Accurate and Interpretable Radar Quantitative Precipitation Estimation with Symbolic Regression

REU Site: Online Interdisciplinary Big Data Analytics in Science and Engineering

Olivia Zhang[1,2*], Brianna Grissom[3*], Julian Pulido[4], Kenia Munoz-Ordaz[5], Jonathan He[6], Mostafa Cham[7], Haotong Jing[1], Weikang Qian[1], Yixin Wen[1], Jianwu Wang[7]
[1]Department of Geography, University of Florida, USA
[2]Department of Statistics, University of Florida, USA
[3]Department of Statistics and Applied Probability, University of California, Santa Barbara, USA
[4]Department of Computer Science, California State University, Sacramento, USA
[5]School of Computing and Design, California State University, Monterey Bay, USA
[6]Atholton High School, Maryland, USA
[7]Department of Information Systems, University of Maryland, Baltimore County, USA
Technical Report HPCF–2024–4, `hpcf.umbc.edu` > Publications

Abstract

Accurate quantitative precipitation estimation (QPE) is essential for managing water resources, monitoring flash floods, creating hydrological models, and more. Traditional methods of obtaining precipitation data from rain gauges and radars have limitations such as sparse coverage and inaccurate estimates for different precipitation types and intensities. Symbolic regression, a machine learning method that generates mathematical equations fitting the data, presents a unique approach to estimating precipitation that is both accurate and interpretable. Using WSR-88D dual-polarimetric radar data from Oklahoma and Florida over three dates, we tested symbolic regression models involving genetic programming and deep learning, symbolic regression on separate clusters of the data, and the incorporation of knowledge-based loss terms into the loss function. We found that symbolic regression is both accurate in estimating rainfall and interpretable through learned equations. Accuracy and simplicity of the learned equations can be slightly improved by clustering the data based on select radar variables and by adjusting the loss function with knowledge-based loss terms. This research provides insights into improving QPE accuracy through interpretable symbolic regression methods.

Key words. Quantitative precipitation estimation, Polarimetric radar, Symbolic regression, Knowledge-based loss terms.

1 Introduction

Accurate estimations of rainfall are crucial for a variety of applications such as extreme weather condition forecasting, flash flood monitoring, and ongoing climate research [1]. Rainfall can be measured directly through rain gauge stations or indirectly calculated through radar remote sensing. Rain gauges provide accurate data but have limited spatial coverage, whereas radars have higher coverage yet more uncertainty as rainfall rate must be calculated from radar data through empirical equations [2]. The traditional radar method for estimating precipitation primarily relied on the Z–R relationship between rainfall rate R and radar reflectivity factor Z. This approach, however, has limitations due to the complex and variable nature of rainfall [3]. There has been significant progress in quantitative precipitation estimation (QPE) with the development of dual-polarimetric radars, which provide additional variables such as differential reflectivity Z_{DR}, specific differential phase K_{DP}, and correlation coefficient ρ_{hv}. The integration of these variables has been shown to

*These authors contributed equally to this work.

reduce uncertainty in QPE by addressing issues related to drop size distribution variability, radar miscalibration, attenuation, and partial beam blockage [3].

Despite these advancements, challenges still remain in accurately modeling the complex relationships between radar measurements and precipitation rate. Symbolic regression is a machine learning method that discovers mathematical relationships from data [4,5], which offers a promising approach to further enhance the interpretability and accuracy of radar-based precipitation estimation. Symbolic regression will create interpretable models, allowing us to explicitly understand how specific variables impact rainfall rate in the form of concrete equations. Our implementation code is publicly available.[1]

This research aims to improve precipitation estimation from dual-polarimetric radar data through the following contributions:

- Testing various symbolic regression models, from genetic programming to deep learning-based methods, in order to understand how symbolic regression could achieve accurate and interpretable QPE.
- Studying whether improved mathematical equations could be learned by subsetting the data prior to applying symbolic regression, considering the complexity of the physics of precipitation and various precipitation types.
- Studying how existing symbolic regression approaches could be extended to embed domain knowledge and to adapt to the complexity of precipitation.

The remainder of this report is organized as follows: Section 2 provides a background of QPE and symbolic regression, Section 3 covers related work, Section 4 describes our dataset, Section 5 covers our methodology for testing the symbolic regression models, subsetting the data prior to applying symbolic regression, and applying knowledge-based loss terms, Section 6 covers the results, Section 7 reviews the equations produced by symbolic regression, and Section 8 provides a conclusion of our research and an overview of future work.

2 Application Background

2.1 Quantitative Precipitation Estimation (QPE)

One method of measuring precipitation is through a network of rain gauges, each of which collects rain at a specific time and location. However, sampling errors, poor gauge placement, wind, clogging of the gauge funnel, and other errors can cause inaccurate precipitation measurements [6]. Another way of measuring precipitation is through radar-derived QPE sources, which provide wider coverage compared to rain gauges as radar can sample large areas in a short amount of time [2]. In particular, polarimetric radars perform real-time and high-resolution QPE the most efficiently [7].

In quantitative estimates of precipitation with radars, conventional methods use a Z–R relationship to estimate rainfall R (mm/hr) via reflectivity Z (mm^6/m^3), given by the equation $Z = aR^b$, where a and b are constants that vary by drop size distribution and precipitation type [2,8], and Z measures the amount of energy reflected back to the radar after hitting a raindrop [9]. However, the Z–R relationship fails to account for nuances in rainfall by precipitation type, region, and season [2].

[1] View our implementation code at `https://github.com/big-data-lab-umbc/big-data-reu/tree/main/2024-projects/team-1`.

Dual-polarization radars gather data from both the horizontal and vertical polarizations, thus being able to better reflect the size, shape, distribution, and phase state of raindrops compared to previous, single-polarization radars [2,9]. In addition to reflectivity, dual-polarization radars provide differential reflectivity Z_{DR}, specific differential phase K_{DP}, and co-polar correlation coefficient ρ_{hv} (also referred to as CC) [9]. Z_{DR} is impacted by the composition or density of raindrops, helping differentiate water drops from ice pellets and snow. As the ratio between reflectivity factors at horizontal and vertical polarizations, Z_{DR} is not impacted by calibration errors, but may become biased with issues such as beam blockage [9]. K_{DP}, a derived variable that represents the change in differential phase shift Φ_{DP}, is useful for identifying heavy precipitation and when hail is mixed with rain, but K_{DP} can be more noisy in light rain [9]. K_{DP}, being immune to radar calibration errors, attenuation, and partial beam blockage, is a reliable factor for rainfall estimation [3]. The relationship between K_{DP} and rainfall has lower sensitivity to variations in drop size distribution than the Z–R relationship [2]. Moreover, as a measure of the variety of how particles affect the radar signals, ρ_{hv} is close to 1.0 during uniform rainfall and decreases with more variability in the types, shapes, and orientations of particles. ρ_{hv} is independent of particle concentrations and is immune to radar miscalibration, attenuation, and beam blockage [9]. As such, dual-polarization radar variables pose significant improvements for accurate precipitation estimation.

2.2 Symbolic Regression

Symbolic regression (SR) is a machine learning technique that finds an interpretable and best-fitting mathematical expression based on the data [4]. Popularity as well as advancements in computing have redefined SR and led to the rapid growth of related published papers in the past decade [10]. SR is usually implemented by evolutionary algorithms, specifically genetic programming, which constructs, compares, and combines different symbolic expressions to form a potential expression while discarding poor-performing combinations [10].

However, SR also presents some limitations for quantitative precipitation estimation. Symbolic regression methods may generate simple to complex equations that disagree with presently accepted knowledge regarding the relationships between radar data and rainfall rate. This prompts further research into the applicability and generalizability of symbolic regression methods to complex datasets.

2.2.1 Genetic Programming Symbolic Regression Models

We tested five genetic programming symbolic regression methods (gpg, gplearn, PySR, Feyn, and pyoperon), each with unique characteristics and approaches. Models such as gplearn [11] and gpg [12], the latter of which is a re-implementation of Gene-pool Optimal Mixing Evolutionary Algorithm (GP-GOMEA) focusing on symbolic regression, generate random sets of expressions and improve them to best fit the data using genetic concepts such as evolution, crossover, mutation, and population fitness. GP-GOMEA excels at generating simpler symbolic expressions through estimating and crossing over interdependencies between model components [12, 13]. In contrast, the C++ framework Operon [14], which is implemented in Python through pyoperon, focuses on execution speed and local search with gradient-based optimization, but tends to produce exceedingly complex models [15]. Packages like PySR [16] have made modifications to classic evolutionary algorithms and use a multi-population evolutionary algorithm that optimizes unknown scalar constants in newly discovered expressions. Similarly, Feyn [17] creates an evolutionary environment to simulate discrete paths from multiple inputs to an output, an idea inspired by Feynman's path integral. Random interactions are sampled onto paths, and with evolution and repeated reinforcement,

the best output is produced [17].

2.2.2 Deep Learning-Based and Other Symbolic Regression Models

We tested three additional symbolic regression methods (DSO, FFX, and RILS-ROLS), each of which applies different approaches to generate expressions learned from the data. Deep Symbolic Optimization (DSO) [18] combines symbolic regression with deep learning to leverage neural networks and a novel risk-seeking policy gradient to generate better-fitting expressions. Another method, Fast Function Extraction (FFX) [19], uses pathwise regularized machine learning to rapidly extract interpretable and simpler models. Moreover, RILS-ROLS [20] is a symbolic regression method that is built upon iterated local search and ordinary least squares to solve combinatorial aspects and determine best-fitting coefficients for equations, respectively.

3 Related Work

Previous research has found that machine learning and deep learning methods have resulted in improved QPE accuracy compared to conventional Z–R relationships. Huangfu et al. [21] found, in a study of twelve deep-learning-based QPE models, rainfall estimates were more accurate when distinguishing rainfall intensity using a K_{DP} threshold and when applying a self-defined loss function that gave varying weights to different intervals of rainfall intensity. Using dual-polarization radar variables as input data, Li et al. [22] and Wang and Chen [23] have found that QPEs derived from convolutional neural networks outperformed those derived from conventional Z–R relationships. Shin et al. [24] found that random forest and regression tree methods were also more accurate at estimating precipitation than the Z–R relationship. Moreover, Verdelho et al. [25] found that combining classification and regression techniques (random forest and gradient boosting) applied on dual-polarimetric radar variables outperformed Z–R relationships. They found that the models had varying accuracy across different rain intensity groups, performing best on moderate rain [25].

