

國立臺灣大學工學院土木工程學系

碩士論文

Department of Civil Engineering

College of Engineering

National Taiwan University

Master's Thesis

探索基於機器學習之天氣類比法於預測極端降水事件
之應用

Development of A Machine Learning-Based Weather
Analog Method for Predicting Extreme Precipitation
Events

周語涵

Yu-Han Jhou

指導教授：張書瑋 博士

Advisor: Shu-Wei Chang Ph.D.

中華民國 114 年 2 月

February, 2025



國立臺灣大學碩士學位論文 口試委員會審定書

NATIONAL TAIWAN UNIVERSITY MASTER'S THESIS ACCEPTANCE CERTIFICATE

探索基於機器學習之天氣類比法於預測極端降水事件之應用

Development of A Machine Learning-Based Weather Analog Method for
Predicting Extreme Precipitation Events

本論文係 周語涵 (R10521229) 在國立臺灣大學土木工程學系結構工程組
完成之碩士學位論文，於民國2025年1月21日承下列考試委員審查通過及
口試及格，特此證明。

The undersigned, appointed by the Department of Civil Engineering, Structural Engineering
Division, on 21 January 2025 have examined a Master's Thesis entitled above presented by
YuHan Jhou (R10521229) candidate and hereby certify that it is worthy of acceptance.

口試委員 Oral examination committee:

張書瑋

(指導教授 Advisor)

汪立本

周佳靚

張書瑋

汪立本

周佳靚

系主任 Director:

葛宇甯

葛宇甯





Acknowledgements

感謝大家的一路相伴，原本這一句話就是我的致謝全文。可能是想到在這裡待得太久，遇到的人太多，酬庸的心態讓我不得不羅列那些人、忘卻某些人，即使真心要感謝的人可能不多。

首先要感謝的是指導教授張書瑋教授、汪立本教授，以及周佳靚教授，感謝他們在研究或是口試過程中的指導。這幾年我投身於研究的能量或是所花時間欠了太多太多，感謝他們能讓我在這樣的狀態下認識自身，探索、蹉跎這些時光，感激之情，無以言表。

再來感謝張老師研究室的夥伴們，原諒我只用三個人做代表。感謝諺靈，為什麼感謝，只能說是因為名字。感謝宏智，如果不是在這裡認識，我們應該可以一起玩更多遊戲、去到更多地方。你在足球隊擔任後衛施展大岩壁時，不要說是我教的。感謝子皓，你在我們尚未熟識時的熱情攀談，給予了我走進研究室結識大家的能量。或許一直以來我就只是想走進來跟你說說話。要是有人感受到了我的偏心，那我是真偏心。此外還有羽白、昶毅、佩澄、正順、泓佑、柏學、冠

綸、筠筑、才津、皇文、重荃，若真有遺漏的名字不用跟我說，也不要跟他說。



這裡還要感謝汪老師研究室與周老師研究室的大家，特別感謝沛均、筠曼研究上的協助，以及有成，有你真好。再來感謝系排的大家，除了一起打球也讓我做了隊長任性一回，泰嘉快點畢業再一起打球吃飯。還有感謝國青研酒室的酒桶們，你們應該算是名符其實的酒肉朋友。

原本我只想用一句話表達我的致謝全文，而我有之所以有能量在這裡感謝這麼多的人、事、物，只是為了在這裡寫出你的名字，謝謝智賢。

真心，感謝大家的一路相伴。



摘要

極端降水事件對防災減災構成重大挑戰，使得準確的預測更為重要。而天氣類比法利用歷史大氣模式尋找相似的過去事件，提供了一種具有物理一致性且經驗驗證的方法，在極端降水的預測中已被證明有效。然而，傳統的天氣類比方法在處理複雜的大氣狀態時，往往難以有效捕捉局部特徵與非線性關係。因此，本研究提出基於 AtmoDist 的機器學習方法，用於評估大氣相似性並提升降水事件的預測能力。

為了提高模型對局部特徵的敏感度，我們引入了空間加權處理與兩階段天氣類比技術，並將 AtmoDist 與傳統方法進行系統比較。此外，利用可解釋人工智慧 (XAI) 技術深入分析模型的決策過程，以揭示重要特徵與影響因素。

研究結果顯示，不同相似性計算方法對大氣變數的表現存在差異，反映變數特定評估的必要性。AtmoDist 模型在降水預測上優於傳統方法，尤其對局部降水事件的預測能力經空間加權後明顯提升。同時，結果也指出模型準確度未必能充分代表其預測能力，需進一步關注特徵重要性分析。透過 DeepSHAP 評估發現，

AtmoDist 對具有高時空變異性的大氣變數表現出更強的關注。



本研究為改善降水預測模型提供了新的視角，並說明可解釋技術對理解模型行為和提升可信度的重要性。此外，結果凸顯機器學習方法在捕捉非線性特徵與空間變異性方面的優勢，為未來替代傳統方法提供了潛在應用價值。

關鍵字：機器學習、天氣類比法、極端降水預測、大氣相似性



Abstract

Extreme precipitation events pose significant challenges for disaster prevention and mitigation, making accurate forecasting essential. Weather analog methods leverage historical atmospheric patterns to identify similar past events, offering a physically consistent and empirically validated approach that has proven effective in predicting extreme precipitation. However, traditional weather analog methods often struggle to effectively capture localized features and nonlinear relationships in complex atmospheric states. To address these challenges, this study proposes a machine learning-based approach using AtmoDist to assess atmospheric similarity and improve precipitation event prediction.

To enhance sensitivity to localized features, we incorporated spatial weighting techniques and a two-stage weather analog framework, systematically comparing AtmoDist with conventional methods. Furthermore, we applied explainable artificial intelligence (XAI) techniques to analyze the model's decision-making process, providing insights into key features and influencing factors.



Our findings reveal that different similarity metrics yield varying performances across atmospheric variables, highlighting the need for variable-specific assessments. The AtmoDist model outperforms traditional approaches in precipitation forecasting, particularly for localized precipitation events, with notable improvements after spatial weighting adjustments. However, results also suggest that accuracy alone may not fully capture predictive performance, emphasizing the importance of feature importance analysis. Using DeepSHAP, we observed that AtmoDist places greater emphasis on atmospheric variables with high spatiotemporal variability.

This study offers a new perspective on improving precipitation forecasting models and underscores the role of explainable techniques in enhancing model interpretability and reliability. Moreover, the results demonstrate the advantages of machine learning methods in capturing nonlinear patterns and spatial variability, highlighting their potential as viable alternatives to traditional approaches.

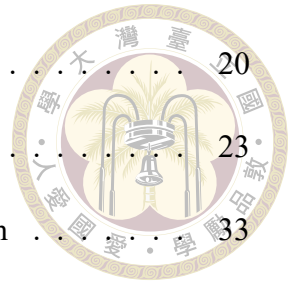
Keywords: Machine Learning, Weather Analog Method, Extreme Precipitation Prediction, Atmospheric Similarity



Table of Contents

	Page
Verification Letter from the Oral Examination Committee	i
Acknowledgements	iii
摘要	v
Abstract	vii
Table of Contents	ix
List of Figures	xiii
List of Tables	xvii
Chapter 1 Introduction	1
1.1 Background and Motivation	1
1.2 Literature Review	5
1.2.1 Machine Learning-based Weather Analog Ensemble Methods	5
1.2.2 Similarity Metrics for Atmospheric Applications	8
1.2.3 Downscaling for Enhanced Localized Climate Projections	12
1.3 Research Objectives	16
1.4 Dissertation Structure	17
Chapter 2 Data	19
2.1 ERA5 (ECMWF Reanalysis v5)	19

2.1.1	Feature Selection	20
2.1.2	Exploratory Data Analysis	23
2.2	Met Office Rain Radar Data from the NIMROD System	33
2.3	Extreme Precipitation Events	36
Chapter 3	Methods	39
3.1	Framework of Two-Phase Weather Analog Method	39
3.2	Traditional Methods for Real-Space Similarity Calculation	44
3.2.1	Weighting for Similarity Calculation	44
3.2.1.1	Formula-Driven Weight Calculation for Atmospheric Data	45
3.2.1.2	Analysis of Weight Calculations for Spatiotemporal and Variable Components	48
3.2.2	Comprehensive Similarity Calculation Formula for Weather Event Analogs	52
3.2.2.1	Different Types of Similarity Metrics	53
3.2.2.2	Similarity Calculation Formula for Weather Events	58
3.3	Machine Learning Approaches for Similarity Estimation	60
3.3.1	Reconstructed Model Structure	63
3.3.2	Data Processing and Feature Analysis	65
3.3.3	Model Performance	70
3.4	Framework and Metrics for Assessing Method Performance	74
3.4.1	Phase I: Evaluation Methods for the Capability of Different Similarity Metrics in Precipitation Event Search	75
3.4.1.1	Methodology for Evaluating Similarity Metrics Across Variables in a Single Event	76



3.4.1.2	Evaluation Methodology for Similarity Search Results Across 157 Precipitation Events	79
3.4.1.3	Computational Approach for Binned Spread Skill Assessment Using Different Similarity Metrics	80
3.4.1.4	Brier Score and Brier Skill Score Computation	83
3.4.2	Phase II: Evaluation Methods for Localized Precipitation Forecast Results	86
3.4.3	Methods for Explaining the Intrinsic Performance of Models Using DeepSHAP	88
3.4.3.1	Theoretical Foundation of DeepSHAP	88
3.4.3.2	Processing Flow of DeepSHAP	90
Chapter 4	Results and Discussion	95
4.1	Phase I: Comparison of Weather Event Similarity Calculation and Prediction	95
4.1.1	Sum of Squared Differences Analysis of Calculation Results for a Single Event	96
4.1.2	Sum of Squared Differences Analysis of Prediction Results for 157 Precipitation Events	100
4.1.3	Spread Skill Evaluation of Models and Methods	103
4.1.4	Brier Skill Score Evaluation	107
4.2	Phase II: Localized Precipitation Forecasting	111
4.3	Explaining Prediction Results Using DeepSHAP	116
Chapter 5	Conclusion and Outlook	123
5.1	Conclusion	123
5.2	Future Work and Outlook	125

References

Appendix A — 157 Extreme Precipitation Events

A.1	Complete List of Precipitation Events	129
-----	---	-----

Appendix B — Model Optimization Process and Accuracy Comparison 153

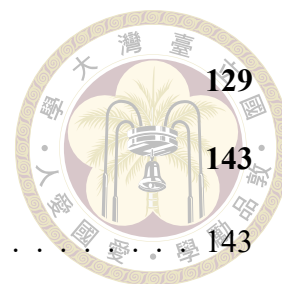
B.1	Model Architectures	153
-----	-------------------------------	-----

