MINISTRY OF ECONOMY
DEPARTMENT OF STATISTICS MALAYSIA

RAINFALL PREDICTION USING HYBRID BAYESIAN METHOD AND STATISTICAL DOWNSCALING BASED MACHINE LEARNING IN SELANGOR

NOOR HAMIZAH BINTI MOHAMAD SANI
UNIVERSITI PENDIDIKAN SULTAN IDRIS

11TH MALAYSIA STATISTICS CONFERENCE
"Data and Artificial Intelligence: Empowering the Future"
19TH September 2024
Organized by:
MINISTRY OF ECONOMY DEPARTMENT OF STATISTICS MALAYSIA
BANK NEGARA MALAYSIA
MALAYSIA INSTITUTE OF STATISTICS
MyStats 202

1. INTRODUCTION

- **Rainfall's Importance**: Rainfall is vital for human life and civilization, influencing various sectors like agriculture and infrastructure.
- **Need for Accurate Prediction**: Reliable rainfall prediction models are essential to improve disaster preparedness by providing early warnings for extreme weather events.
- **Data Challenges & Solutions**: Handling large, high-dimensional atmospheric data requires dimensionality reduction to simplify model training without losing predictive accuracy.
- **Advanced Modeling Techniques**: Combining Bayesian method with machine learning improves prediction accuracy by addressing non-linearity and excess zeros in the data.
- Thus, a prediction modeling a hybrid of Machine Learning-based Statistical Downscaling model and Bayesian method is proposed, potentially advancing predictive accuracy and reliability for better decision-making in agriculture, water management, and disaster preparedness

2. METHODOLOGY

•**Data Overview:** Data from 33 rainfall stations in Selangor (2008-2018).
Daily rainfall (predictand) and atmospheric variables (predictors).