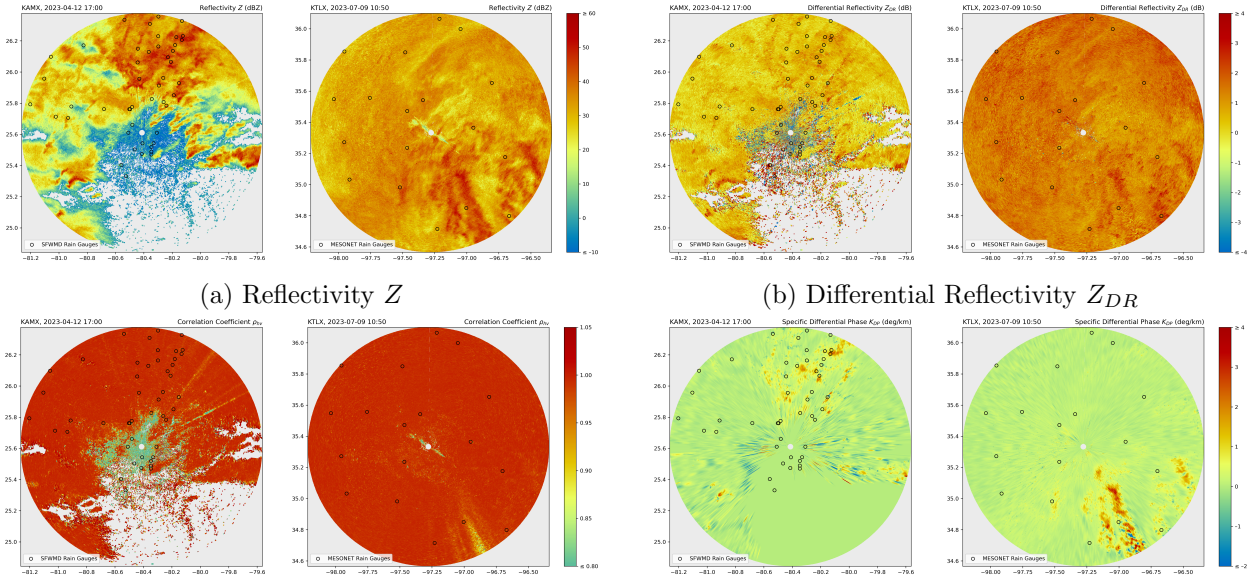
Previous studies have explored the role of clustering the data prior to applying symbolic regression. Sofos et al. [26] used k-medoids and agglomerative hierarchical clustering to separate a fluid simulation dataset into gas, liquid, and supercritical states prior to applying symbolic regression with PySR on each cluster. They found that clustering may reveal underlying nuances in the dataset for which symbolic regression identifies specific equations [26]. Building upon previous research, we test different symbolic regression methods and evaluate whether separating the data into groups based on the radar variables and rainfall intensity may reveal relationships between rainfall observations that can strengthen estimation accuracy.

4 Data

The source of dual-polarimetric radar data we used for this research is the Weather Surveillance Radar, 1988 Doppler (WSR-88D), also referred to as Next Generation Weather Radar (NEXRAD), operated across the United States by the National Weather Service. Our dataset consists of reflectivity Z_H (dBZ), differential reflectivity Z_{DR} (dB), specific differential phase K_{DP} (deg/km), and co-polar correlation coefficient ρ_{hv}, collected from the WSR-88D radars at Level II and 0.5 elevation angle. We explore radar data from South Florida, USA and Central Oklahoma, USA across three dates with significant precipitation. The data for South Florida are retrieved from the Miami KAMX WSR-88D radar and the data for Central Oklahoma are retrieved from the Oklahoma City KTLX WSR-88D radar.

The radar data are spatially and temporally merged with the rainfall rate (mm/hr) collected from rain gauge stations from the Oklahoma Mesonet and the South Florida Water Management District's DBHYDRO. We keep only observations where the rainfall rate is equal to 1 mm/hr or greater due to data quality concerns with trace amounts of rainfall. The range of rainfall rates are 1.010 mm/hr to 101.600 mm/hr, and the median value is 5.334 mm/hr. As shown in Figure 4.2, the rainfall rate has a right-skewed distribution with a few extreme values.

For Florida, we have data for every 15 minutes on April 12, 2023 (totalling 1,324 observations after cleaning the data), and for Oklahoma, we have data for every 5-10 minutes on July 9, 2023 and June 8, 2022 (totalling 1,406 observations after cleaning the data). Figure 4.1 shows an example of the radar data from the Miami WSR-88D radar (KAMX) and the Oklahoma WSR-88D radar (KTLX) for rainfall events on April 12, 2023 at 17:00 UTC and July 9, 2023 at 10:50 UTC, respectively. Rain gauge stations providing the ground truth data are displayed as black circular outlines. While the radar data are available over the observed area of Central Oklahoma and South Florida, the rain gauge data are only available at specific stations. We only use data located at these points for symbolic regression, which totals 2730 observations across the three days.



(a) Reflectivity Z

(b) Differential Reflectivity Z_{DR}

(c) Co-Polar Correlation Coefficient ρ_{hv}

(d) Specific Differential Phase K_{DP}

Figure 4.1: WSR-88D Radar Data of Rainfall Events in South Florida (KAMX) and Central Oklahoma (KTLX)

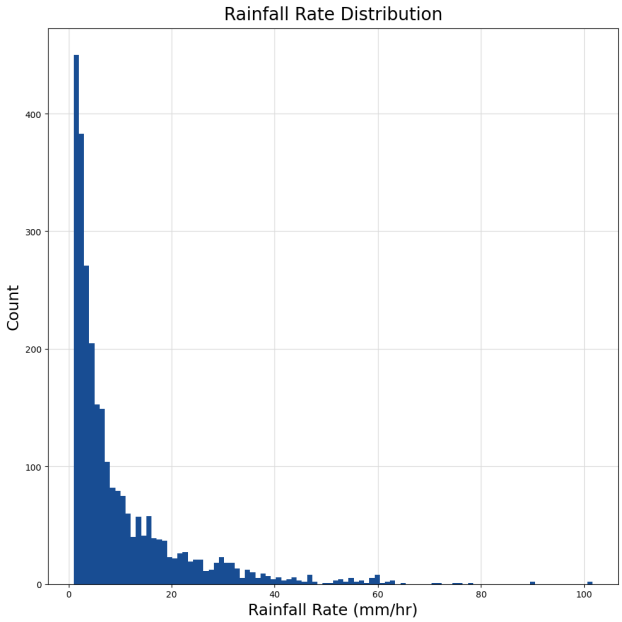


Figure 4.2: Distribution of rainfall rate

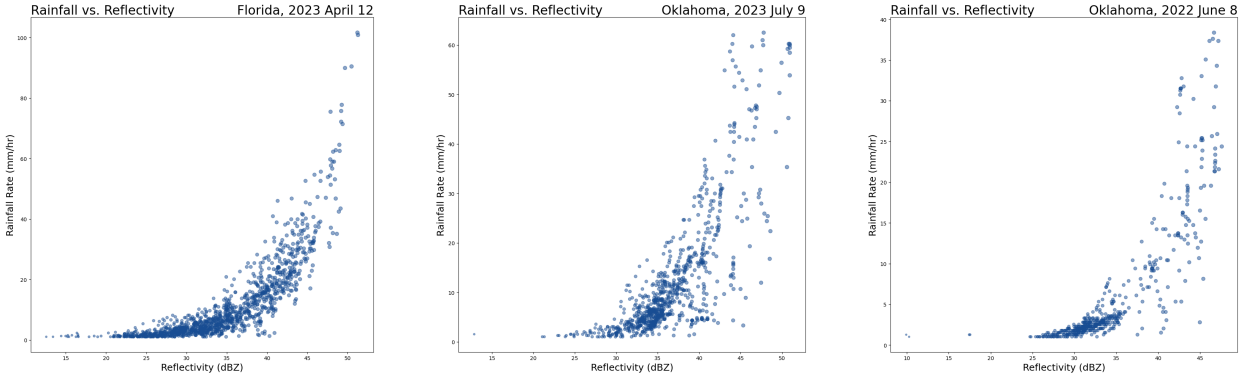


Figure 4.3: Reflectivity-rainfall relationship for three rainfall events

5 Methodology

5.1 Benchmarking Procedure

The procedure used to apply symbolic regression to the dual-polarimetric radar data was informed by SRBench [4], a benchmark of 14 different methods on over 200 datasets.

Following the procedure in SRBench, each symbolic regression model was trained on 75% of the data and tested on 25% of the data in ten repeated trials, with a different train-test split in each trial. The models were trained on the radar variables (Z, Z_{DR}, K_{DP}, ρ_{hv}) without scaling. Each model's hyperparameters were selected based on SRBench [4] and the model's respective documentation. We report two sets of metrics for PySR: one without restrictions on the complexity of the equation, and one with restrictions, the latter of which is subsequently referred to as PySR (simplified).

The training and testing R^2 scores, normalized root mean square error (NRMSE) scores, and simplicity scores were reported for each trial using the following equations, where k is the training

or testing size, y_i is the actual rainfall rate for observation i collected from rain gauges, $\hat{y}_i$ is the predicted rainfall rate for observation i from the learned equation, $\bar{y}$ is the mean rainfall rate for the training or testing set, and s is the number of components in the learned equation.

$$R^2 = 1 - \frac{\sum_{i=1}^{k}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{k}(y_i - \bar{y}_i)^2} \tag{5.1}$$

$$NRMSE = \frac{\sqrt{\frac{1}{k}\sum_{i=1}^{k}(y_i - \hat{y}_i)^2}}{\bar{y}} \tag{5.2}$$

$$simplicity = -log_5(s) \tag{5.3}$$

Metrics are reported based on the R^2 score (Equation 5.1) on the test data. NRMSE score (Equation 5.2) was used instead of root mean squared error to ensure that this metric was comparable across different training and testing sets that could have different distributions of rainfall rates. The simplicity score metric (Equation 5.3) was chosen from the SRBench Competition 2022 [5] to assign higher scores to simpler equations. The simplest equation (with only one component) has a simplicity score of 0, and as the equations become more complex, the simplicity decreases. Our objective is to find symbolic expressions with both high test R^2 scores and simplicity scores close to 0, representing an accurate yet interpretable equation. For our model performance to be comparable to existing models, we hope to achieve R^2 scores over 0.85.