B.2	Different Data Preprocessing Methods	155
-----	--	-----

B.3	Effect of Temporal Resolution	156
-----	---	-----

Appendix C — Evaluation of Different Weather Analog Methods 160

C.1	Single-Event SSR Evaluation Across Variables for Multiple Methods	160
-----	---	-----

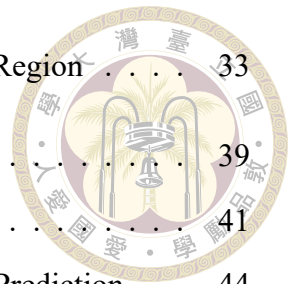




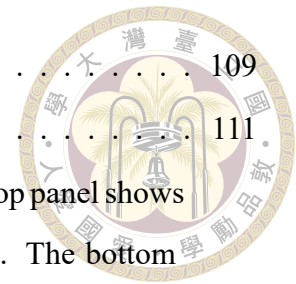
List of Figures

1.1	Weather Analog Method	2
1.2	Deep Analog Network for Feature Space Embedding	7
1.3	AtmoDist Methodology Overview	11
1.4	NORA System's Two-Stage Analog Selection Strategy	14
2.1	Geographic Area of the Study	20
2.2	Distribution Characteristics of the Five Variables	24
2.3	Spatial distribution of the statistical properties (mean, standard deviation, maximum, minimum) of $d2m$ averaged over entire period of 2001-2020 .	25
2.4	Spatial distribution of the statistical properties (mean, standard deviation, maximum, minimum) of u component of wind averaged over entire period of 2001-2020	25
2.5	Spatial distribution of the statistical properties (mean, standard deviation, maximum, minimum) of v component of wind averaged over entire period of 2001-2020	26
2.6	Spatial distribution of the statistical properties (mean, standard deviation, maximum, minimum) of msl averaged over entire period of 2001-2020 .	26
2.7	Spatial distribution of the statistical properties (mean, standard deviation, maximum, minimum) of r averaged over entire period of 2001-2020 . . .	27
2.8	Seasonal Decomposition Plot of $d2m$	28
2.9	Seasonal Decomposition Plot of u	29
2.10	Seasonal Decomposition Plot of u	29
2.11	Seasonal Decomposition Plot of msl	30
2.12	Seasonal Decomposition Plot of r	30

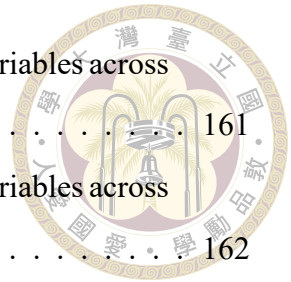
2.13	Nimrod Radar Precipitation Coverage for the Birmingham Region	33
3.1	Framework of Two-Phase Weather Analog Method	39
3.2	Workflow of Weather Analog Method	41
3.3	Weighted Multi-Factor Framework for Atmospheric Event Prediction	44
3.4	Weights Derived from Precipitation Event Data	48
3.5	Wind Speed and Direction Distribution	50
3.6	Atmodist Model Structure for Atmospheric Event Similarity Assessment	60
3.7	Revised Model Structure for Enhanced Small-Scale Patch Analysis	63
3.8	Distribution Characteristics of the Five Variables After Spatial Weighting	66
3.9	Variance Analysis of Temporal Lags using Dew Point Temperature ($d2m$)	67
3.10	L2 Norm Difference of Atmospheric Variables Over Time Lag	69
3.11	Validation Performance of Atmodist Model Using Weighted and Unweighted Data	70
3.12	Confusion Matrices of Unweighted and Weighted Models for Time Difference Prediction	72
3.13	Evaluation Framework of Similarity Metrics and Precipitation Forecasts	74
3.14	Selected Local Data Points Based on Correlation Coefficients	78
3.15	Processing Flowchart of DeepSHAP	90
4.1	Sum of Squared Residuals Analysis for Five Atmospheric Variables across Different Analogues Based on AtmoDist Results	96
4.2	Sum of Squared Residuals for Five Variables across Five Methods	98
4.3	Comparison of First-Place Counts for Different Methods Across Atmospheric Variables	99
4.4	Comparison of Sum of Squared Residuals Distribution across Five Methods for 157 Precipitation Events	101
4.5	Global and Local SSR Ranking Distribution Across Methods	102
4.6	Binned Spread Skill Comparison for Five Variables Using Different Methods	104
4.7	Binned Spread-Skill Relationship for Precipitation	106
4.8	Brier Skill Score for Different Variables	107



4.9	Brier Skill Score for Precipitation Prediction	109
4.10	Observed Precipitation Patterns	111
4.11	Localized precipitation forecasts using AtmoDist-SW. The top panel shows the mean precipitation rate derived from 12 analog events. The bottom panels present the deterministic mean forecast (left) and the probabilistic precipitation forecast (right), both representing the expected precipitation over the next 2–3 hours.	112
4.12	Localized Precipitation Forecasts Using Cosine Similarity. The top panel shows the mean precipitation rate derived from 12 analog events. The bottom panels present the deterministic mean forecast (left) and the probabilistic precipitation forecast (right), both representing the expected precipitation over the next 2–3 hours.	113
4.13	Comparison of Proportional SHAP Values Between AtmoDist-SW and AtmoDist	116
4.14	Comparison of Standard Deviation Distributions Under Two Data Processing Methods	117
4.15	Spatial Distribution of SHAP Values for AtmoDist-SW and AtmoDist Across Variables	119
4.16	Mean SHAP Value Over Time Steps for AtmoDist-SW and AtmoDist Across 157 Events	120
B.1	Validation Loss and Accuracy Comparisons for Different Structural Modifications	154
B.2	Validation Loss and Accuracy Comparisons for MaxPooling and ResBlock Modifications	154
B.3	Validation Loss and Accuracy Comparisons for Different Data Preprocessing Methods	155
B.4	Validation Loss and Accuracy Comparisons for Different Temporal Resolutions and Time Intervals	156
C.5	Sum of Squared Residuals Analysis for Five Atmospheric Variables across Different Analogues Based on AtmoDist Results	160



C.6	Sum of Squared Residuals Analysis for Five Atmospheric Variables across Different Analogues Based on Euclidean Results	161
C.7	Sum of Squared Residuals Analysis for Five Atmospheric Variables across Different Analogues Based on Cosine Results	162
C.8	Sum of Squared Residuals Analysis for Five Atmospheric Variables across Different Analogues Based on SSIM Results	163





List of Tables

2.1	Overview of the ERA5 Data	23
2.2	Settings for Seasonal Decomposition	32
2.3	Overview of the Nimrod Radar Precipitation Data	35
2.4	Extreme Precipitation Events (First Five)	37
3.1	Summary of Variables, Resolutions, and Forecast Settings	43
3.2	Layer-wise Comparison of Original Atmodist Model and Modified At- moDist Model with Corresponding Shapes	65
4.1	Comparison of BS and BSS values for different variables between AtmoDist- SW and Cosine methods.	108
4.2	Comparison of BS and BSS values for different precipitation thresholds between AtmoDist-SW and Cosine methods.	110





Chapter 1 Introduction

1.1 Background and Motivation

The frequency and intensity of extreme rainfall events are increasing globally due to climate change [1], resulting in more severe flooding and widespread damage. To address these challenges, reliable prediction methods are essential for mitigating impacts and reducing losses.

The Weather Analog Method is a forecasting technique based on historical data, which predicts future weather conditions by comparing current weather patterns with similar historical events. As shown in Figure 1.1, this method involves searching for historical analogs that closely resemble the current atmospheric state. Compared to other forecasting methods, the Weather Analog Method offers the advantages of being data-driven, simple, intuitive, and requiring fewer computational resources [2]. By leveraging a rich historical weather database, this approach effectively utilizes past weather patterns to predict future states, making it a valuable tool for forecasting.

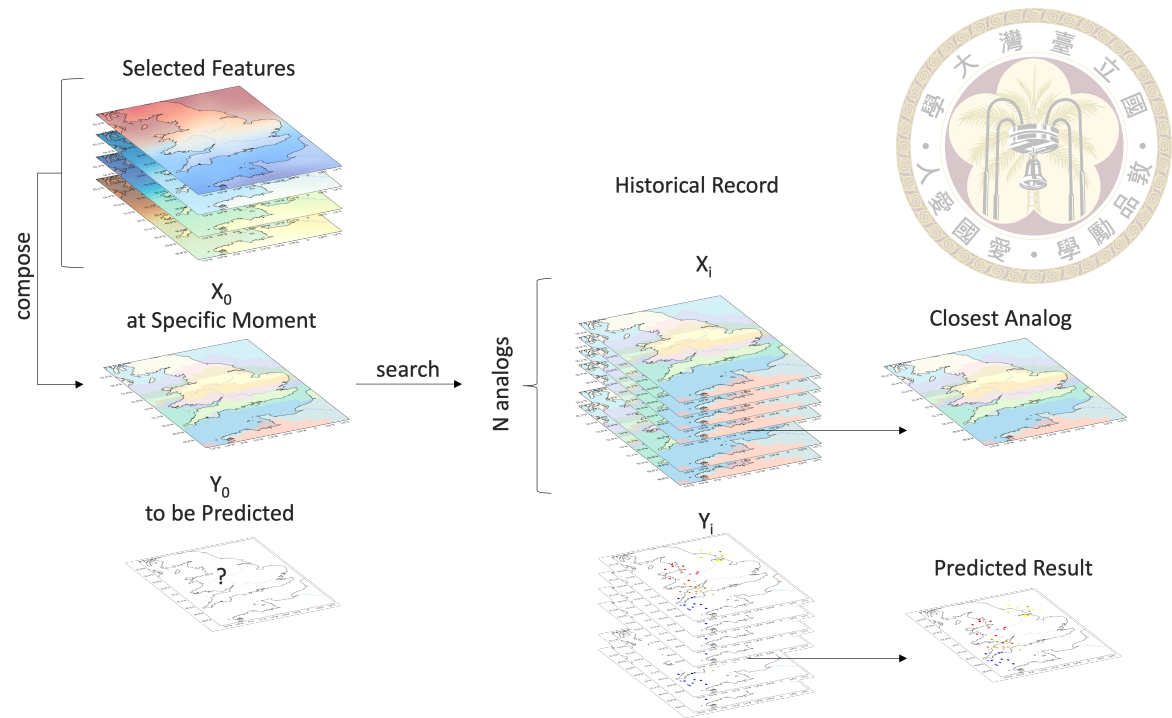


Figure 1.1 Weather Analog Method

Atmospheric features at a specific time (X_0) are extracted and compared against a historical record (X_i) to find N analogs that are closest in resemblance. The closest analog's result (Y_i) is then used to predict the outcome (Y_0) for the moment of interest. This approach relies on matching past weather patterns to the present state to make informed forecasts.