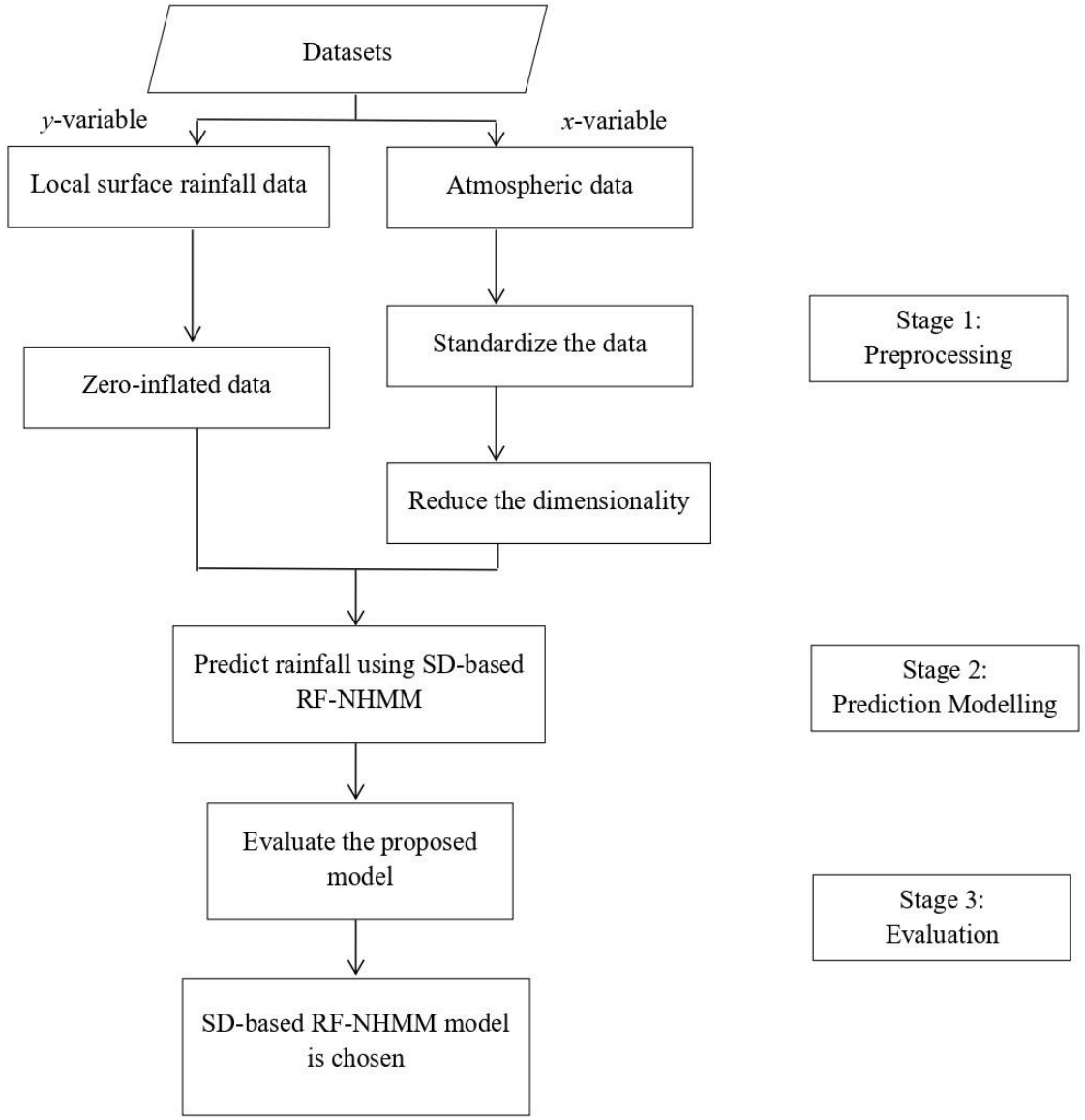
• **Methods:**

- PCA (Principal Component Analysis): Standardize the data and reduce dimensionality of the data.

- Non-Homogeneous Markov Model (NHMM): Captures zero-inflated data patterns.

- Random Forest (RF): Handles non-linearity in data and improves prediction accuracy.

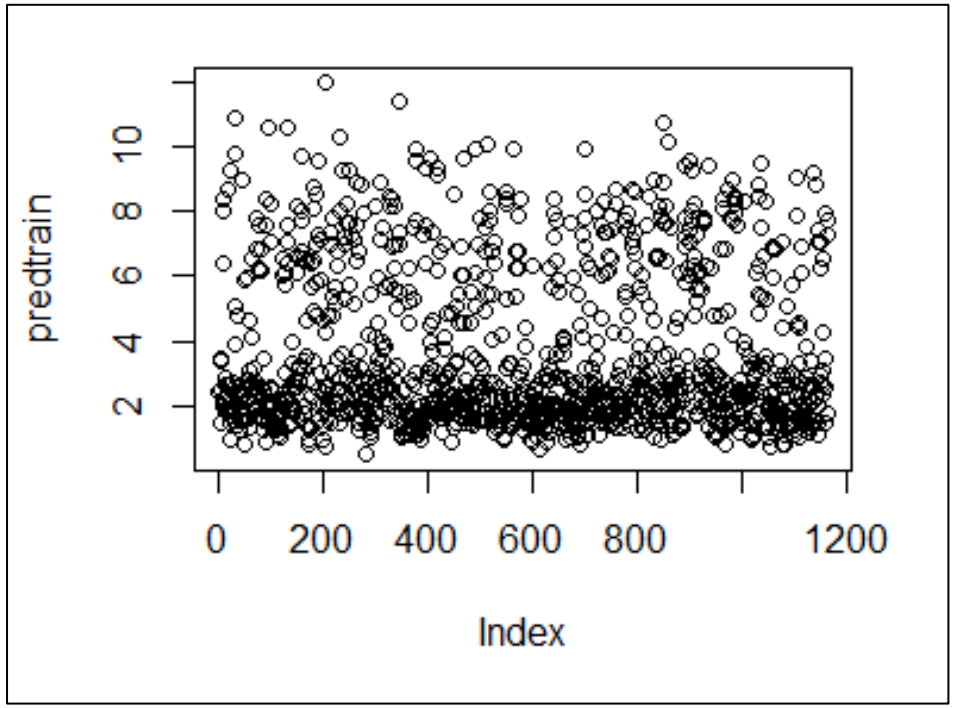
FRAMEWORK OF THE STUDY



3. RESULTS

- The Random Forest (RF) method, utilizing 500 decision trees, was applied to forecast rainfall trends in Selangor.
- The dataset was divided into two sections, with 70% of the data used for the calibration phase and the remaining 30% for validation purposes.
- The RF-NHMM model's performance was evaluated using metrics such as RMSE, NSE, MAE, and MFE for both the calibration and validation periods.