5.2 Benchmarking Existing Symbolic Regression Models

To identify the most effective model for estimating precipitation rates from radar data, we tested various symbolic regression models by systematically comparing their performance. Benchmarking is crucial for ensuring that the selected models perform well on both training and unseen data. It allows us to understand the trade-offs between model complexity and prediction accuracy. We benchmarked five genetic programming symbolic regression algorithms (gpg, gplearn, PySR, Feyn, and pyoperon) and three non-genetic programming symbolic regression algorithms (DSO, FFX, and RILS-ROLS). We compared results from these symbolic regression methods to results from ordinary least squares linear regression, which creates a linear expression to fit the data by minimizing the sum of squares between the predicted rainfall rate and the ground-truth rainfall rate.

5.3 Symbolic Regression on Subsets of Data

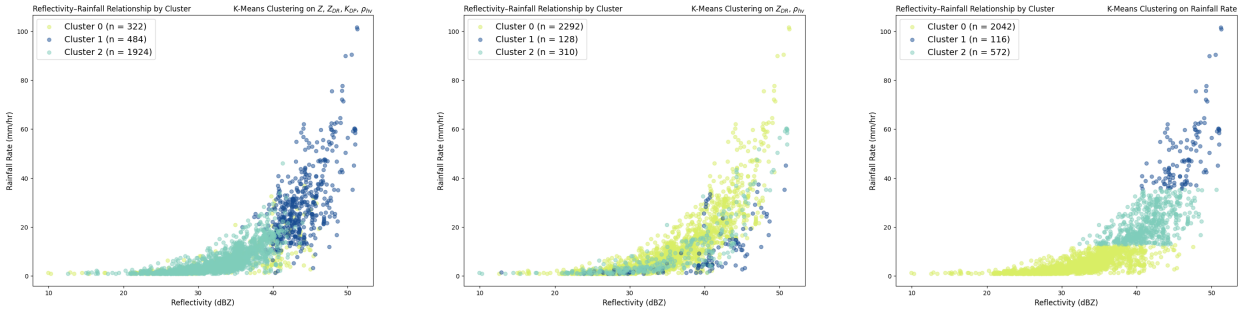
One significant challenge to quantitative precipitation estimation is the applicability of methods to different precipitation types and intensities. For example, the Z–R relationship to estimate rainfall from radar reflectivity varies across geographical region and type of rain. Previous research into quantitative precipitation estimation using machine learning and deep learning have found successful results when distinguishing the intensity of rainfall [21,25]. To test whether the accuracy and interpretability of the learned equations improve, we applied three methods to subset the data before applying symbolic regression on the separate subsets.

5.3.1 Symbolic Regression on Clusters

Using k-means clustering, bisecting k-means clustering, and agglomerative hierarchal clustering, we divided the data into three clusters based on all four radar variables (Z, Z_{DR}, K_{DP}, ρ_{hv}), based only on Z_{DR} and ρ_{hv}, and based on the rainfall rate. Figures 5.1, 5.2, and 5.3 show the Z–R relationship by cluster for all clustering methods. The clustering methods were implemented

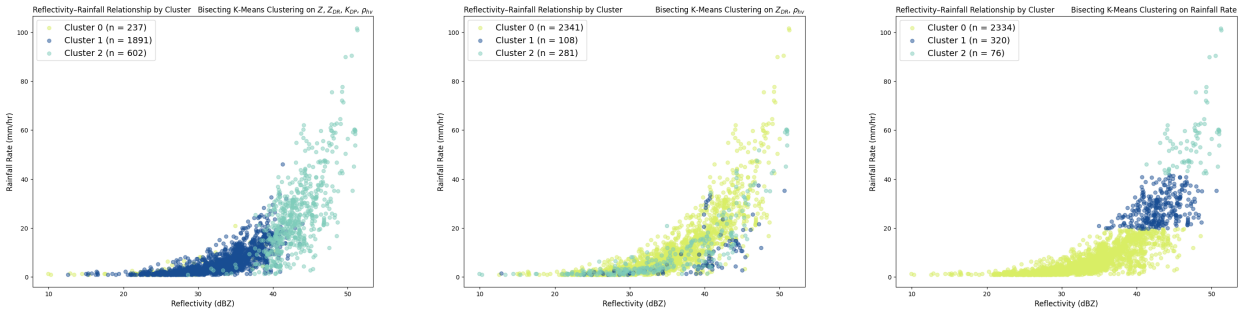
using the Python library `scikit-learn` [27]. K-means clustering is an algorithm that separates data into equal-variance subsets by minimizing the sum of squared distances between observations and the mean of their cluster [28]. Bisecting k-means clustering is a hybrid approach combining k-means and hierarchical clustering, where the entire dataset is split into two clusters, which are then split again until the set number of clusters is reached [28]. On the other hand, agglomerative clustering starts from the bottom with each data point being a cluster, which are consecutively merged together until a set number of clusters is reached. We applied agglomerative clustering using ward as the linkage criterion, which minimizes the increase in variance when two clusters are combined [28, 29].

After clustering the data into three clusters with each method, symbolic regression using Feyn was applied on the clusters separately following the benchmarking procedure from Section 5.1. Clustering was performed on data scaled to unit variance, while symbolic regression was applied on unscaled data to maintain the respective units of each variable.



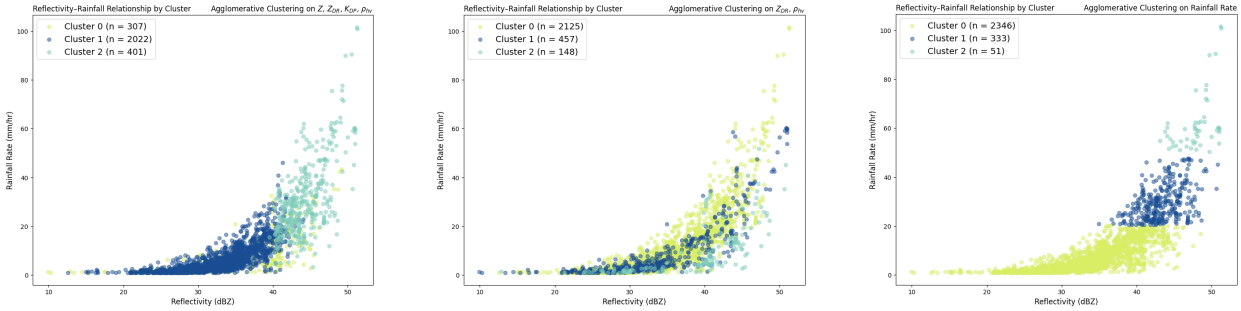
(a) With Z, Z_{DR}, K_{DP}, and ρ_{hv} (b) With Z_{DR} and ρ_{hv} (c) With rainfall rate

Figure 5.1: Reflectivity-rainfall relationship with k-means clustering



(a) With Z, Z_{DR}, K_{DP}, and ρ_{hv} (b) With Z_{DR} and ρ_{hv} (c) With rainfall rate

Figure 5.2: Reflectivity-rainfall relationship with bisecting k-means clustering



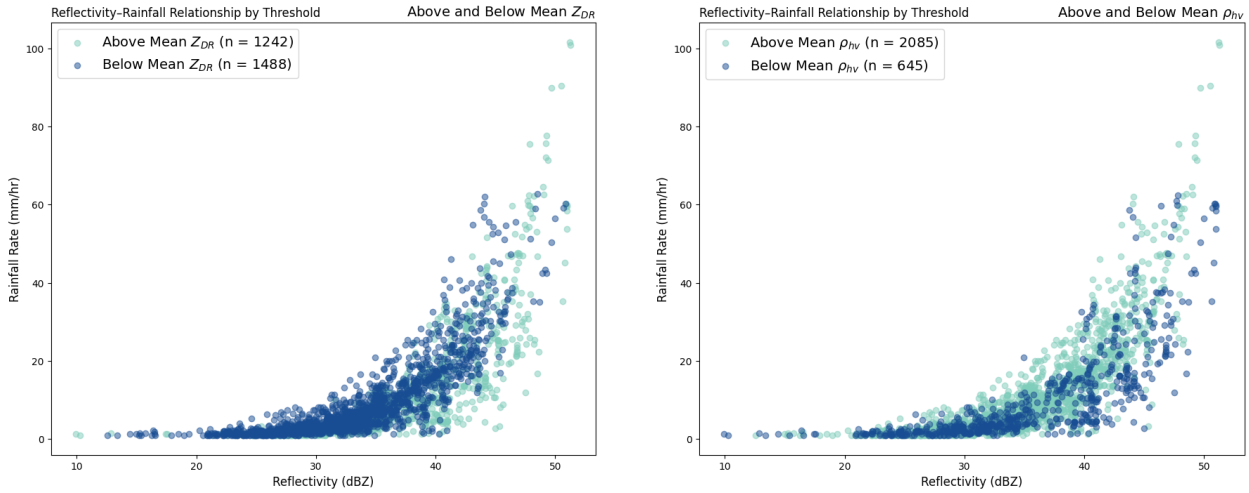
(a) With Z, Z_{DR}, K_{DP}, and ρ_{hv} (b) With Z_{DR} and ρ_{hv} (c) With rainfall rate

Figure 5.3: Reflectivity-rainfall relationship with agglomerative hierarchical clustering

5.3.2 Symbolic Regression by Radar Variables

Another method to subset the data prior to applying symbolic regression is based on the mean of the radar variables. We grouped the data into observations below and observations above the mean Z_{DR} (0.7641 dB), and followed the benchmarking procedure from Section 5.1 on the two groups. We then repeated this process for mean ρ_{hv} (0.9830).

Figure 5.4 shows the two groups for both Z_{DR} and ρ_{hv}. For both, separate patterns could be distinguished between the above-mean group and the below-mean group in the Z–R relationship. It is possible that dividing the data this way will allow for better-fitting equations to be learned within the two groups. Other radar variables, Z and K_{DP}, were not used to divide the data because there were no prominent patterns seen after separating the data based on the mean of these variables. Similarly, there were no prominent patterns seen after separating the data based on the median for all four radar variables.