For instance, Horton [3] highlighted that the simplicity and efficiency of the analog method make it particularly useful in hydrometeorological applications. By identifying relevant analog cases, it enables fast and effective predictions at low computational costs, with operations that are both straightforward and easy to understand. Furthermore, according to Chattopadhyay et al. [4], while Numerical Weather Prediction (NWP) models demand increasing computational time and resources, they still face challenges in forecasting extreme weather events. In contrast, ensemble forecasting using the analog method has demonstrated potential advantages in predicting heavy precipitation events compared to more complex NWP models.



With advancements in machine learning, its application to analog search within the Weather Analog Method shows significant potential, as it excels in identifying patterns and relationships within complex and variable weather data, enabling more accurate and efficient predictions [5]. However, the Machine Learning-Based Weather Analog Method faces several challenges in practical applications. First, it relies on analog searches, which require datasets with explicitly defined similarity metrics for atmospheric states during model training [6]. Defining similar atmospheric states demands both a high level of expertise and careful consideration of the methods used, as these can significantly influence the determination of similarity. Additionally, numerical similarity does not always correspond to physical similarity, a limitation that is particularly relevant in weather forecasting [7].

Extreme precipitation events in urban areas are highly localized, whereas atmospheric parameters are typically observed and modeled at much larger scales. This scale mismatch poses challenges, as large-scale atmospheric conditions often fail to capture the fine-scale variability needed for accurate predictions of localized precipitation. Addressing this issue requires integrating higher-resolution observational data and refining similarity labeling methods when applying machine learning to weather analog forecasting. These enhancements are crucial for improving the accuracy and reliability of predictions at the urban

scale.

These challenges highlight the need for a deeper understanding of the strengths and limitations of the machine learning-based Weather Analog Method, particularly when integrated with machine learning for extreme precipitation prediction.



1.2 Literature Review



This literature review delves into the machine learning-based Weather Analog Method, examining its potential and addressing critical challenges. Defining similarity metrics for data labeling in model training remains a key issue, as the accuracy of these metrics significantly influences prediction reliability. Furthermore, the inherent scale mismatch between large-scale atmospheric observations and city-scale precipitation events poses additional hurdles. By exploring existing studies, this review aims to provide insights into these challenges and outline potential strategies for improving the method's applicability and effectiveness.

1.2.1 Machine Learning-based Weather Analog Ensemble Methods

The Weather Analog Method is an essential tool in weather forecasting. Delle Monache et al. [2] were the first to propose using a 21-member analog ensemble to forecast 10-meter wind speed and 2-meter temperature within 48 hours in the United States. Compared to other deterministic and probabilistic forecasting methods such as the Regional Ensemble Prediction System (REPS), Ensemble Model Output Statistics (EMOS), and Logistic Regression (LR), the analog method has demonstrated superior performance in capturing extreme events.

Bochenek et al. [8], through an examination of 500 scientific articles published since 2018 on the application of machine learning in climate and numerical weather prediction, demonstrated the advantages of machine learning in handling high-dimensional data, accounting for variations across different spatial and temporal scales, and addressing chaotic dynamic systems. Machine learning methods in weather analog forecasting can generate ensemble members by capturing complex atmospheric patterns, thus outperforming traditional methods in dealing with the nonlinear relationships within atmospheric data.

However, compared to traditional methods, machine learning demonstrates less flexibility and transparency when generating ensemble members [9]. In contrast, traditional analog methods predict by matching historical similar events, offering a straightforward and easy-to-understand approach. Additionally, machine learning models often require large amounts of high-quality training data, yet data for extreme weather events are typically scarce, which can limit model performance. Furthermore, when faced with unseen data, machine learning models are more prone to poor generalization due to their lack of exposure to rare events [10, 11]. Since extreme weather events usually occupy the tail end of variable distributions, predicting them is naturally more difficult [12]. Moreover, ensemble members generated by machine learning models may produce physically unrealistic predictions due to the difficulty of incorporating physical constraints [13]. These

considerations highlight the disadvantages of applying machine learning methods to extreme weather event prediction and underscore the advantages of traditional analog methods, which rely on real historical data for forecasting.

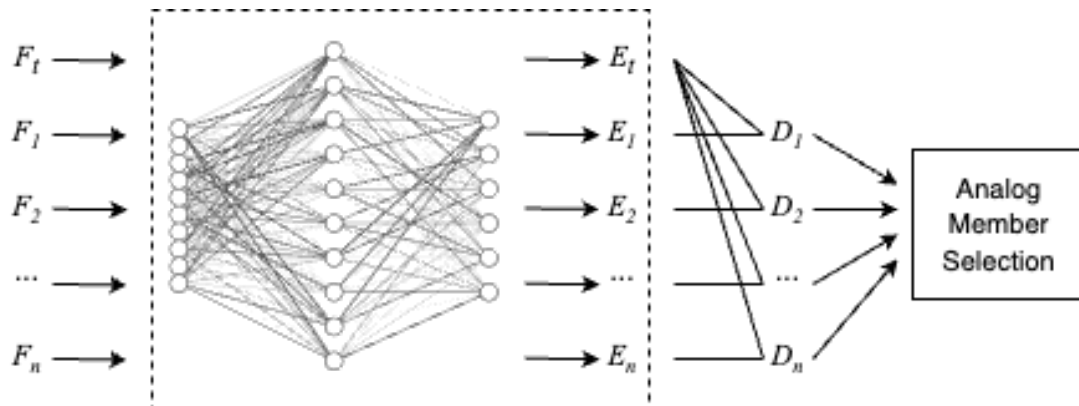


Figure 1.2 Deep Analog Network for Feature Space Embedding

A schematic showing how multivariate forecast time series (F) are mapped into a feature space (E) to calculate distances (D) for similarity estimation. This process enables the comparison of atmospheric patterns by transforming raw data into a learned embedding space. (Adapted from [14])

Hu et al. [14] introduced a method that leverages machine learning techniques to improve near-surface weather analog forecasting. As illustrated in Figure 1.2, this approach uses machine learning to map atmospheric states into neural network space, optimizing the feature space through a process of embedding. A triplet approach is employed to improve analog identification, thereby overcoming the computational limitations of traditional analog ensemble (AnEn) methods. Their study demonstrated that machine learning is more effective in processing and analyzing multivariate data, enhancing forecasting accuracy and efficiency. This method not only retains the traditional advantage of matching

historical events to generate ensemble members but also addresses the limitations of traditional computational methods in handling nonlinear relationships, making it particularly suitable for dealing with complex weather patterns and improving the quantification of forecast uncertainty.



Building on the above, our research aims to apply machine learning techniques to similarity assessment, specifically by mapping atmospheric states into neural network space to optimize the feature space. We will focus on defining the analogs in the training dataset and gaining a deeper understanding of traditional similarity measures. With these improvements, we expect to significantly enhance the accuracy and efficiency of weather forecasting and achieve breakthroughs in the prediction and management of extreme weather events.

1.2.2 Similarity Metrics for Atmospheric Applications

The performance of weather analog methods largely depends on the techniques used for selecting analog members, with similarity algorithms playing a pivotal role. The calculation of atmospheric similarity has undergone extensive development and found various applications.

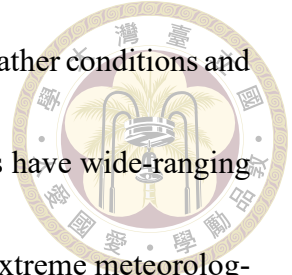
Elmore and Richman [15] conducted an in-depth analysis of the application of princi-

pal component analysis (PCA) in meteorological studies. They emphasized that selecting an appropriate similarity matrix is crucial for enhancing physical insight in analysis. The study highlighted that there is no fixed formula for determining the best similarity coefficient, as it depends on the analyst's specific needs for physical interpretation. Their findings indicated that a Euclidean distance-based similarity matrix provides insights consistent with the physical factors controlling spatial gradients in pressure, rainfall, and temperature fields within meteorological datasets.

Moreover, Zhou and Zhai [16] introduced a novel analog forecasting model, KISAM, which integrates large-scale atmospheric variables, a weighted cosine similarity method, and a cuckoo search optimization algorithm to predict extreme precipitation events. When compared with the deterministic model output from the European Centre for Medium-Range Weather Forecasts (ECMWF), KISAM exhibited higher skill scores and lower root mean square errors (RMSE) for forecasts exceeding three days, particularly in predicting heavy rainfall events.

Additionally, Hoffmann [17] advanced the use of the Structural Similarity Index Method (SSIM) to analyze and quantify the persistence of large-scale atmospheric flow patterns. This approach enables researchers to identify structural differences and similarities between two images, assigning a score between 0 and 1. The research demonstrated

that SSIM helps improve understanding and forecasting of extreme weather conditions and their future changes. These studies indicate that similarity algorithms have wide-ranging applications and significant potential for addressing and forecasting extreme meteorological conditions.



However, traditional methods such as Euclidean distance encounter significant computational challenges when applied to large-scale, high-dimensional data. These methods require pairwise distance calculations between data points, and as the volume and dimensionality of the data increase, the computational complexity grows exponentially [14]. Furthermore, many traditional similarity measures assume linear relationships between variables, whereas in real-world atmospheric systems, there are often complex, nonlinear interactions between different variables [18]. Traditional methods struggle to capture and interpret these nonlinear relationships, leading to inefficiencies in predictions and reduced accuracy in representing the true physical phenomena.