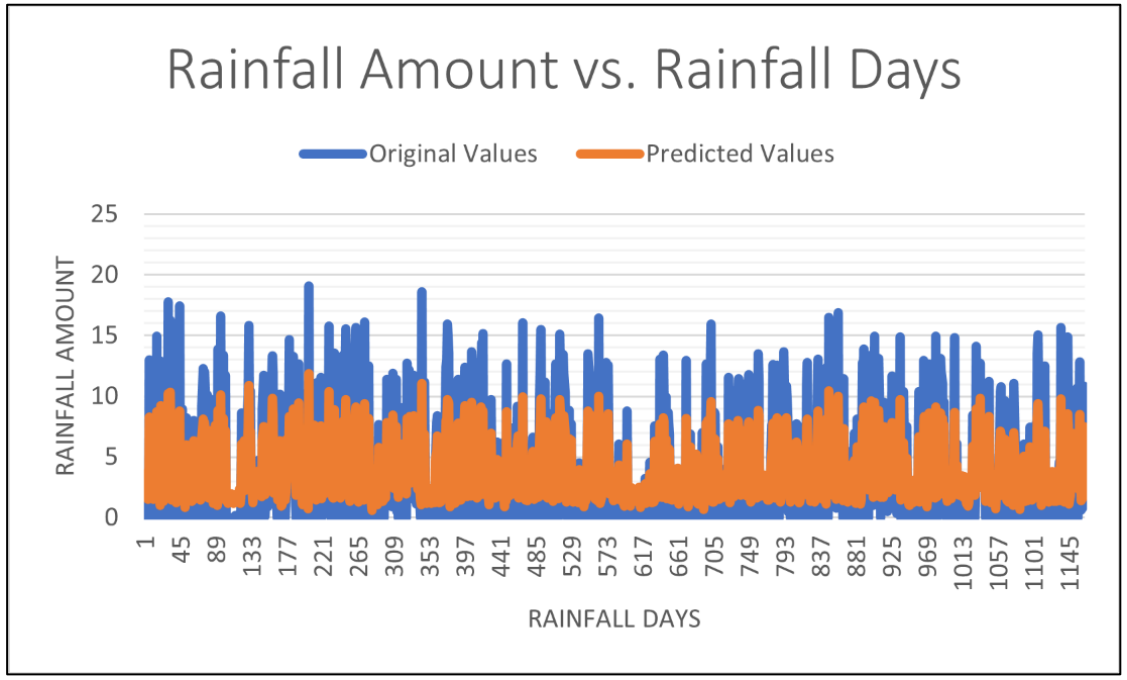
3. RESULTS



The scatter plot shows predicted rainfall values against their index, **revealing a tendency for predictions** to cluster around consistent values with some variability.

3. RESULTS

Calibration				Validation			
RMSE	NSE	MAE	MFE	RMSE	NSE	MAE	MFE
2.298	0.752	1.900	-0.049	4.937	0.006	3.851	0.867



- RMSE of 2.298 and MAE of 1.900 indicate moderate prediction accuracy
- NSE of 0.752 demonstrates decent predictive capability
- MFE of -0.049 suggests minimal bias in the predictions

- The model captures the overall trend but there are some noticeable discrepancies, especially during higher rainfall periods

4. DISCUSSION AND CONCLUSION

- Key Findings

- The RF-NHMM model provides superior accuracy in predicting rainfall, minimizing bias and error.

- Applications

- Can be used in climate change assessments, agricultural planning, water resource management, and disaster risk reduction.

- Future Research

- Further refinement of the model to improve prediction of extreme rainfall events.
- Expand the model to other regions with similar climate patterns.

Thank you