(a) Mean Z_{DR} (b) Mean ρ_{hv}

Figure 5.4: Reflectivity-rainfall relationship for above and below mean Z_{DR} and mean ρ_{hv}

5.3.3 Symbolic Regression on Decision Tree Leaf Nodes

Decision trees are a type of supervised learning algorithm that recursively subsets data to create nodes with increasingly similar target values within each node and increasingly different target

values between nodes by learning decision rules based on the predictor values [28, 30]. Decision trees may work particularly well with symbolic regression due to their interpretability, breaking down complex relationships into simpler segments to train symbolic regression models.

Using decision trees in `scikit-learn`, we set the parameters to create three leaf nodes—subsets of the data that cannot be further divided—with at least 400 values in each. The decision trees first divided the data based on $K_{DP} \leq 0.367$, then based on $Z \leq 35.9$, as shown in Figure 5.5. We then followed the benchmarking procedure from Section 5.1 on the three leaf nodes.

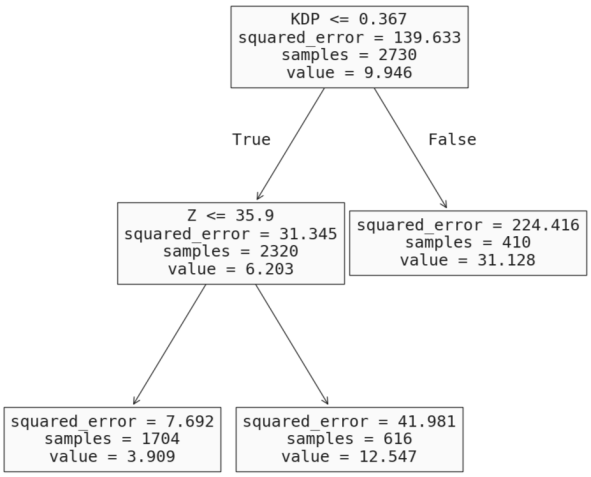


Figure 5.5: Steps taken by decision tree to separate the data into three leaf nodes

5.4 Exploring New Symbolic Regression Models

Machine learning models can be improved by incorporating prior knowledge into the training process [31]. This is especially useful in contexts of limited data where deep learning models capable of high accuracy are less feasible [32]. Adding knowledge-based loss terms to a model's loss function penalizes models that stray away from meeting the criteria that the target variable should follow. In addition to measuring the discrepancy between the ground-truth target variable and the predicted target variable, the updated loss function measures the deviation between the predicted target variable and the knowledge-based criteria,

$$\arg\min_{f} \text{Loss}\left(Y, \hat{Y}\right) + \lambda_D \text{Loss}_D\left(\hat{Y}\right) \tag{5.4}$$

where f is the model the machine learning algorithm is testing, Y is the ground-truth target variable, $\hat{Y}$ is the predicted target variable, and λ_D is the weight for the knowledge-based loss term [31]. The goal is to find the model f that minimizes this expression. In addition to using QPE-specific knowledge in loss functions, we incorporated knowledge learned from the data such as information from clustering the data prior to applying symbolic regression and from binning observations into groups by magnitude of rainfall rate. We integrated these new loss functions into the source code of gpg [12] with the goal of generating simpler symbolic expressions that perform better on the testing data. After applying these new loss terms, all final results were obtained following the benchmarking procedure from Section 5.1.

5.4.1 Implementing Z–R Relation Into the Loss Function

One format of knowledge-based loss terms is in the form of algebraic equations [31]. In the domain of QPE, this would mean we use the Z–R relation, $Z = aR^b$. Prior to implementing a new loss term for the Z–R relation, we needed to choose optimal values of a and b for our dataset through the following process:

1) Gather empirically derived values of a and b from previous studies.

2) Use optimization methods to find additional optimal values of a and b.

3) Evaluate the performance of each Z–R relationship gathered from the previous two steps on the entire dataset of size n. This involves computing $(\frac{Z_i}{a})^{(\frac{1}{b})}$ and comparing this result to the ground-truth value of R_i, $i = 1, 2, ..., n$.

4) Select the values of a and b with the highest performance. Incorporate these values into a Z–R relation loss term for a symbolic regression model.

Following the process above, we used the Marshall-Palmer relationship ($a = 200$ and $b = 1.6$) [33] and the Fulton et al. relationship ($a = 300$ and $b = 1.4$) [34] for 1), scipy's `optimize()` function ($a = 134$ and $b = 1.6$) for 2), and R^2 and NRMSE score for 3) (Table 5.1). We selected $a = 134$ and $b = 1.6$ as final values for a and b for 4), with R^2 and NRMSE scores of 0.7614 and 0.5804, respectively.

Values of a and b	R^2 Score	NRMSE Score
$a = 200$ and $b = 1.6$	0.6958	0.6552
$a = 300$ and $b = 1.4$	0.7389	0.6071
$a = 134$ and $b = 1.6$	**0.7614**	**0.5804**

Table 5.1: Results from using three different Z–R relations to predict rainfall rate

We then inserted the optimal values of a and b into a Z-R based loss function,

$$loss_f = \frac{1}{2k}\sum_{i=1}^{k}(y_i - \hat{y}_i)^2 + \frac{\lambda}{2k}\sum_{i=1}^{k}(\hat{y}_i - (\frac{Z_i}{134})^{\frac{1}{1.6}})^2 \tag{5.5}$$

where k is the training size, y_i is the ground-truth rainfall rate for observation i, $\hat{y}_i$ is the predicted rainfall rate for observation i, Z_i is the reflectivity for observation i, and λ is a weight parameter.

5.4.2 Implementing Silhouette Score Into the Loss Function

We incorporated the radar-based and rainfall-based clusters generated from k-means clustering, bisecting k-means clustering, and agglomerative hierarchical clustering into the training process of symbolic regression models. To reward models that predict rainfall rates aligning with the predetermined clusters, we subtracted the silhouette score multiplied by a weight term (λ) from the mean squared error term of the loss function. The silhouette score, a value from -1 to 1, measures how well each predicted rainfall rate is assigned to the cluster it is supposed to be a part of [35]. A higher score indicates better cluster assignment. This means the models that predict rainfall rates matching the clusters closely will have a lower error term associated with them, making them more

likely to be selected as the best model. Incorporating the silhouette score between the predicted rainfall rates and the predetermined cluster labels in the loss function, we obtain

$$loss_f = \frac{1}{2k}\sum_{i=1}^{k}(y_i - \hat{y}_i)^2 - \lambda * silhouette\_score(\hat{Y}, L) \tag{5.6}$$

as the updated loss term, where L denotes the cluster labels.

5.4.3 Implementing Deviations From Rainfall Group Into the Loss Function

Approximation constraints introduce reasonable ranges of the target variable to help train higher quality models [32]. Approximation constraints can be added to the loss function of machine learning models to penalize models that predict values of the target variables violating realistic bounds. Similar to approximation constraints, we binned precipitation rates in our dataset into three roughly equally sized groups: low precipitation in the range [1.00, 3.05), medium precipitation in the range [3.05, 9.15), and high precipitation in the range [9.15, 102.00). The bounds were chosen arbitrarily so that each group is of similar size, making each group evenly represented in the training process of symbolic regression models. We then included terms that increase the loss for models that predict rainfall rates not aligning with these three predetermined groups. Expanding on the work from [32], the approximation constraint loss function is

$$loss_f = \frac{1}{2k}\sum_{i=1}^{k}(y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^{3}\left[\sum_{i=1}^{k_j} ReLU(L_j - \hat{y_i^j}) + \sum_{i=1}^{k_j} ReLU(\hat{y_i^j} - U_j)\right] \tag{5.7}$$

where k_j is the number of observations from group $j = 1, 2, 3$, L_j is the lower bound of group j, U_j is the upper bound of group j, $\hat{y_i^j}$ is the predicted rainfall rate for observation i from group j, and $ReLU(x) = max(0, x)$. Only one weight term of λ was included for simplicity.

6 Results

6.1 Benchmarking Existing Symbolic Regression Models

The metrics for each model from the trial with the 5th highest train R^2 score[2] are shown in Table 6.1. Metrics from the trial with the highest test R^2 score are listed in Table 6.2, and the corresponding equations are listed in Table 6.3. All metrics are associated with one equation generated by the symbolic regression model. Distributions of test R^2 scores and the comparison with simplicity are shown in Figure 6.1.

The linear regression model achieving a low test R^2 score justifies the need for more complex expressions generated from symbolic regression. The model with the best performance regardless of model simplicity is pyoperon with a best test R^2 score of 0.9214. However, the equation has too low of a simplicity value to be used in operational QPE. Feyn and gpg also resulted in some of the highest test R^2 scores, but with complex equations. Thus, we decided that the best-performing model taking model simplicity into account is RILS-ROLS with a best test R^2 score of 0.9145 and a simplicity score of -1.9. Figure 6.3 shows the estimated rainfall rate across Central Oklahoma and South Florida using the RILS-ROLS equation.

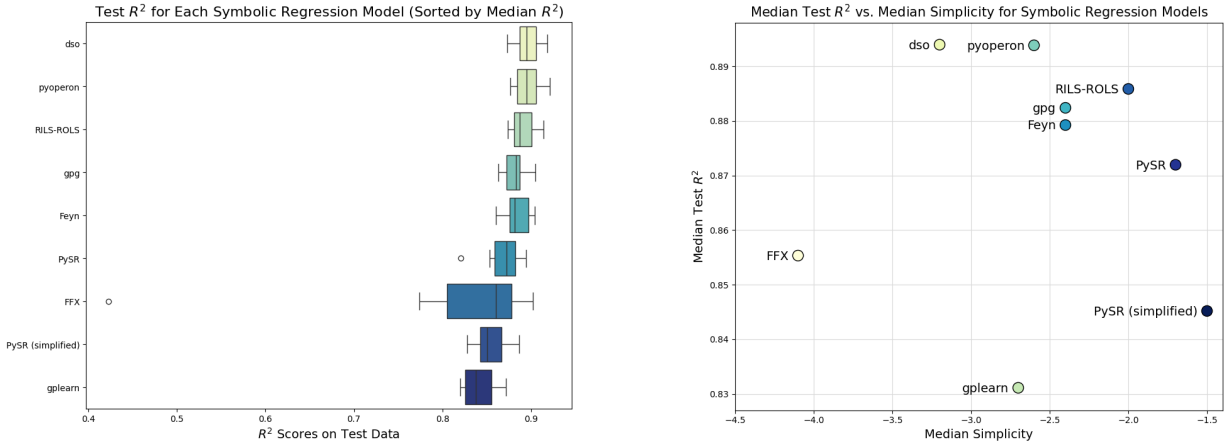
[2]The 5th highest train R^2 score is used as an approximation of the median of all ten trials while ensuring metrics from only one equation are reported.