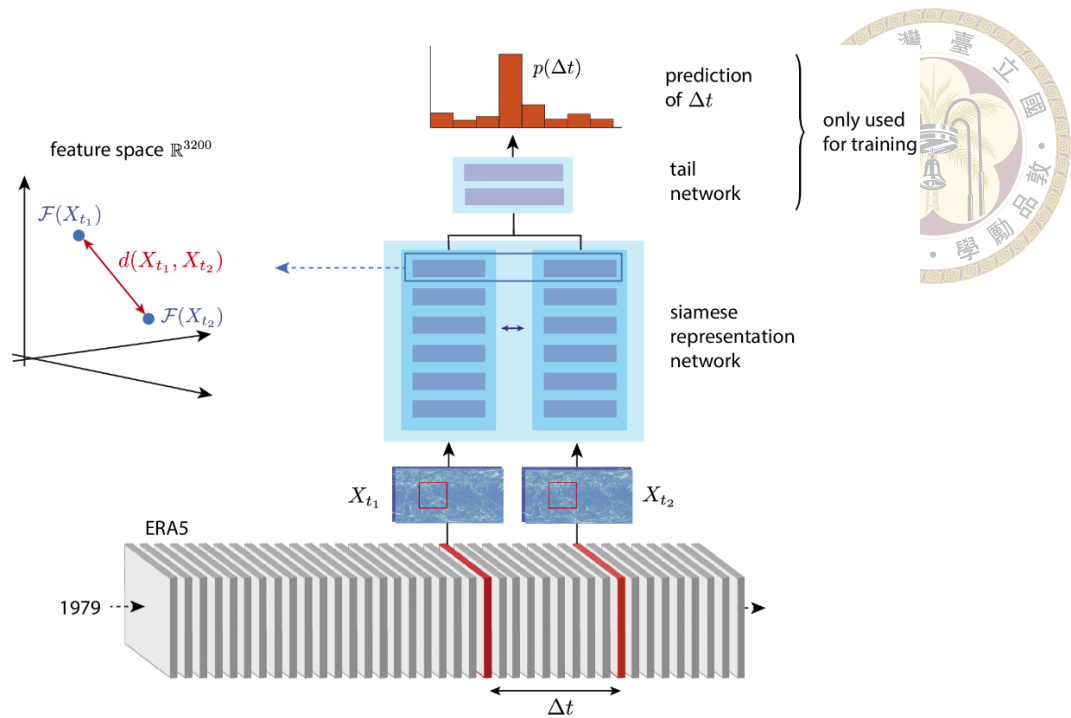


Figure 1.3 AtmoDist Methodology Overview

Two temporally close atmospheric fields are embedded into a high-dimensional space through a shared network. The embeddings are used to predict their temporal separation, Δt , aiding in identifying the feature space distance. The model is trained end-to-end for accurate temporal mapping. Source: [19]

Compared to traditional methods, machine learning offers the ability to dynamically adjust model parameters, making it more adaptable to changing weather patterns. Deep learning methods, in particular, can automatically learn complex nonlinear relationships within data, significantly enhancing prediction accuracy. For instance, as shown in Figure 1.3, the AtmoDist model proposed by Hoffmann and Lessig [19] employs self-supervised learning to capture intrinsic representations of atmospheric dynamics. By leveraging unlabeled data, it defines a data-driven distance metric between atmospheric states. The figure illustrates how the model maps atmospheric state data from different time points into a feature space and subsequently defines distances between states based on their representa-

tions within this space. This approach improves forecasting performance in applications such as high-resolution downscaling and missing data imputation. It demonstrates that machine learning methods not only improve prediction accuracy and efficiency but also provide new perspectives for understanding atmospheric predictability, offering powerful tools for future weather forecasting and climate analysis.



Building on this, our research aims to apply the AtmoDist model to analog member selection in weather analog methods and compare its performance with traditional techniques. We anticipate that this will enhance forecast accuracy and provide further scientific support for related studies.

1.2.3 Downscaling for Enhanced Localized Climate Projections

In meteorological forecasting, atmospheric parameter observations typically cover large areas, making it difficult to accurately reflect localized rainfall conditions, such as those involved in urban flooding. Atmospheric parameters, such as temperature, humidity, and wind speed, provide valuable macro-level meteorological information but often lack precision in smaller, complex regions. Using these parameters directly for small-scale rainfall forecasting can increase prediction errors and reduce accuracy. Therefore, more refined methods are needed to enhance prediction precision.



To address this issue, Matulla et al. [20] proposed a downscaling method for daily precipitation forecasting using the Weather Analog Method. This approach improves precipitation forecasts by introducing and comparing various weighted similarity metrics. The Weather Analog Method predicts future weather conditions by using similar cases from historical data, and the key lies in how these analog cases are selected and weighted. The study demonstrated that selecting appropriate similarity metrics significantly improves forecast accuracy. Specifically, this method defines weights by calculating the correlation between the target location of the event and the surrounding areas, thereby increasing prediction precision. The weighted similarity metric not only considers the correlation between different meteorological variables but also dynamically adjusts weights to adapt to varying weather conditions, leading to a more accurate reflection of localized rainfall conditions.

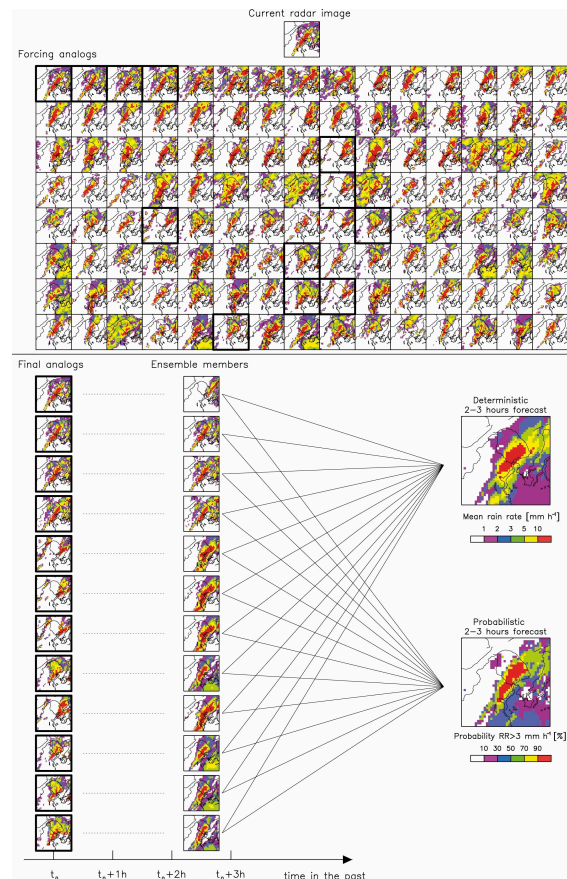


Figure 1.4 NORA System's Two-Stage Analog Selection Strategy

In the first stage, 120 analogs are identified based on current atmospheric parameters by comparing them to historical records. In the second stage, these 120 analogs are refined using their high-resolution rainfall data to generate more accurate precipitation forecasts. This process results in deterministic and probabilistic forecasts, providing both mean rain rate and probability of exceeding specific rain rates for a short-term forecast window. Source: [21]

Furthermore, Panziera et al. [21] developed the NORA (Nowcasting of Orographic Rainfall by means of Analogues) system, which further refines the Weather Analog Method, focusing on improving short-term rainfall forecasts. The NORA system is optimized for mountainous regions, where complex terrain significantly influences rainfall patterns. It employs a two-stage analog selection strategy, as shown in Figure 1.4: first, a preliminary filtering of analogs based on forecast variables; second, a refined selection using Eu-

clidean distance similarity metrics. This dual process allows the NORA system to more accurately capture the effects of terrain on rainfall patterns, providing more detailed and reliable precipitation forecasts.



In the first stage of analog selection, the NORA system uses multiple forecast variables from historical meteorological data, such as rainfall rate, temperature, humidity, and wind speed, to select cases that are most similar to the current meteorological conditions. This process ensures that the initially filtered analogs closely match the current situation across multiple parameters. The second stage further refines this selection process by using Euclidean distance similarity metrics to identify the most suitable analogs for the predicted rainfall. This refined process enables the forecast to better reflect real conditions, particularly in mountainous areas where terrain significantly impacts rainfall patterns.

In summary, by combining weighted similarity metrics and a two-stage analog selection strategy, these studies highlight the significant potential of the Weather Analog Method for improving small-scale rainfall forecasts. These methods not only enhance forecast accuracy but also address the meteorological challenges posed by complex terrain in localized areas, providing strong technical support for future meteorological predictions.

1.3 Research Objectives

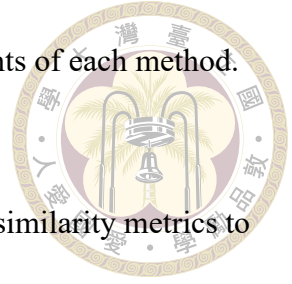


Building upon the insights gleaned from the preceding literature review, this study aims to address the limitations of current machine learning applications to weather analog methods, particularly in terms of similarity calculations. Our primary objective is to develop a machine learning-based weather analog forecasting method specifically designed for predicting extreme precipitation events. Through expanding the scope of similarity computations, this research also deepens our understanding of the complex dynamics driving extreme climate phenomena. The key objectives and contributions are as follows:

1. Developing a machine learning-based weather analog method specifically tailored for predicting extreme precipitation events.
2. Enhancing weather similarity computation and incorporating a self-supervised model (AtmoDist) to learn feature-space distances, enabling effective classification of atmospheric states while offering deeper insights into the intricate mechanisms behind extreme climate events.
3. Employing explainable artificial intelligence (XAI) to improve model transparency and trustworthiness.
4. Comparing traditional real-space and machine learning feature-space approaches

for similarity calculation to assess the advantages and constraints of each method.

5. Proposing a two-phase analog selection strategy and weighted similarity metrics to enhance forecast accuracy for localized precipitation.



1.4 Dissertation Structure

This dissertation is organized into five chapters, each addressing different aspects of the study.

Chapter 1 introduces the background, motivation, and objectives of the study. It also reviews relevant literature, including machine learning-based weather analog ensemble methods, similarity metrics for atmospheric applications, and downscaling techniques for improving localized climate projections. The chapter concludes with a statement of the research purpose and significance.

Chapter 2 describes the datasets used in this study. It covers feature selection and exploratory data analysis of the ERA5 reanalysis dataset, along with radar precipitation data from the Met Office NIMROD system. Additionally, it introduces extreme precipitation events employed for evaluation purposes.

Chapter 3 details the methodologies adopted in this study. It begins with the frame-

work of the two-phase weather analog method, followed by discussions on traditional similarity calculation methods, weighting strategies, and machine learning approaches for similarity estimation. This chapter also explains the evaluation frameworks and metrics used to assess method performance, including precipitation event searches and localized forecast evaluations. Furthermore, DeepSHAP is introduced to interpret model predictions.

Chapter 4 presents the results and discussions. It evaluates the performance of various similarity metrics and methods in precipitation event predictions. The results are analyzed for both single and multiple precipitation events, and the localized forecast performance is assessed. DeepSHAP is used to explain the internal logic behind model predictions.

Chapter 5 summarizes the findings and contributions of this study. It provides a conclusion and proposes recommendations for future research directions.





Chapter 2 Data

This study aims to explore the use of weather analog methods and a self-supervised atmospheric distance computation model to predict extreme precipitation events in the UK. To achieve this goal, a comprehensive analysis is conducted using multiple datasets, including extreme precipitation event data over the pilot area in the UK, ERA5 global atmospheric reanalysis data, and Nimrod radar precipitation data. The selection and purpose of these datasets will be discussed further in the following sections.

