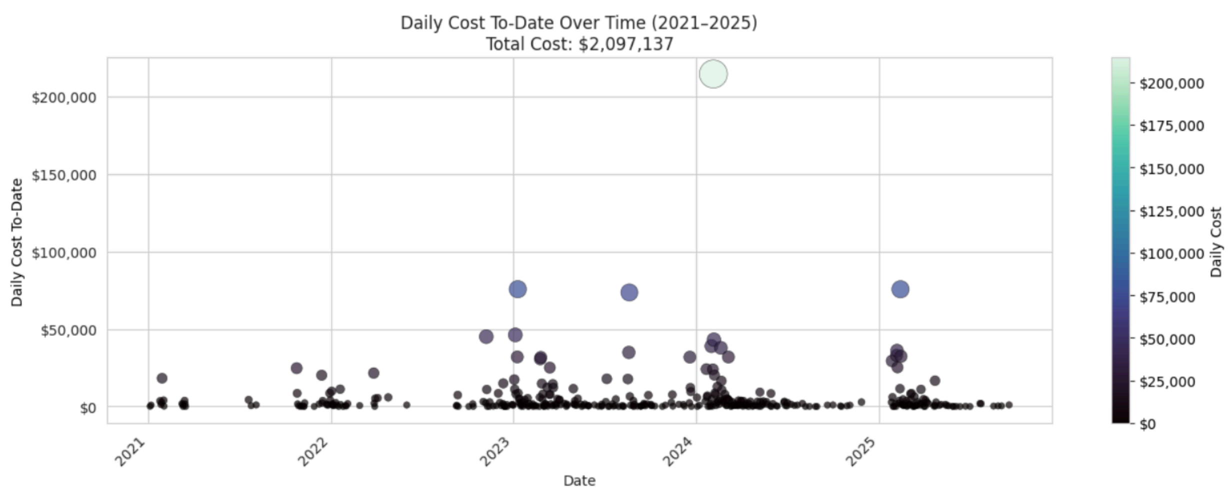

Fig 1: Distribution of daily maintenance costs for storm events (2021–2025)

2. Data and Pre-processing

This analysis integrates two primary datasets: (1) daily precipitation observations from NOAA for the UCLA station, spanning January 2021 to September 2025, and (2) UCLA Facilities Management's Maximo Raincall work order records for the same period. The NOAA dataset includes daily precipitation in millimeters, which was converted to inches to improve interpretability for a U.S. campus context and to align with common reporting units. The Facilities dataset required parsing the "Reported Date" field into a standardized datetime format.

After cleaning both datasets, precipitation values were merged onto the Facilities dataset using shared calendar dates, allowing each day's work order activity to be directly paired with measured rainfall intensity.

The distribution of daily precipitation is highly skewed, with many small rainfall events and only a few very large storms. This creates a long right tail in the original precipitation data, as shown in the (Figure 2). Such skew can reduce model stability and make it difficult for regression techniques to learn meaningful relationships. Applying a log10 transformation produces a more balanced and compact distribution, shown in the second histogram (Figure 3), which better represents the range of storm intensities. This transformation makes the data more suitable for regression modeling by reducing the influence of extreme rainfall values and improving the ability of the models to detect patterns in storm-driven work order activity.

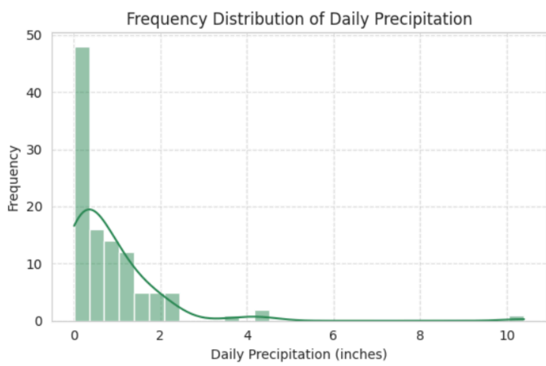

Fig 2: Distribution of NOAA Precipitation

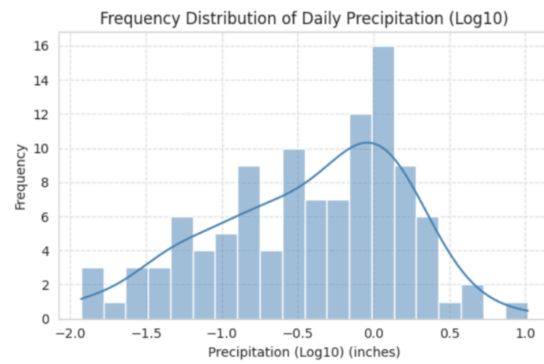

Fig 3: Log10 Precipitation Distribution

3. Modeling:

To evaluate how daily precipitation influences Facilities Management work order activity, supervised regression methods were applied. Since the objective was to predict work order volume from measurable rainfall intensity, a supervised approach was appropriate for identifying patterns within the merged NOAA and FM datasets. After data cleaning and transformation, several regression models were tested to assess their ability to capture the relationship between storm intensity and maintenance demand.