Model	Train R^2	Test R^2	Train NRMSE	Test NRMSE	Simplicity
Linear Regression	0.8354	0.8521	0.4875	0.4393	-1.6
gplearn	0.8296	0.8311	0.4889	0.4903	-1.8
gpg	0.8824	0.8824	0.4018	0.4246	-2.4
PySR	0.8885	0.8720	0.3916	0.4420	-1.9
PySR (simplified)	0.8409	0.8452	0.4723	0.4695	**-1.5**
FFX	0.8971	0.8553	0.3761	0.4698	-4.2
Feyn	0.8940	0.8792	0.3864	0.4142	-2.4
RILS-ROLS	0.8944	0.8859	0.3904	0.3894	-2.0
DSO	0.8998	**0.8940**	0.3743	0.3920	-3.1
pyoperon	**0.9003**	0.8938	**0.3701**	**0.3782**	-2.5

Table 6.1: Results from the trial with the median test R^2 score for each model

Model	Train R^2	Test R^2	Train NRMSE	Test NRMSE	Simplicity
Linear Regression	0.8229	0.8753	0.4912	0.4364	-1.6
gplearn	0.8117	0.8721	0.5065	0.4421	-1.8
gpg	0.8744	0.9049	0.4115	0.3842	-2.5
PySR	0.8533	0.8943	0.4446	0.4050	**-1.5**
PySR (simplified)	0.8291	0.8864	0.4825	0.4166	**-1.5**
FFX	0.8840	0.9020	0.3955	0.3899	-4.0
Feyn	0.8757	0.9046	0.4116	0.3817	-2.4
RILS-ROLS	0.8824	0.9145	0.4003	0.3900	-1.9
DSO	**0.8909**	0.9188	**0.3856**	**0.3523**	-3.2
pyoperon	0.8882	**0.9214**	0.3902	0.3787	-2.6

Table 6.2: Results from the trial with the highest test R^2 score for each model



(a) Distribution of test R^2 scores

(b) Test R^2 scores compared with simplicity

Figure 6.1: Accuracy and simplicity of symbolic regression models

Model	Equation
Linear Regression	$R = 0.5070Z - 2.082Z_{DR} + 35.16K_{DP} + 22.99\rho_{hv} - 34.48$
gplearn	$R = \sqrt{\left\lvert K_{DP}(\rho_{hv}^4)(Z)\sqrt{\left\lvert K_{DP}(\rho_{hv}^2)(Z^2)\right\rvert}\right\rvert}$
GP-GOMEA	$R = -0.0095ZZ_{DR}\rho_{hv}\frac{(Z_{DR}-5.0785)}{(\rho_{hv}-4.5774)} + 0.0047Z\rho_{hv}^2(Z + K_{DP})(K_{DP} + 0.5812)(\rho_{hv} + cos(Z_{DR}) + cos(K_{DP} - \rho_{hv}) + 0.8610) + (0.0047K_{DP} + 0.0070)\frac{-cos(\rho_{hv})+\frac{K_{DP}}{Z_{DR}}}{cos(Z)} - 4.5493$
PySR	$R = \left\lvert 0.8496^{Z_{DR}}(Z)(K_{DP} + 0.4819) - 10.92\right\rvert$
PySR (simplified)	$R = Z^{\rho_{hv}}(K_{DP} + 0.1387) - Z_{DR}$
FFX*	$R = -4.78 - 28000\max(0, K_{DP} - 1.46)\max(0, \rho_{hv} - 0.987) - 17400\max(0, 0.940 - \rho_{hv})\max(0, 0.956 - \rho_{hv}) - 16900\max(0, \rho_{hv} - 1)\max(0, \rho_{hv} - 0.987) + 14900\max(0, 0.0308 - K_{DP})\max(0, \rho_{hv} - 1) + ...$
Feyn*	$R = 103.7e^{-2.381(0.34K_{DP}+0.9164e^{-34.62(0.8061\rho_{hv}-1)^2-0.5622(Z_{DR}-0.16)^2}-1)^2+...}$
RILS-ROLS	$R = 1.208Z(K_{DP})\rho_{hv}^3 - 20.09K_{DP} + 2(10^{-6})\rho_{hv}^4Z^4e^{\cos(Z_{DR})} - 0.6427$
DSO*	$R = \log(\rho_{hv} + e^{\rho_{hv}(5.355K_{DP}^3-56.22K_{DP}^2(\rho_{hv})-1.577Z(K_{DP}^2)+6.587Z_{DR}(K_{DP}^2)+...}$
pyoperon	$R = -0.044 + (-39.35K_{DP}(\rho hv) - 760.3K_{DP} + \frac{797.1K_{DP}-0.052Z}{\sqrt{0.004Z_{DR}^2+1}} - \sin(0.149Z))(\frac{1}{\sqrt{0.487Z_{DR}^2+1}})(1.002(\frac{-4.72K_{DP}-0.168Z}{\sqrt{0.8118Z_{DR}^2+1}} - \sin(0.208Z)))$

*Equations truncated due to length.

Table 6.3: Equation from the trial with the highest test R^2 score for each model

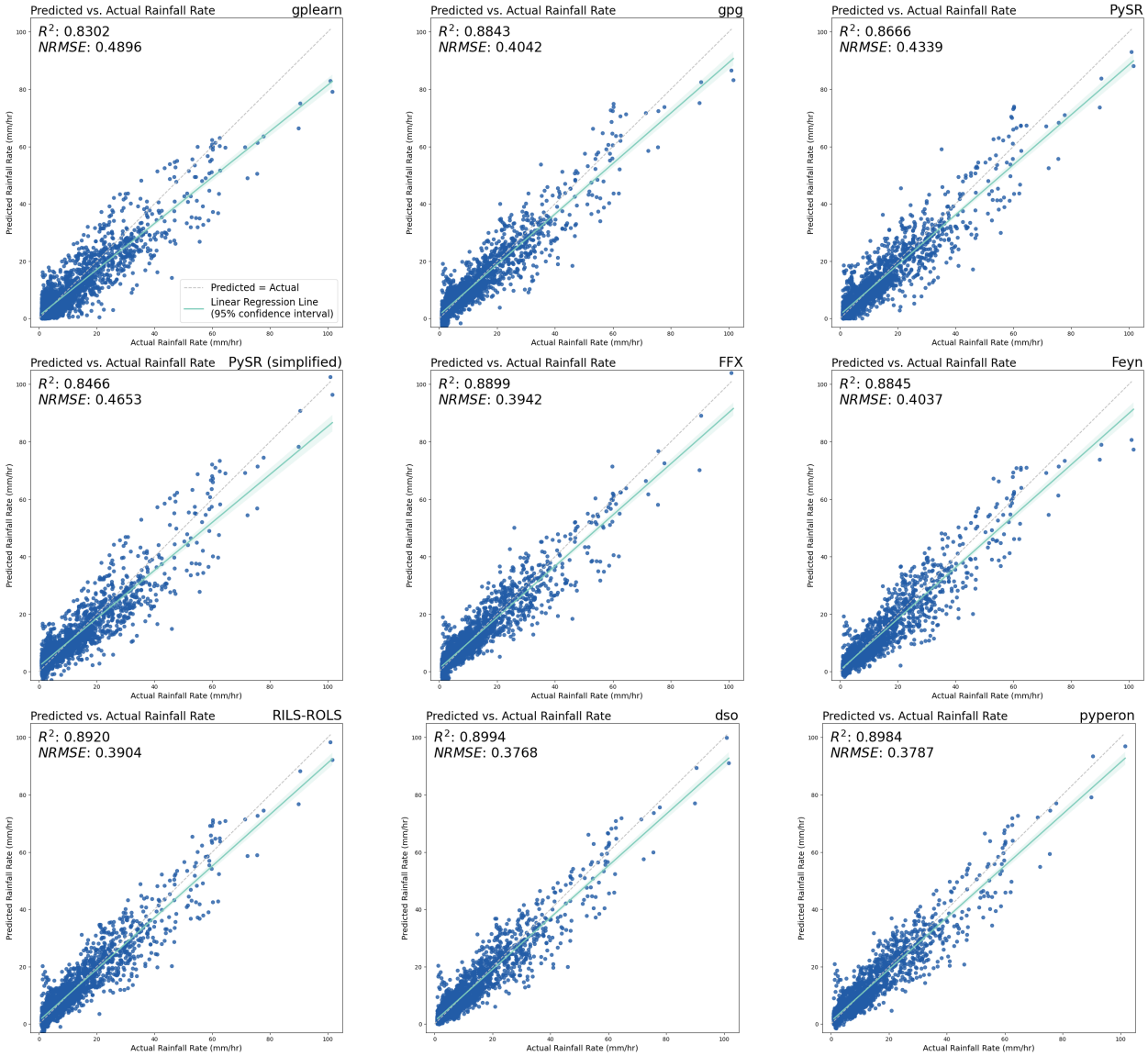


Figure 6.2: Predicted vs. actual rainfall rate across different symbolic regression methods

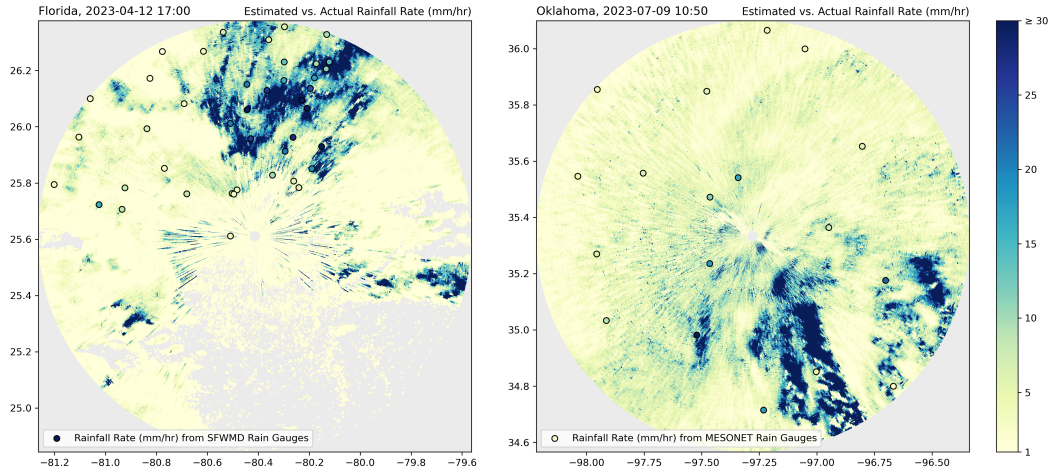


Figure 6.3: Map of estimated rainfall rate from RILS-ROLS equation (Table 6.3)

6.2 Symbolic Regression on Subsets of Data

In Section 6.1, we identified that Feyn achieved one of the highest test R^2 scores out of the models but had a lower simplicity score, so we used Feyn for symbolic regression on subsets of data to see if the R^2 and simplicity could be improved.