2.1 ERA5 (ECMWF Reanalysis v5)

This study utilizes the ERA5 global reanalysis dataset provided by the European Centre for Medium-Range Weather Forecasts (ECMWF) as the primary source of meteorological data. The ERA5 dataset, produced by ECMWF through the integration of a data assimilation system and historical observations, provides hourly temporal resolution data from 1940 to the present, with a spatial resolution of 31 kilometers (approximately 0.25°

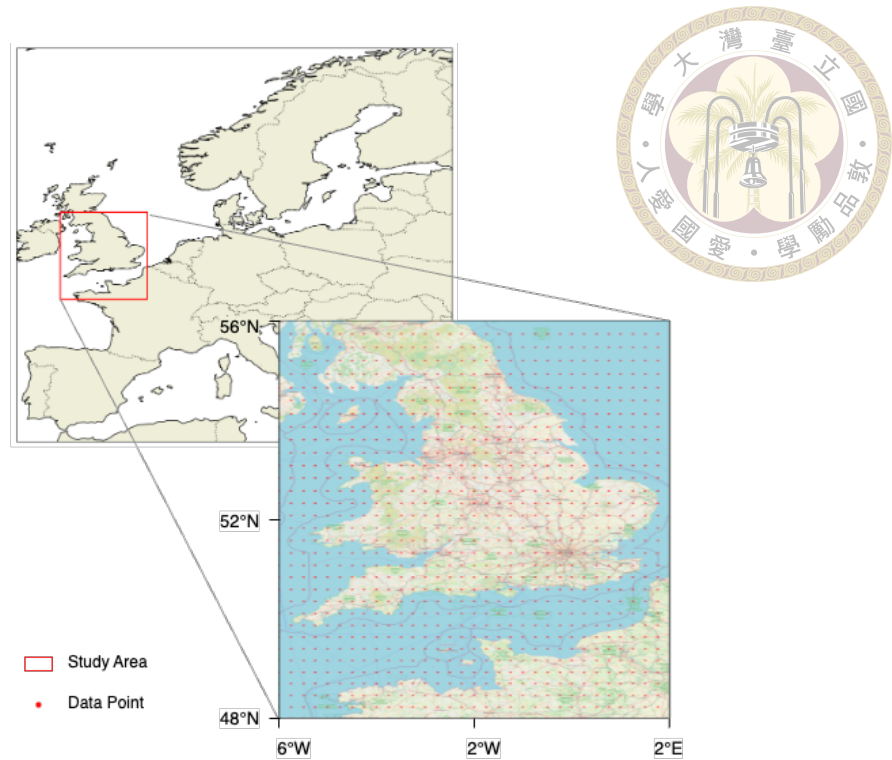
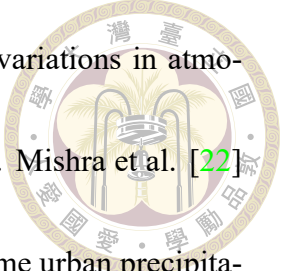


Figure 2.1 Geographic Area of the Study

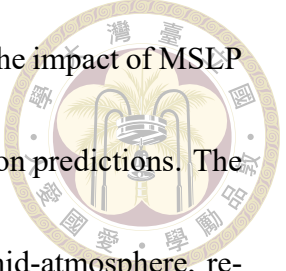
x 0.25° latitude-longitude), covering atmospheric, land, and ocean-related variables. For this research, data from 2001-2020 is used, focusing on the geographical area of the United Kingdom, defined by the latitude-longitude range (56°N, 6°W, 48°N, 2°E), as shown in Figure 2.1, providing the necessary spatial coverage for subsequent analysis.

2.1.1 Feature Selection

To effectively predict extreme precipitation events, it is crucial to select relevant and highly correlated meteorological variables. One such variable is the 2-meter dew point temperature, representing the moisture content in the air. A higher dew point temperature indicates more moisture, which is a precursor for precipitation. When it approaches or reaches the actual air temperature, the air becomes nearly saturated, increasing the like-



likelihood of precipitation. Changes in dew point temperature indicate variations in atmospheric moisture, which is crucial for forecasting precipitation events. Mishra et al. [22] emphasized the key role of dew point temperature in simulating extreme urban precipitation events, highlighting its significance in precipitation forecasting. Wind speed at the 850 hPa level is another critical variable for analyzing precipitation and storm movement. This level effectively reflects the horizontal transport of moisture and heat in the atmosphere, offering insights into moisture direction. As winds at 850 hPa are less influenced by surface friction, they provide consistent wind field data, aiding in understanding atmospheric fronts and pressure systems. Studies have shown that wind patterns at 850 hPa greatly influence moisture transport and distribution, affecting regional precipitation [23]. Additionally, Gregory et al. [24] noted the importance of 850 hPa wind speed in tropical cyclone tracking and its role in forecasting storm intensity and movement. Mean sea level pressure provides insights into the pressure field adjusted for sea level, removing elevation effects on pressure observations. MSLP is vital for understanding weather systems, as low-pressure areas typically bring rising air and precipitation, while high-pressure areas correlate with clearer weather. Changes in MSLP are used to identify cyclones and anticyclones, offering information on precipitation forecasts. Lavers and Villarini [25] discussed the relationship between atmospheric rivers and extreme precipitation events, noting the



critical role of MSLP in these occurrences. Wang et al. [26] analyzed the impact of MSLP changes on winter droughts, underscoring its importance in precipitation predictions. The 700 hPa level relative humidity is selected for its relevance in the mid-atmosphere, reflecting moisture distribution within convective systems. Humidity at this level provides insight into the availability of moisture for precipitation formation. A moist 700 hPa layer suggests favorable conditions for convective cloud development, which aids in predicting the intensity and extent of precipitation. Literature indicates that moisture availability at this level is crucial for understanding precipitation, particularly in convective systems that may result in extreme rainfall [27, 28]. In summary, this study extracts relevant variable data from the ERA5 dataset for the UK region between 1940 and 2020, including 2-meter dew point temperature, 850 hPa wind speed, mean sea level pressure, and 700 hPa relative humidity. Data extraction and preprocessing steps were carried out using Python scripts, including data download, spatial cropping, and time series generation (see Table 2.1 for details).

Table 2.1 Overview of the ERA5 Data

Variables (Symbol)	2-meter dew point temperature ($d2m$), 850 hPa wind speed (u, v), mean sea level pressure (msl), 700 hPa relative humidity (r)
Period	2001-2020
Temporal Resolution	1 hour
Spatial Coverage	6°W - 2°E / 48°N - 56°N (United Kingdom)
Spatial Resolution	0.25° (31 km)
Patch-size	32 × 32
Data Provider	ECMWF (European Centre for Medium-Range Weather Forecasts)

2.1.2 Exploratory Data Analysis

We aim to develop a model that uses time differences as a basis for measuring the similarity gap between atmospheric states, where the relationship between variables and time will have a direct impact on the model's training performance [29]. Specifically, we can expect that the more pronounced the variations over time or the stronger the relationships between variables, the more effectively the model can capture these changes, thereby successfully predicting the temporal distance between different atmospheric states [30]. Furthermore, the variations in atmospheric variables often exhibit characteristics such as long-term trends, seasonality, or randomness. In the following sections, we will apply several methods to observe the relationships between data and time. This will enable us to explore how these relationships influence the model's training outcomes in subsequent

research phases.

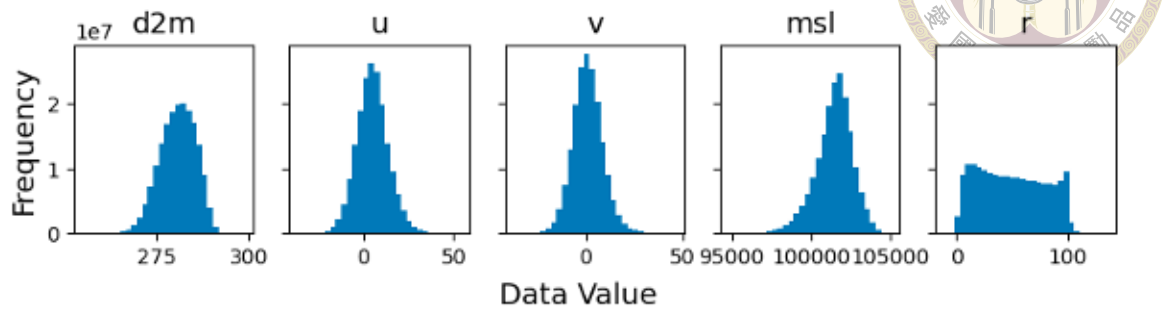


Figure 2.2 Distribution Characteristics of the Five Variables

Figure 2.2 illustrates the distribution characteristics of five variables. The first four variables exhibit a shape that is approximately normal; however, it is worth noting that the first and fourth variables display slight right-skewness, where the data extends somewhat to the right tail, creating a minor asymmetry. Despite this, such slight skewness is generally not expected to significantly impact model performance, particularly when the dataset is large, allowing the model to effectively learn the underlying patterns. In contrast, the fifth variable presents a “plateau-like” distribution, with values being more uniformly spread over a specific range, lacking a clear concentration or peak. This type of distribution could potentially hinder the model’s ability to leverage this feature for discrimination. Therefore, during model training, it is crucial to take into account both the distribution characteristics of the variables and their correlation with the target variable to determine if further handling of skewed or plateau-like distributions is necessary.

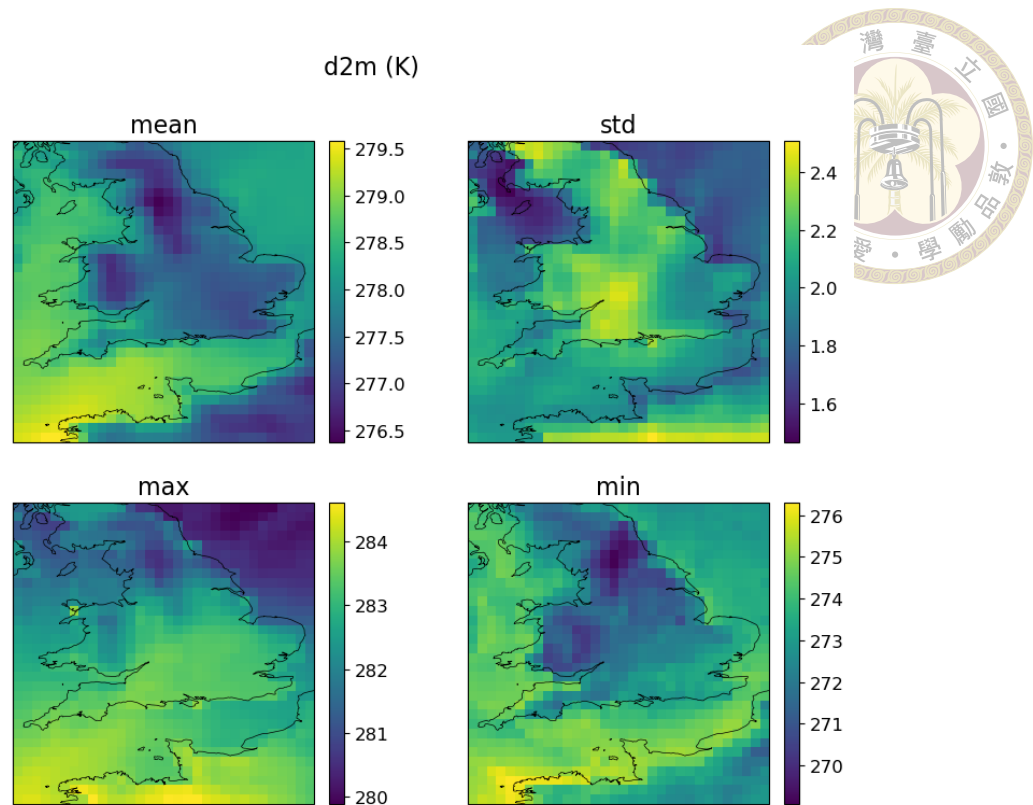


Figure 2.3 Spatial distribution of the statistical properties (mean, standard deviation, maximum, minimum) of $d2m$ averaged over entire period of 2001-2020