For this project, four regression models were examined:

1. Linear Regression
2. Polynomial Regression
3. Ridge Regression

4. Support Vector Regression (SVR)

A modeling dataset was created by grouping each rainy day and calculating the total number of work orders alongside the corresponding precipitation values. These variables were then split into training and testing sets to support model development and to evaluate predictive performance on data not used during training.

3.1. Linear Regression:

Linear Regression provides a straightforward baseline model for assessing the relationship between daily precipitation and work order volume. As a simple linear approach, it offers an initial benchmark against which more flexible or regularized models can be compared. This model assumes a direct linear relationship between log-scaled precipitation and the number of work orders, making it useful for establishing reference performance before introducing more complex regression techniques.

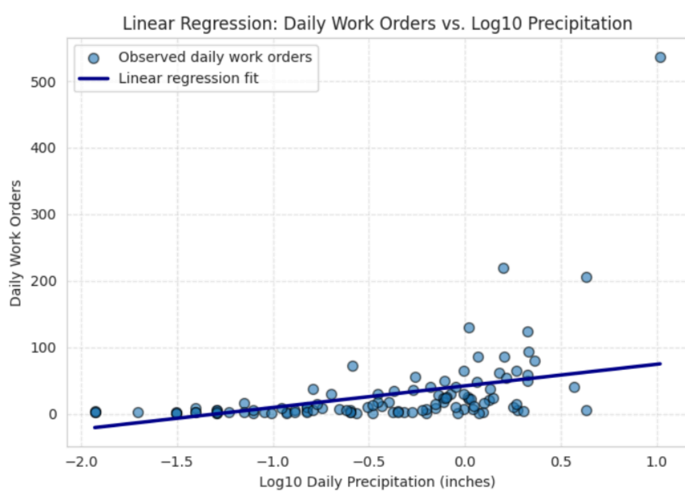

Fig 4: Linear Regression Fit for Log10 Precipitation and Work Orders

3.2. Polynomial Regression:

Polynomial Regression builds on the linear baseline by allowing the predictor to curve in ways that better represent nonlinear responses often seen in environmental systems. Since precipitation impacts on work orders are unlikely to increase at a constant rate, polynomial terms provide a controlled way to introduce curvature into the model. Degrees 2 through 6 were tested to examine how additional complexity influenced the model's ability to capture these nonlinear patterns while also revealing how higher-degree terms can increase sensitivity to noise and potential overfitting. This range allowed for a systematic evaluation of model flexibility and provided insight into how much curvature is supported by the underlying physical and operational relationships in the dataset.