6.2.1 Symbolic Regression on Clusters

Table 6.4 shows the results from the trial with the highest test R^2 score; the metrics reported are the mean from the three clusters for each clustering method.

The best result came from k-means clustering based on Z_{DR} and ρhv, which resulted in a mean test R^2 of 0.9200. The train and test R^2 score and simplicity score are slightly higher than the metrics from Feyn applied on the entire dataset without initial clustering. This indicates that clustering based on Z_{DR} and ρhv may break down the data into sections with stronger relationships that are learned by symbolic regression. However, it may be harder to generalize to new data within these clusters.

Cluster	Variable	Train R^2	Test R^2	Train NRMSE	Test NRMSE	Simplicity
All Data (Without Clustering)		0.8757	0.9046	0.4116	0.3817	-2.4
K-Means	All Radar	0.7382	0.7826	0.4142	0.3784	-2.1
	Z_{DR} and ρ_{hv}	0.9048	**0.9200**	0.3605	0.3250	-2.2
	Rain	0.6318	0.6764	0.2456	0.2434	-2.1
Bisecting K-Means	All Radar	0.7129	0.7527	0.4300	0.3936	**-2.0**
	Z_{DR} and ρ_{hv}	**0.9064**	0.8980	0.3554	0.3722	-2.1
	Rain	0.6106	0.6069	0.2214	0.2265	**-2.0**
Agglomerative	All Radar	0.7556	0.7887	0.4006	0.3727	-2.1
	Z_{DR} and ρ_{hv}	0.8973	0.8746	0.3758	0.3980	-2.3
	Rain	0.6466	0.6623	**0.2201**	**0.2248**	**-2.0**

Table 6.4: Mean metrics of three clusters from the trial with the highest test R^2 score using Feyn

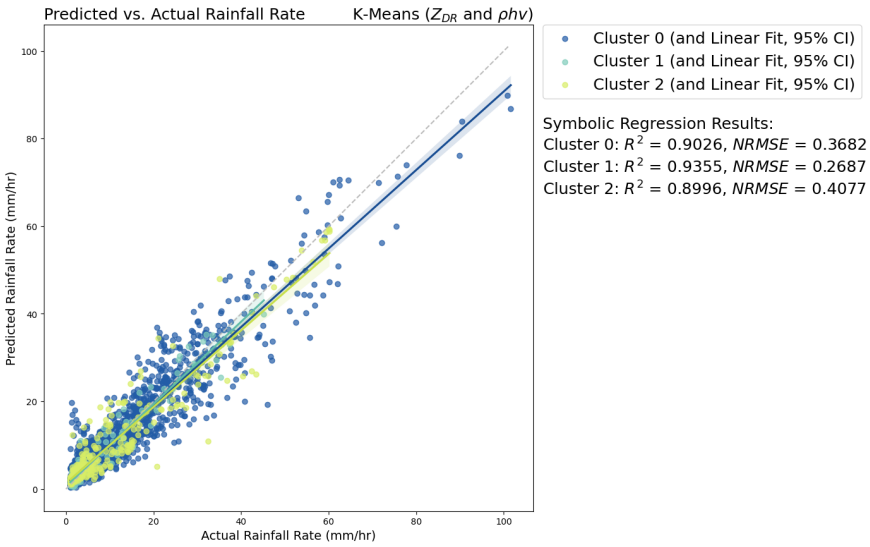


Figure 6.4: Actual vs. predicted rainfall of each cluster identified by k-means clustering on Z_{DR} and ρhv

Cluster	Size	Train R^2	Test R^2	Train NRMSE	Test NRMSE	Simplicity
All Data	2730	0.8757	0.9046	0.4116	0.3817	-2.4
0	2292	0.8979	0.9125	0.3767	0.3458	-2.5
1	128	0.9282	0.9581	0.2849	0.2134	-2.2
2	310	0.8778	0.9473	0.4446	0.3010	-1.7

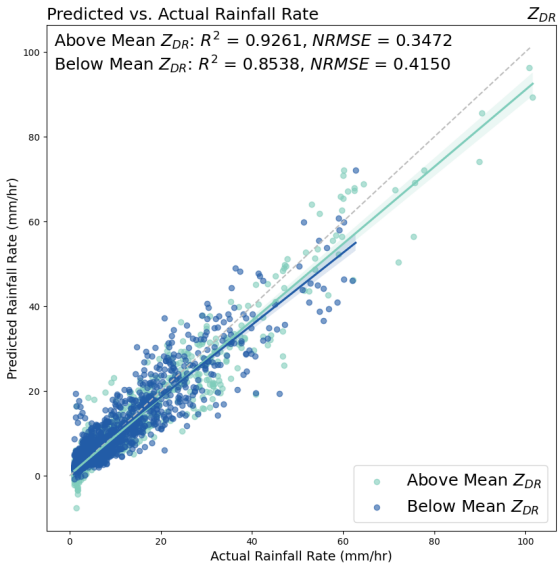
Table 6.5: Metrics from trial with highest test R^2 score for each cluster identified by k-means clustering on Z_{DR} and ρhv

6.2.2 Symbolic Regression by Radar Variables

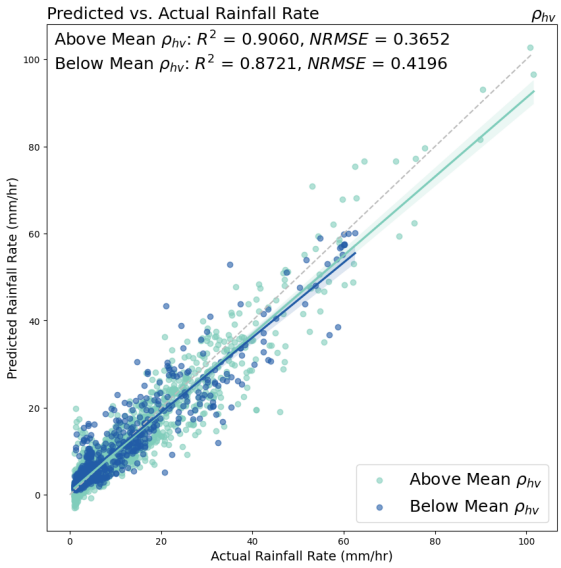
Table 6.6 shows the results from applying symbolic regression on the data divided into two groups based on mean Z_{DR}, and the same for ρ_{hv}. The above mean Z_{DR} group achieved the highest test R^2 score of 0.9519, indicating that radar observations in this group had a strong relationship with the rainfall rate. However, the below mean Z_{DR} group performed much worse with a test R^2 score of 0.8538. This may indicate that in our dataset, the radar data associated with higher Z_{DR} values estimate rainfall rates better than those with lower Z_{DR} values. Similarly, there were slight improvements in the above mean ρ_{hv} group, and worse performance in the below-mean group, compared with symbolic regression on the ungrouped data.

Variable	Group	Size	Train R^2	Test R^2	Train NRMSE	Test NRMSE	Simplicity
All Data		2730	0.8757	0.9046	0.4116	0.3817	-2.4
Z_{DR}	Above Mean	1242	**0.9161**	**0.9519**	0.3620	**0.2976**	-2.2
	Below Mean	1488	0.8538	0.8538	0.4096	0.4314	-2.3
ρ_{hv}	Above Mean	2085	0.9023	0.9132	**0.3554**	0.3898	-2.1
	Below Mean	645	0.8586	0.9025	0.4268	0.3985	-2.4

Table 6.6: Metrics from the trial with the highest test R^2 score for each group using Feyn



(a) Mean Z_{DR}



(b) Mean ρ_{hv}

Figure 6.5: Predicted vs. actual rainfall rate by threshold set by mean Z_{DR} and mean ρ_{hv}

6.2.3 Symbolic Regression on Decision Tree Leaf Nodes

The metrics for each leaf node from the trial with the highest test R^2 score are listed in Table 6.7, and a comparison of the actual and predicted rainfall values is shown in Figure 6.6.

Subset	Size	Train R^2	Test R^2	Train NRMSE	Test NRMSE	Simplicity
All Data	2730	0.8757	0.9046	0.4116	0.3817	-2.4
Node 1	1704	0.5775	0.6330	0.4620	0.4271	-2.1
Node 2	616	0.4510	0.5624	0.3923	0.3123	-2.1
Node 3	410	0.7773	0.7826	0.2277	0.2199	-2.0

Table 6.7: Metrics from the trial with the highest test R^2 score for each node using Feyn

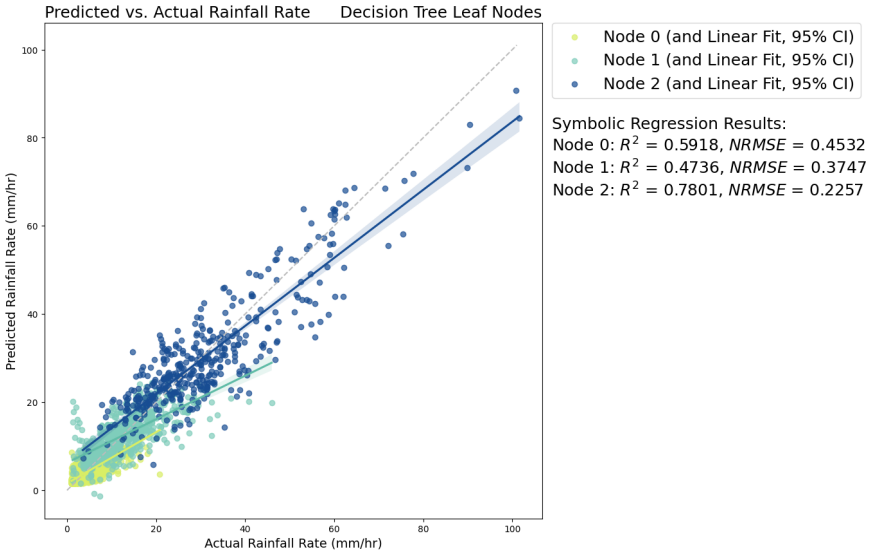


Figure 6.6: Actual vs. predicted rainfall of each node identified by decision trees

The R^2 scores for the nodes are significantly lower than the score achieved without subsetting the data, while the learned equations are simpler. Although grouping the data using decision trees prior to applying symbolic regression failed to achieve a higher R^2 score, decision trees introduced an interesting approach to dividing the data, using both K_{DP} and Z as described in Section 5.3.3, which may be valuable for further study.