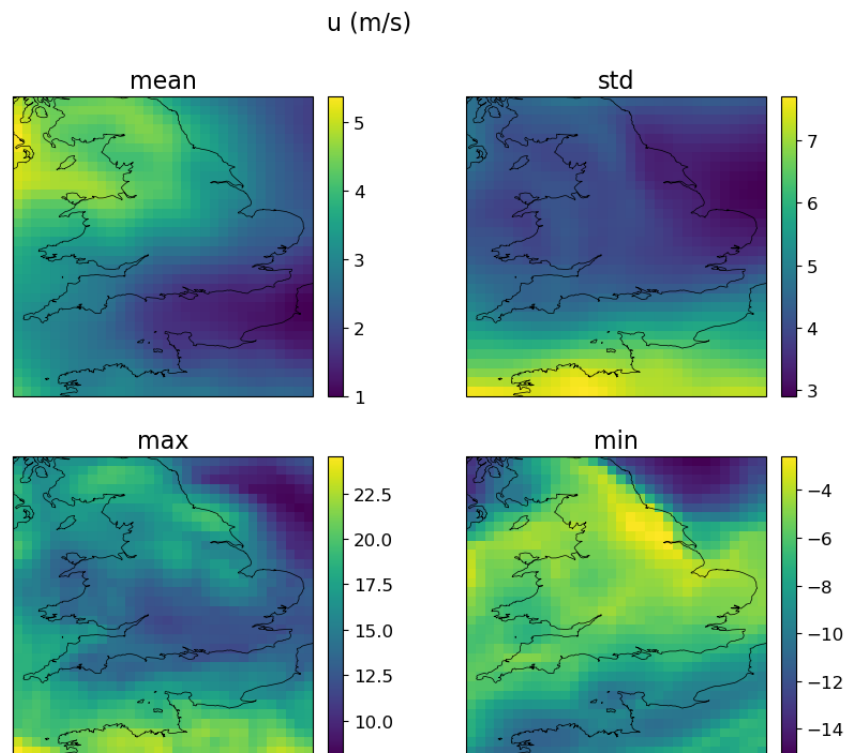


Figure 2.4 Spatial distribution of the statistical properties (mean, standard deviation, maximum, minimum) of u component of wind averaged over entire period of 2001-2020

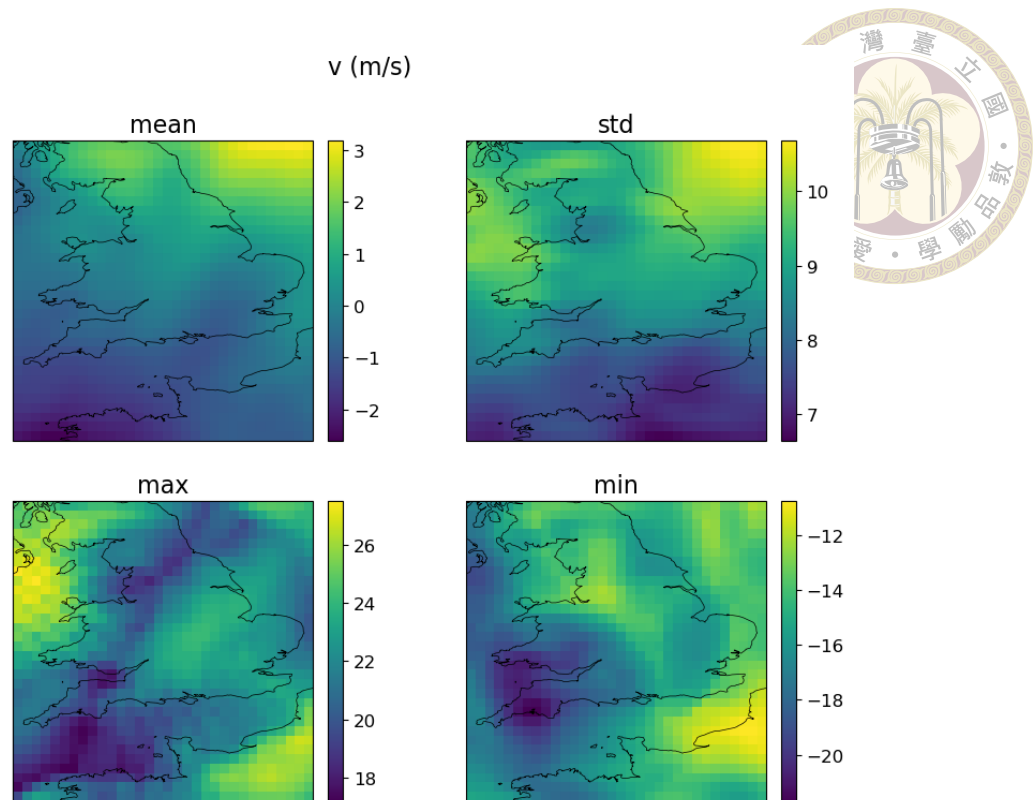


Figure 2.5 Spatial distribution of the statistical properties (mean, standard deviation, maximum, minimum) of v component of wind averaged over entire period of 2001-2020

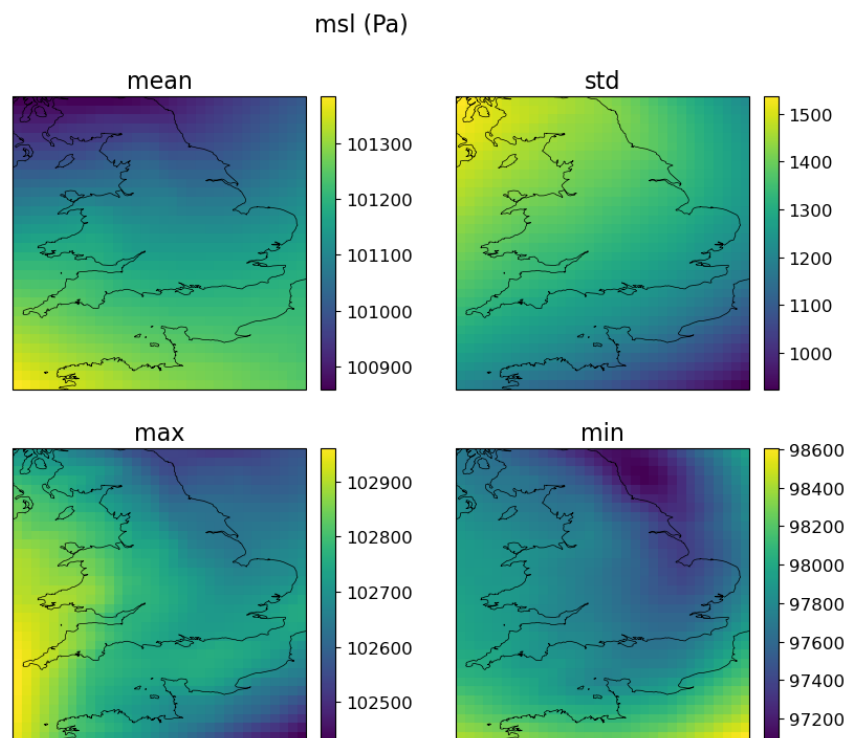


Figure 2.6 Spatial distribution of the statistical properties (mean, standard deviation, maximum, minimum) of msl averaged over entire period of 2001-2020

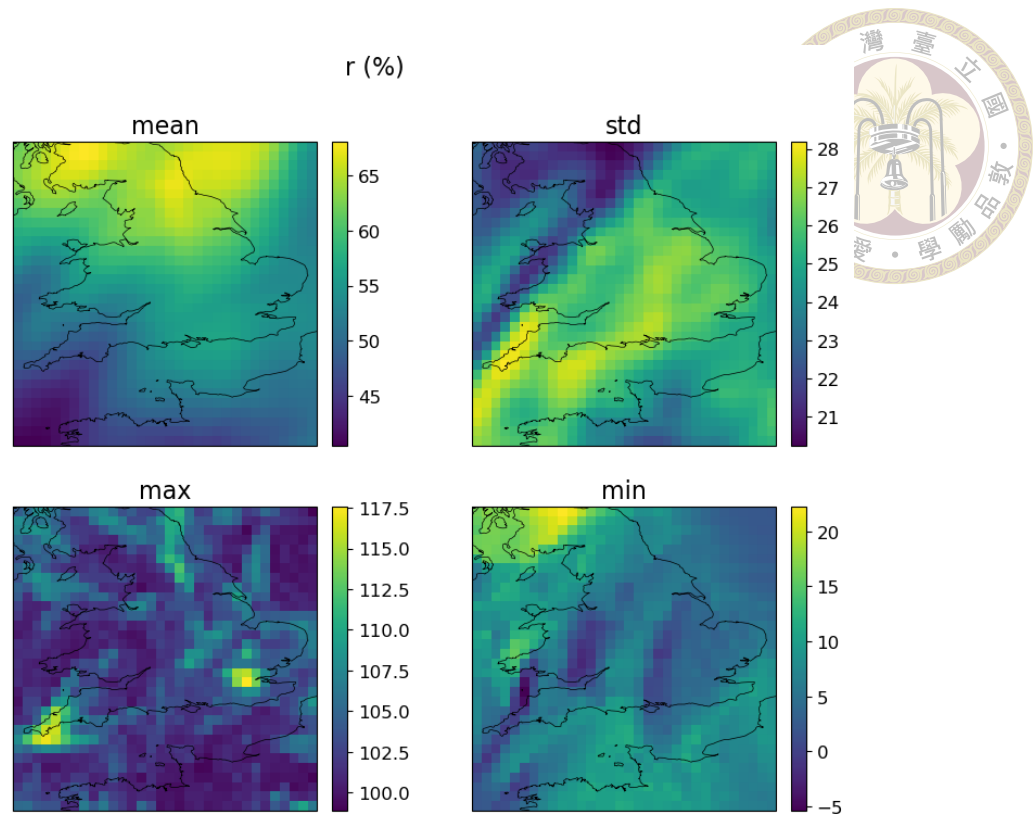
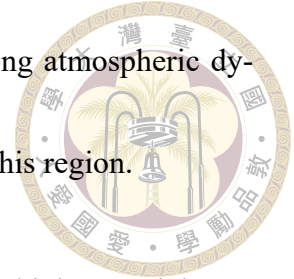


Figure 2.7 Spatial distribution of the statistical properties (mean, standard deviation, maximum, minimum) of r averaged over entire period of 2001-2020

The spatial variability of the five atmospheric variables can be clearly illustrated in Figures 2.3–2.7. The dew point temperature ($d2m$) and mean sea level pressure (msl) exhibit notable spatial stability across the considered region, but their spatial characteristics differ significantly. Dew point temperature ($d2m$) shows strong correlations with terrain, reflecting localized variations influenced by topography. In contrast, mean sea level pressure (msl) demonstrates higher spatial continuity and lower variability, highlighting a smoother spatial distribution. Relative humidity (r) and the zonal wind component (u) experience moderate spatial fluctuations, reflecting a certain degree of regional variability without extreme changes. Meanwhile, the meridional wind component (v) shows signif-

icant spatial variability, with abrupt color transitions indicating strong atmospheric dynamic processes that produce substantial wind speed fluctuations in this region.