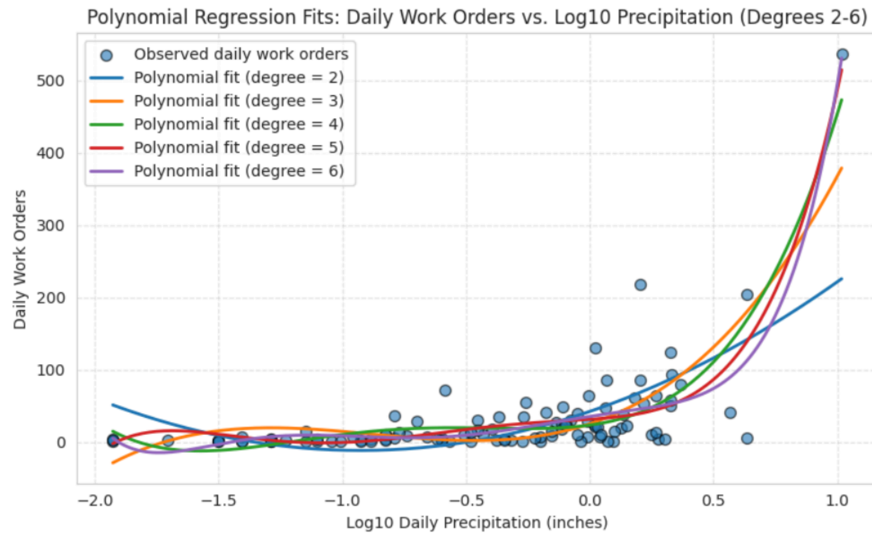

Fig 5: Polynomial Regression Fits for Log10 Precipitation and Work Orders

3.3. Ridge Regression:

Ridge Regression extends the linear framework by adding a regularization term that shrinks coefficient values and reduces sensitivity to noise in the data. This method is particularly useful when models become more flexible, such as after introducing polynomial features, because increased complexity can amplify variance and lead to unstable predictions. By penalizing large coefficients, Ridge Regression helps stabilize the relationship between log-scaled precipitation and work order volume, offering a more controlled fit in situations where nonlinear patterns exist but the underlying signal is noisy. This approach supports more reliable generalization and provides an opportunity to test whether a regularized linear model can outperform the unregularized baseline.

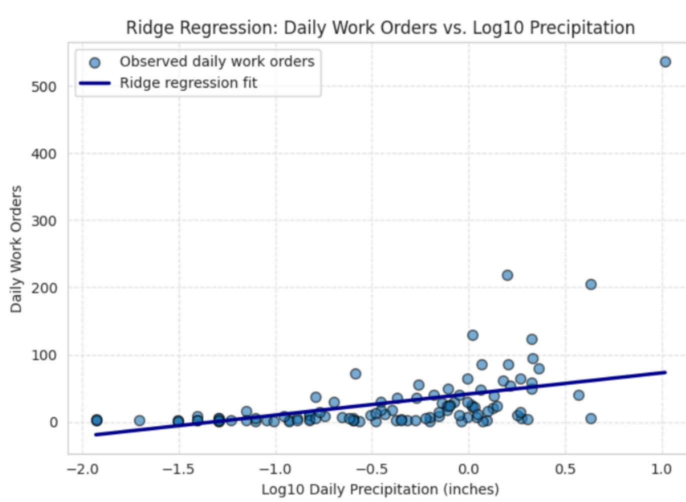

Fig 6: Ridge Regression Fits for Log10 Precipitation and Work Orders

3.4. Support Vector Regression (SVR)

Support Vector Regression was applied to explore nonlinear patterns in the relationship between precipitation and work order volume. An RBF kernel was selected because it allows the model to capture curved relationships that simpler regression methods may miss. The parameter C controls how strongly the model penalizes errors, epsilon sets the tolerance around the regression line where small deviations are ignored, and gamma determines how far the influence of each data point extends. Together, these parameters shape how flexible or smooth the final function becomes. Since SVR is sensitive to differences in scale, both predictors and targets were standardized prior to training. This approach tests whether a kernel-based method can better represent storm-driven maintenance responses in comparison to the linear and polynomial models.

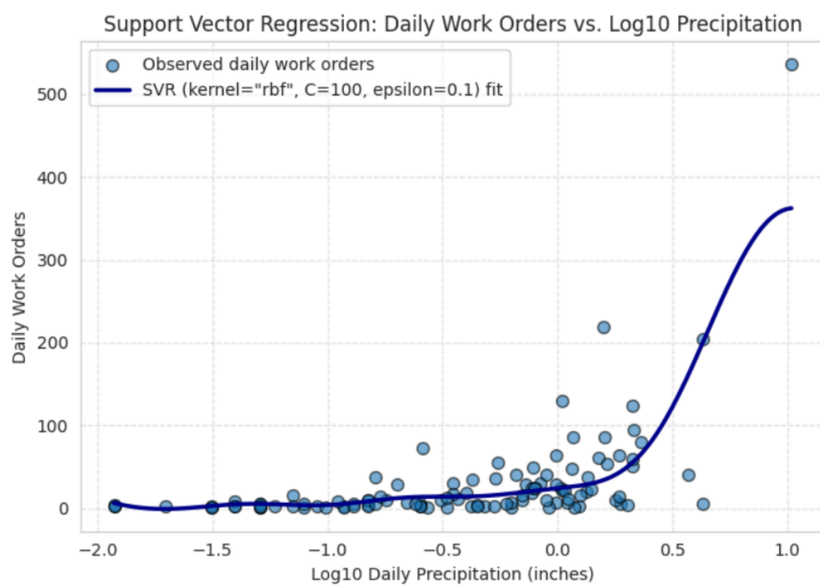

Fig 6: SVR Fits for Log10 Precipitation and Work Orders

4. Results:

Model performance was evaluated using RMSE and R^2 for both training and testing datasets, allowing comparison of how each regression approach captured the relationship between log-scaled precipitation and daily work order volume. The linear baseline produced the highest testing RMSE (83.65) and the lowest testing R^2 (0.18), indicating a limited ability to represent the nonlinear response observed in the data. Polynomial Regression showed clear improvements in predictive skill as model degree increased, with testing RMSE decreasing from 72.85 at degree 2 to 32.90 at degree 6. Correspondingly, testing R^2 increased from 0.38 to 0.87, demonstrating that higher-order polynomial terms captured the curvature in the precipitation–response relationship more effectively than linear methods.