6.3 Exploring New Symbolic Regression Models

The results from the trial with the best test R^2 score from modifying the loss function of GP-GOMEA are found in Table 6.8 and Table 6.9.

Including the Z–R relation in the loss function with $\lambda = 1$ worsened the original results. Different values of λ produced similar results. This is likely because the Z–R relationship is not very accurate in predicting rainfall to begin with, thus adding unnecessary noise to the loss function.

Including the silhouette score term in the loss function from k-means clustering by Z_{DR} and ρ_{hv} as in Section 6.2.1 slightly improved the testing metrics while leaving the training metrics relatively unchanged. By identifying different clusters and predicting rainfall rates in the appropriate magnitudes, this clustering-based term was able to generate a model that fit slightly better to the data. Including the silhouette score term from other methods of clustering and clustered by different variables did not change or worsened the original results.

Including the binned rainfall term in the loss function with $\lambda = 0.01$ was able to slightly improve model performance in addition to generate a simpler equation. Other values of λ either worsened or did not change the original results. Incorporating bounds that the rainfall rates should follow within the loss function could have guided the models in the right direction to predict rainfall rates from unseen data in the appropriate magnitudes. As shown in Table 6.8, when the model generated from the binned rainfall loss term was evaluated on the whole dataset, it produced test R^2 and NRMSE scores of 0.9067 and 0.3804, compared to test R^2 and NRMSE scores of 0.9049 and 0.3842 for the original method, respectively.

Loss Function	Train R^2	Test R^2	Train NRMSE	Test NRMSE	Simplicity
Original	0.8744	0.9049	0.4115	0.3842	-2.5
Z-R ($\lambda = 1$) (Equation 5.4.1)	0.8546	0.8900	0.4427	0.4132	**-2.3**
Silhouette score ($\lambda = 20$) (Equation 5.4.2)	0.8746	0.9060	0.4110	0.3819	-2.5
Binned rainfall ($\lambda = 0.01$) (Equation 5.4.3)	**0.8748**	**0.9067**	**0.4108**	**0.3804**	**-2.3**

Table 6.8: Results from the model with the highest test R^2 score using gpg

Loss Function	Equation
Original	$R = -0.0095ZZ_{DR}\rho_{hv}\frac{(Z_{DR}-5.0785)}{(\rho_{hv}-4.5774)} + 0.0047Z\rho_{hv}^2(Z + K_{DP})(K_{DP} + 0.5812)(\rho_{hv} + cos(Z_{DR}) + cos(K_{DP} - \rho_{hv}) + 0.8610) + (0.0047K_{DP} + 0.0070)\frac{-cos(\rho_{hv})+\frac{K_{DP}}{Z_{DR}}}{cos(Z)} - 4.5493$
Z–R relation ($\lambda = 1$)	$R = 0.0016Z\rho_{hv}^2(Z - K_{DP})(K_{DP} + 0.59340)(\rho_{hv} + cos(Z_{DR}) + cos(K_{DP} - \rho_{hv}) + 6.9488) - 0.3357sin(cos(1) + \frac{6.5855}{(Z_{DR}-2.7016)}) - 4.3605$
Silhouette score ($\lambda = 20$)	$R = -0.0149ZZ_{DR}\rho_{hv}\frac{(Z_{DR}-4.9152)}{(\rho_{hv}-4.5719)} + 0.0043Z\rho_{hv}^2(Z + K_{DP})(K_{DP} + 0.6023)(\rho_{hv} + cos(Z_{DR}) + cos(K_{DP} - \rho_{hv}) + 1.2032) - (0.0137K_{DP} + 0.0017)\frac{(cos(\rho_{hv})-\frac{K_{DP}}{Z_{DR}})}{cos(Z)} - 4.7574$
Binned rainfall ($\lambda = 0.01$)	$R = 0.0051Z\rho_{hv}^2(Z - K_{DP})(K_{DP} + 0.4857)(\rho_{hv} + cos(Z_{DR}) + cos(K_{DP} - \rho_{hv}) + 0.9265) - 0.7527sin(cos(1) + \frac{5.9838}{(Z_{DR}-2.6914)}) - 4.5027$

Table 6.9: Equation from the model with the highest test R^2 score using GP-GOMEA

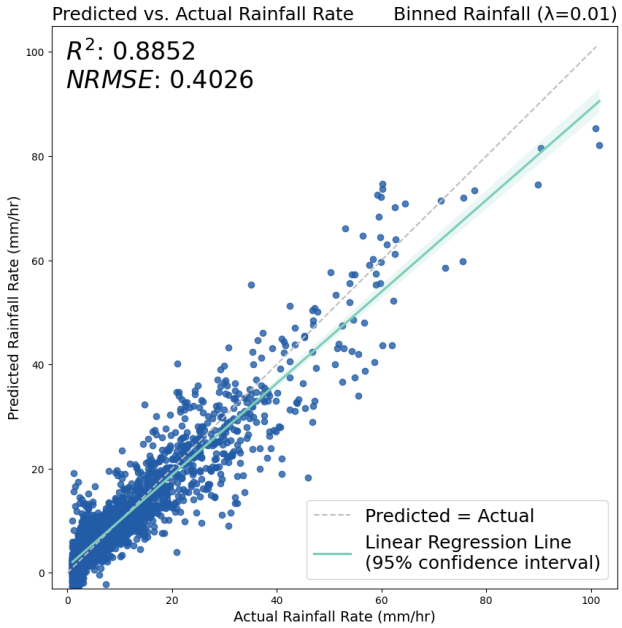


Figure 6.7: Results from using the model generated from a binned rainfall loss term ($\lambda = 0.01$)

7 Evaluation from Radar Meteorology

7.1 Potential Equations for Operational Applications

Among all the SR algorithms, the simplest equations were produced by gplearn, PySR, and RILS-ROLS, as shown in Table 6.3. Compared with the conventional Z–R relationships shown in Table 5.1, the equation from gplearn performed slightly better, and the equations from PySR and RILS-ROLS showed significant improvement, with acceptable test R^2 values of 0.8943 and 0.9145, respectively.

The PySR equation consists of two stages. When the first term is less than 10.92, the QPE tends to decrease as Z and K_{DP} increase, which contradicts the expected relationship between rainfall and radar measurements. Further research could explore the significance of the absolute operation and whether it consistently aligns with physical understanding. When the first term exceeds 10.92, the equation's relationship aligns with domain knowledge: the QPE increases as Z and K_{DP} increase. For Z_{DR}, it is represented as a decreasing exponential function in this equation. These trends are consistent with previous studies [36, 37].

The PySR (simplified) equation, which includes all four radar variables, indicates that the QPE increases as Z and K_{DP} increase, which similarly aligns with domain knowledge. However, the equation indicates that as Z_{DR} increases, QPE decreases the same amount, which does not fully capture the relationship between Z_{DR} and rainfall rate.

In the RILS-ROLS equation, all variables are utilized. The equation consists of three terms: the first two terms reflect the interaction between Z and K_{DP}, while ρhv influences the strength of this interaction. Since ρhv is crucial for identifying precipitation types, its role in the interaction between Z and K_{DP} can be seen as a representation of a fuzzy classification strategy within the equation [38]. The third term is more complex, and further evaluation is needed to understand its contribution to the final estimation.

The equations demonstrate promising accuracy and lower errors compared to traditional QPE equations. Their potential for operational applications is evident, provided that contradictions within the equations with domain knowledge are addressed.

7.2 Insights from Complex Equations

The equations with more complex terms provide us with new insights about the relationship between radar and rainfall. The presence of trigonometric functions suggests that new combinations and transformations on existing radar products may play an important role. While the FFX equation is more complex, most of the terms are related to the max operation between 0 and the radar variables. The equation can be manually simplified based on the classification of different thresholds set by the radar variables ρhv and K_{DP} [39]. This classification strategy indeed aligns with the ability of ρhv and K_{DP}. This emphasizes the importance of the classification of precipitation types, as well as determining the drop size and density.

8 Conclusions and Future Work

Symbolic regression is an effective approach to quantitative precipitation estimation given its interpretability and accuracy. All of the symbolic regression methods we tested (Table 6.1) resulted in higher accuracy than conventional Z–R relationships (Table 5.1). Of the symbolic regression models, we decided that the best model for estimating rainfall rate from the four radar variables is RILS-ROLS due to its high test R^2 score and simplicity.

We tested Feyn symbolic regression on different subsets of the data with the goal of better quantifying relationships among different precipitation intensities and types. Applying Feyn on three clsters resulting from k-means clustering based on Z_{DR} and ρhv achieved improved R^2 scores, lower NRMSE scores, and slightly simpler equations. Symbolic regression on clustered data by all four radar variables, on clustered data by rainfall rate, and on decision tree-identified nodes resulted in worse scores, but slightly simpler equations. Dividing the data based on mean Z_{DR} resulted in improved performance for observations in the above-mean group, but worse performance in the below-mean group. Overall, the radar variables provide insights into complex relationships with rainfall rate, which can be learned by symbolic regression.