Ridge Regression performed similarly to the linear baseline, with testing RMSE of 83.87 and testing R^2 of 0.17, suggesting that regularization did not improve predictive skill in the absence of additional nonlinear features. Support Vector Regression showed moderate improvement relative to lower-order polynomial models, producing a testing RMSE of 47.77 and a testing R^2 of 0.73. This indicates that the nonlinear kernel in SVR more effectively represented the structure in the data than unregularized linear methods, though it did not outperform the highest-degree polynomial models.

Overall, the results show that methods capable of modeling nonlinear relationships achieved substantially better predictive accuracy. Polynomial Regression of degree 6 achieved the strongest performance across all metrics, suggesting that the relationship between storm intensity and work order demand is highly nonlinear and best represented by models with sufficient flexibility to capture this curvature.

Model	**Training RMSE**	**Testing RMSE**	**Training R^2**	**Training R^2**
Linear	35.41	83.65	0.24	0.18
Polynomial (2nd degree)	33.25	72.85	0.33	0.38
Polynomial (3rd degree)	32.39	60.63	0.37	0.57
Polynomial (4th degree)	32.04	47.58	0.38	0.73
Polynomial (5th degree)	32.03	42.37	0.38	0.79
Polynomial (6th degree)	32.01	32.90	0.38	0.87
Ridge	35.41	83.87	0.24	0.17
SVR	34.18	47.77	0.29	0.73

5. Discussion

The modeling results show that daily work order volume does not respond linearly to increasing precipitation. The poor performance of the linear baseline indicates that storm-driven maintenance demand accelerates in a way that simple linear structures cannot capture. Higher-degree Polynomial Regression produced the strongest performance across all models, suggesting that the relationship between storm intensity and operational impacts is strongly nonlinear. This pattern is consistent with physical expectations, as heavier rainfall can exceed drainage capacity, trigger compounding failures, and generate multiple work orders on the same day.

Ridge Regression did not improve predictive skill, which reflects the limited benefit of regularization when only a single predictor is used and no additional nonlinear terms are included. Support Vector Regression, which incorporates a nonlinear kernel, performed reasonably well and demonstrated its ability to model complex relationships in environmental datasets. However, its performance did not surpass the highest-degree polynomial models, indicating that the polynomial structure aligned more closely with the curvature present in the precipitation response.

Despite the strong performance of higher-degree models, several limitations remain. Work order volume is influenced by many additional factors, including antecedent soil moisture, building vulnerability, deferred maintenance, seasonal staffing, and reporting behavior. Precipitation intensity alone cannot account for these sources of variability, which likely contributes to the remaining scatter in the data and the moderate predictive accuracy of even the most flexible models. In addition, higher-degree polynomial models may capture noise in the dataset when applied outside the historical range of events, which should be considered when interpreting the results.

Overall, the analysis demonstrates that precipitation intensity is a meaningful but incomplete predictor of storm-related work order activity. The findings highlight the value of incorporating nonlinear techniques into climate resilience planning and point to opportunities for future modeling that includes additional environmental and operational variables.

6. Conclusion

This analysis provides evidence that machine learning models can capture meaningful patterns between storm intensity and daily work order volume, but the results also show that precipitation is only one part of a more complex operational system. The nonlinear models improved predictive skill compared to the linear baseline, suggesting that higher-intensity storms generate maintenance demands that escalate more rapidly than lighter events. However, even the best-performing models leave substantial unexplained variability, reflecting the influence of factors not included in this study, such as building conditions, storm duration, reporting behavior, and pre-existing vulnerabilities.

The findings indicate that rainfall intensity alone cannot fully predict operational impacts, yet it still offers a useful signal for understanding broad trends in storm-related workload. This work establishes an initial framework that can be expanded with additional environmental and institutional variables, potentially enabling more reliable forecasts of maintenance demand during severe weather. Rather than offering definitive predictions, the models demonstrate both the potential and the limitations of data-driven approaches when applied to real campus operations.