The models generated from adding knowledge-based loss terms to the loss function of GP-GOMEA was able to improve the original results slightly. This demonstrates that including cluster-based loss terms generated from unsupervised algorithms prior to training symbolic regression models as well as including penalties when the model predicts a rainfall rate outside of its appropriate range in the overall loss function have potential for improving the accuracy of machine learning models. However, improved models were not always able to be reproduced when retraining the model due to variation in the GP-GOMEA model itself. Additionally, whether modifying the loss function of other symbolic regression algorithms will improve the original results is unknown.

This study can be built upon through future work by testing symbolic regression on a larger dataset encompassing more geographic regions and dates, as well as considering the time of the observations when training symbolic regression models. Our dataset for this study is limited and only includes data from Florida and Oklahoma. Future research will aim to train symbolic regression models with robust datasets that can be generalizable to other regions. There is also more work to be done in analyzing how to incorporate domain knowledge into the loss function of symbolic regression models to improve learned equations.

Acknowledgment

This work is supported by the grant "REU Site: Online Interdisciplinary Big Data Analytics in Science and Engineering" from the National Science Foundation (grant no. OAC–2348755).

The hardware used in the computational studies is part of the UMBC High Performance Computing Facility (HPCF). The facility is supported by the U.S. National Science Foundation through the MRI program (grant nos. CNS–0821258, CNS–1228778, OAC–1726023, and CNS–1920079) and the SCREMS program (grant no. DMS–0821311), with additional substantial support from the University of Maryland, Baltimore County (UMBC). See `hpcf.umbc.edu` for more information on HPCF and the projects using its resources.

References

[1] D. Wijayarathne, P. Coulibaly, S. Boodoo, and D. Sills, "Use of radar quantitative precipitation estimates (QPEs) for improved hydrological model calibration and flood forecasting," *Journal of Hydrometeorology*, vol. 22, no. 8, pp. 2033–2053, 2021.

[2] R. Cifelli and V. Chandrasekar, *Dual-Polarization Radar Rainfall Estimation.* American Geophysical Union (AGU), 2010, pp. 105–125.

[3] A. Ryzhkov, P. Zhang, P. Bukovčić, J. Zhang, and S. Cocks, "Polarimetric radar quantitative precipitation estimation," *Remote Sensing*, vol. 14, no. 7, p. 1695, 2022.

[4] W. La Cava, B. Burlacu, M. Virgolin, M. Kommenda, P. Orzechowski, F. O. de França, Y. Jin, and J. H. Moore, "Contemporary symbolic regression methods and their relative performance," *Advances in Neural Information Processing Systems*, vol. 2021, no. DB1, p. 1, 2021.

[5] ——, "2022 SRBench competition: Judging criteria," 2022. [Online]. Available: https://cavalab.org/srbench/judging-criteria/

[6] L. C. Sieck, S. J. Burges, and M. Steiner, "Challenges in obtaining reliable measurements of point rainfall," *Water Resources Research*, vol. 43, no. 1, Jan 2007.

[7] S. Trömel, C. Chwala, C. Furusho-Percot, C. C. Henken, J. Polz, R. Potthast, R. Reinoso-Rondinel, and C. Simmer, "Near-realtime quantitative precipitation estimation and prediction (RealPEP)," *Bulletin of the American Meteorological Society*, vol. 102, no. 8, pp. E1591–E1596, 2021.

[8] W. Krajewski and J. Smith, "Radar hydrology: Rainfall estimation," *Advances in Water Resources*, vol. 25, no. 8, pp. 1387–1394, 2002.

[9] M. R. Kumjian, "Principles and applications of dual-polarization weather radar. Part I: Description of the polarimetric radar variables," *Journal of Operational Meteorology*, vol. 1, 2013.

[10] D. Angelis, F. Sofos, and T. E. Karakasidis, "Artificial intelligence in physical sciences: Symbolic regression trends and perspectives," *Archives of Computational Methods in Engineering*, vol. 30, no. 6, pp. 3845–3865, 2023.

[11] T. Stephens, "Genetic programming in Python, with a scikit-learn inspired API," https://gplearn.readthedocs.io/en/stable/intro.html.

[12] M. Virgolin, T. Alderliesten, C. Witteveen, and P. A. N. Bosman, "Improving model-based genetic programming for symbolic regression of small expressions," *Evolutionary Computation*, vol. 29, no. 2, pp. 211–237, 2021.

[13] M. Virgolin, T. Alderliesten, C. Witteveen, and P. A. Bosman, "Scalable genetic programming by gene-pool optimal mixing and input-space entropy-based building-block learning," in *Proceedings of the Genetic and Evolutionary Computation Conference*, 2017, pp. 1041–1048.

[14] B. Burlacu, G. Kronberger, and M. Kommenda, "Operon C++: An efficient genetic programming framework for symbolic regression," in *Proceedings of the 2020 Genetic and Evolutionary Computation Conference Companion*, ser. GECCO '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 1562–1570.

[15] B. Burlacu, "Gecco'2022 symbolic regression competition: Post-analysis of the Operon framework," in *Proceedings of the Companion Conference on Genetic and Evolutionary Computation*, 2023, pp. 2412–2419.

[16] M. Cranmer, "Interpretable machine learning for science with PySR and SymbolicRegression.jl," *arXiv preprint arXiv:2305.01582*, 2023.

[17] K. R. Broløs, M. V. Machado, C. Cave, J. Kasak, V. Stentoft-Hansen, V. G. Batanero, T. Jelen, and C. Wilstrup, "An approach to symbolic regression using Feyn," *arXiv preprint arXiv:2104.05417*, 2021.

[18] B. K. Petersen, M. Landajuela, T. N. Mundhenk, C. P. Santiago, S. K. Kim, and J. T. Kim, "Deep symbolic regression: Recovering mathematical expressions from data via risk-seeking policy gradients," in *Proc. of the International Conference on Learning Representations*, 2021.

[19] T. McConaghy, "FFX: Fast, scalable, deterministic symbolic regression technology," *Genetic Programming Theory and Practice IX*, pp. 235–260, 2011.

[20] A. Kartelj and M. Djukanović, "RILS-ROLS: Robust symbolic regression via iterated local search and ordinary least squares," *Journal of Big Data*, vol. 10, no. 71, 2023.

[21] J. Huangfu, Z. Hu, J. Zheng, L. Wang, and Y. Zhu, "Study on quantitative precipitation estimation by polarimetric radar using deep learning," *Advances in Atmospheric Sciences*, pp. 1–14, 2024.

[22] W. Li, H. Chen, and L. Han, "Polarimetric radar quantitative precipitation estimation using deep convolutional neural networks," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–11, 2023.

[23] L. Wang and H. Chen, "Machine learning for polarimetric radar quantitative precipitation estimation," in *2023 United States National Committee of URSI National Radio Science Meeting (USNC-URSI NRSM)*, 2023, pp. 298–299.

[24] K. Shin, J. J. Song, W. Bang, and G. Lee, "Quantitative precipitation estimates using machine learning approaches with operational dual-polarization radar data," *Remote Sensing*, vol. 13, no. 4, 2021.

[25] F. F. Verdelho, C. Beneti, L. G. Pavam, L. Calvetti, L. E. S. Oliveira, and M. A. Zanata Alves, "Quantitative precipitation estimation using weather radar data and machine learning algorithms for the southern region of Brazil," *Remote Sensing*, vol. 16, no. 11, 2024.

[26] F. Sofos, A. Charakopoulos, K. Papastamatiou, and T. E. Karakasidis, "A combined clustering/symbolic regression framework for fluid property prediction," *Physics of Fluids*, vol. 34, no. 6, 2022.

[27] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[28] ——, "Scikit-learn: User guide," 2024. [Online]. Available: https://scikit-learn.org/1.5/user_guide.html

[29] Vijaya, S. Sharma, and N. Batra, "Comparative study of single linkage, complete linkage, and ward method of agglomerative clustering," in *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*, 2019, pp. 568–573.

[30] B. de Ville, "Decision trees," *WIREs Computational Statistics*, vol. 5, no. 6, p. 448–455, Oct 2013.

[31] L. Von Rueden, S. Mayer, K. Beckh, B. Georgiev, S. Giesselbach, R. Heese, B. Kirsch, J. Pfrommer, A. Pick, R. Ramamurthy *et al.*, "Informed machine learning – a taxonomy and survey of integrating prior knowledge into learning systems," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 1, pp. 614–633, 2021.

[32] N. Muralidhar, M. R. Islam, M. Marwah, A. Karpatne, and N. Ramakrishnan, "Incorporating prior domain knowledge into deep neural networks," in *2018 IEEE International Conference on Big Data*. IEEE, 2018, pp. 36–45.

[33] J. S. Marshall and W. M. K. Palmer, "The distribution of raindrops with size," *Journal of Atmospheric Sciences*, vol. 5, no. 4, pp. 165–166, 1948.

[34] R. A. Fulton, J. P. Breidenbach, D.-J. Seo, D. A. Miller, and T. O'Bannon, "The WSR-88D rainfall algorithm," *Weather and Forecasting*, vol. 13, no. 2, pp. 377–395, 1998.

[35] L. Lovmar, A. Ahlford, M. Jonsson, and A.-C. Syvänen, "Silhouette scores for assessment of SNP genotype clusters," *BMC Genomics*, vol. 6, pp. 1–6, 2005.

[36] S. E. Giangrande and A. V. Ryzhkov, "Estimation of rainfall based on the results of polarimetric echo classification," *Journal of Applied Meteorology and Climatology*, vol. 47, no. 9, pp. 2445–2462, 2008.

[37] E. A. Brandes, G. Zhang, and J. Vivekanandan, "Experiments in rainfall estimation with a polarimetric radar in a subtropical environment," *Journal of Applied Meteorology and Climatology*, vol. 41, no. 6, pp. 674–685, 2002.

[38] H. S. Park, A. V. Ryzhkov, D. Zrnić, and K.-E. Kim, "The hydrometeor classification algorithm for the polarimetric WSR-88D: Description and application to an MCS," *Weather and Forecasting*, vol. 24, no. 3, pp. 730–748, 2009.

[39] A. V. Ryzhkov, T. J. Schuur, D. W. Burgess, P. L. Heinselman, S. E. Giangrande, and D. S. Zrnic, "The joint polarization experiment: Polarimetric rainfall measurements and hydrometeor classification," *Bulletin of the American Meteorological Society*, vol. 86, no. 6, pp. 809–824, 2005.