In summary, $d2m$ is strongly influenced by terrain, msl exhibits higher spatial continuity and lower variability, r and u display moderate variability, whereas v is the most dynamic. These distinctions underscore the multi-scale spatial variability of different atmospheric variables within the same region.

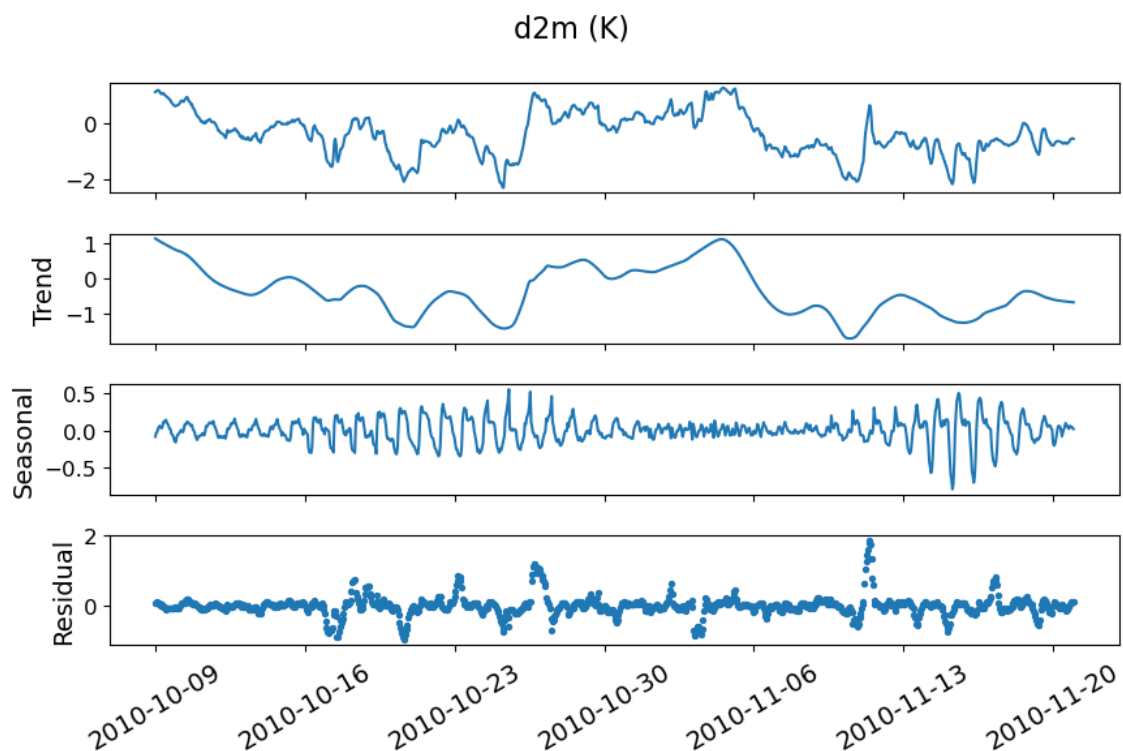


Figure 2.8 Seasonal Decomposition Plot of $d2m$

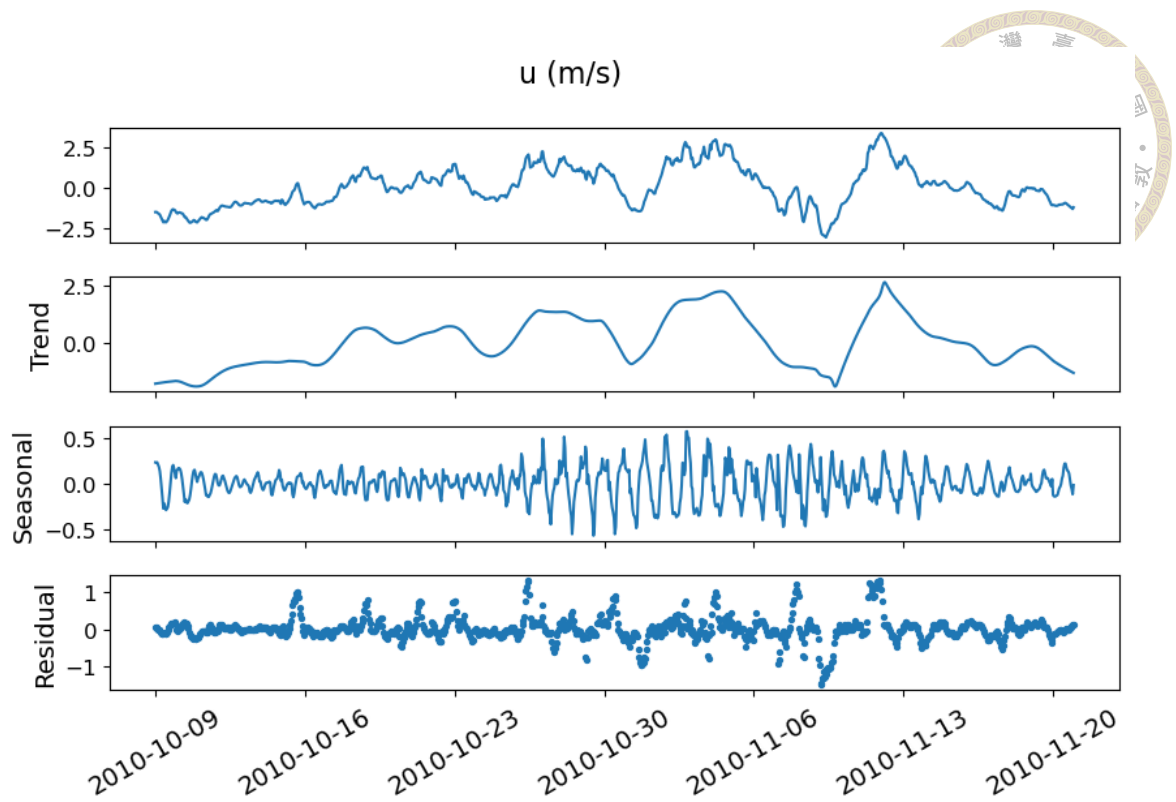


Figure 2.9 Seasonal Decomposition Plot of u

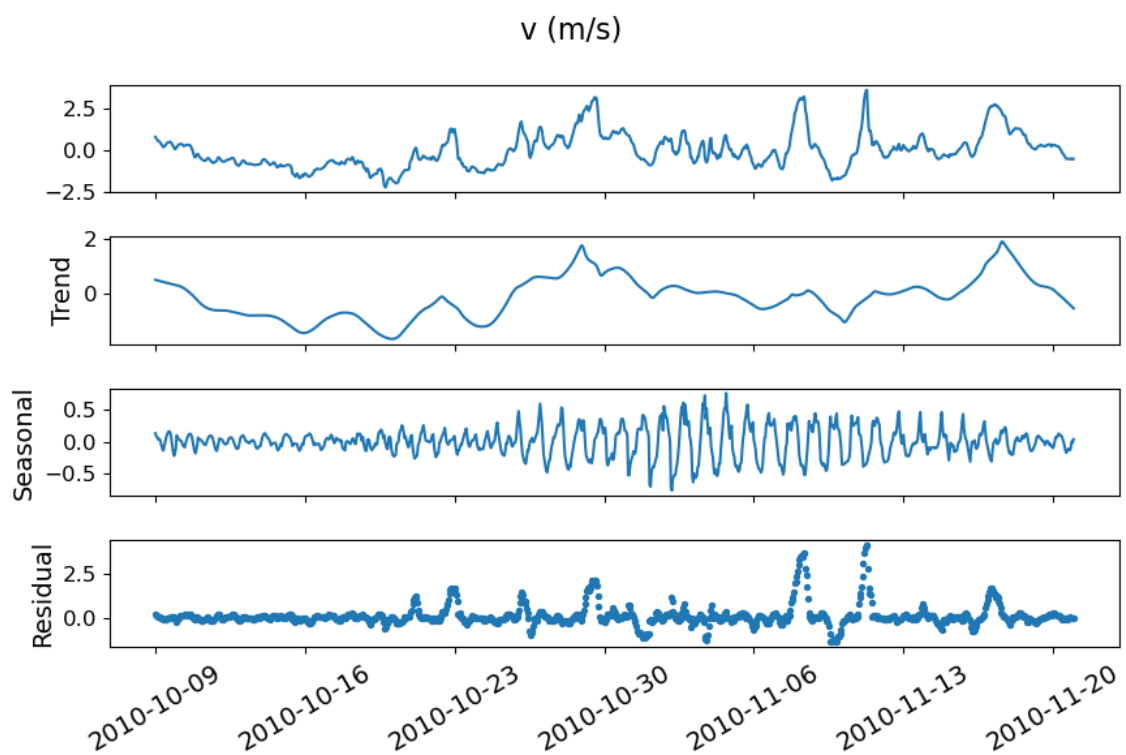


Figure 2.10 Seasonal Decomposition Plot of v

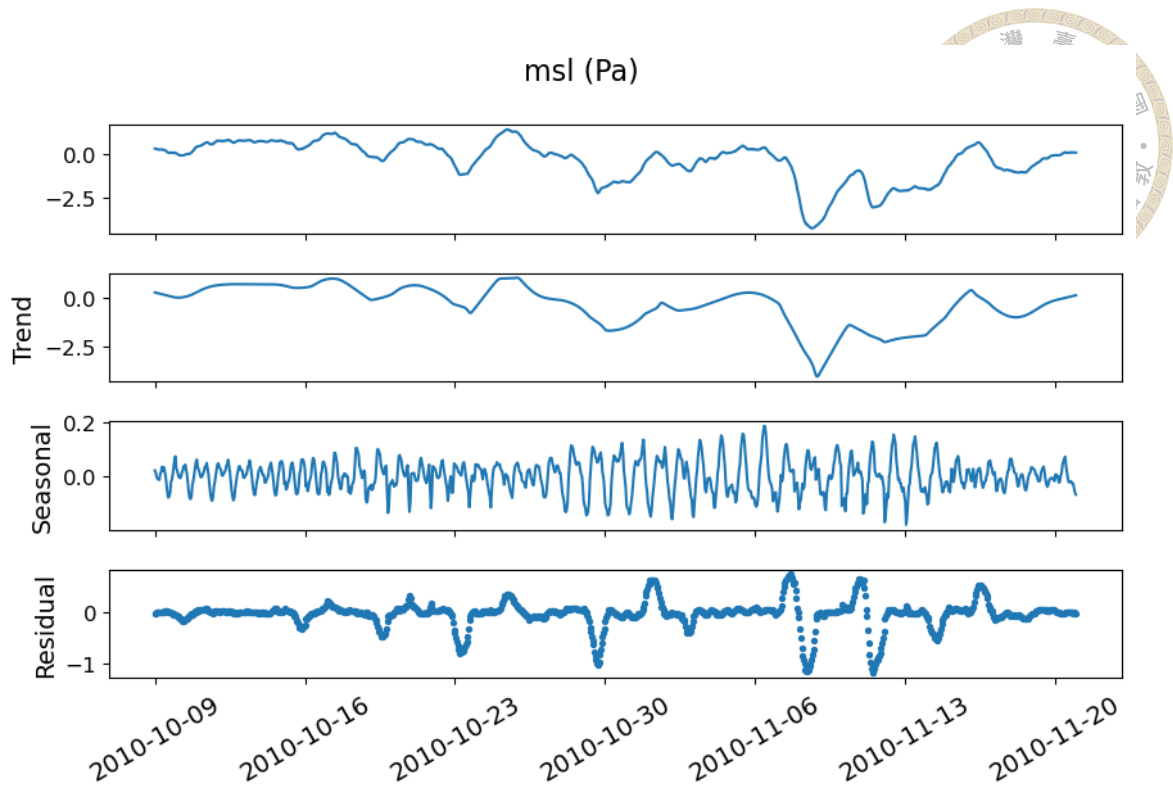


Figure 2.11 Seasonal Decomposition Plot of msl

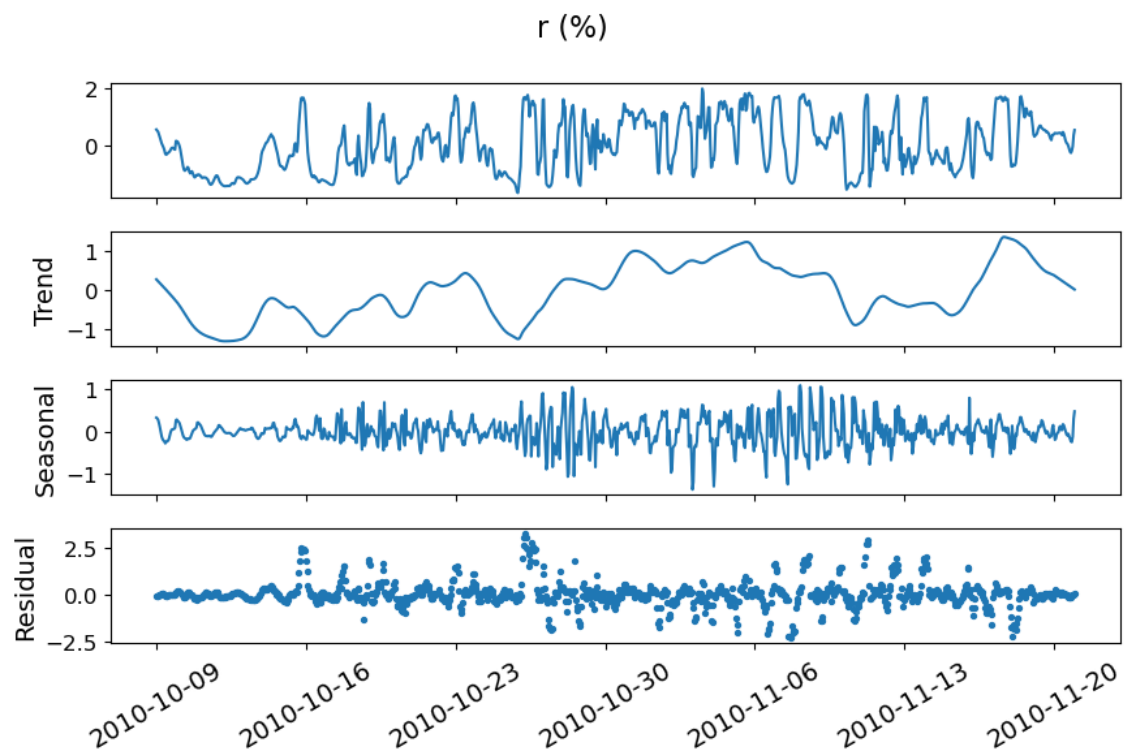


Figure 2.12 Seasonal Decomposition Plot of r

To gain deeper insights into temporal variations, seasonal decomposition is applied

to analyze the target region over the period from October 9, 2010, to November 20, 2010.

The parameters used for the seasonal decomposition are summarized in Table 2.2.



Based on the time series decomposition shown in the Figures 2.8–2.12, the five meteorological variables exhibit different levels of dispersion in their observed values, long-term trends, periodic components, and residuals. Relative humidity (r) has the widest observational range, spanning from 0% to 100%. Its periodic amplitude accounts for approximately one-fourth of the total variation, while residuals can fluctuate sharply over short periods, indicating high sensitivity to local weather conditions. Wind speeds (u, v) show periodic amplitudes covering about 20% of the total variation, but residuals can reach ± 25 m/s, nearly covering the entire observational range. This highlights the potential for extreme wind changes caused by fronts or convective systems.

Sea level pressure (msl), ranging between 97500 Pa and 102500 Pa, exhibits a relatively stable trend. Its diurnal periodic amplitude accounts for about 10% of the total variation, but residuals can reach up to half of the range, implying noticeable fluctuations when fronts or low-pressure systems pass through. In comparison, $d2m$ varies between 270K and 285K, with a range of about 15K. Its periodic component accounts for approximately one-third of the variation, and residuals reach up to two-thirds, indicating greater stability than u and r .

Overall, wind speeds (u and v) and r exhibit the highest temporal dispersion, followed by msl , while $d2m$ is relatively stable. These results reflect their varying sensitivity to transient weather changes and highlight the need to pay special attention to rapid fluctuations in wind fields (u and v) and r for future forecasting or control strategies.

Table 2.2 Settings for Seasonal Decomposition

Parameter	Value
Time Resolution	1 Hour
Period (period)	24 (1 Day)
Seasonal Window (seasonal)	7 (Default)
Trend Window (trend)	Auto-calculated: $1.5 \cdot period / (1 - 0.15)$
Robust (robust)	True
Selected Time Interval	2010/10/09 – 2010/11/20

2.2 Met Office Rain Radar Data from the NIMROD System

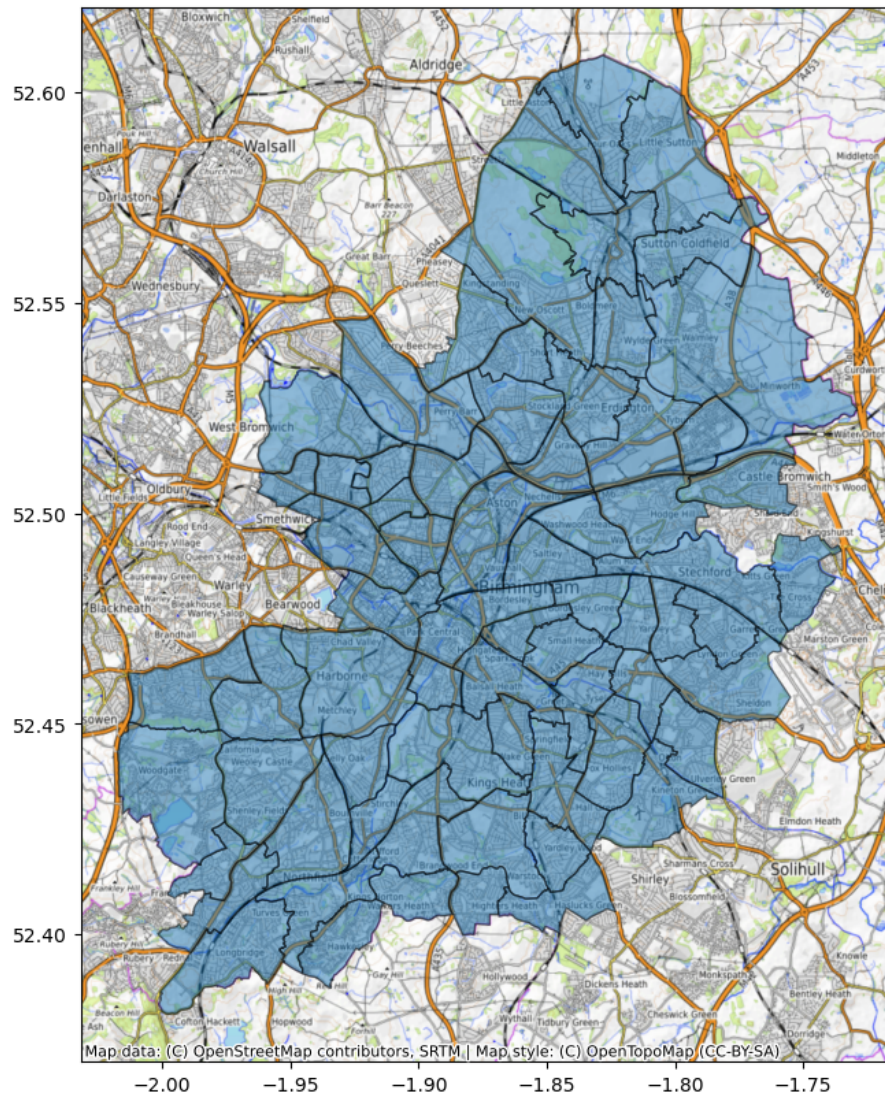


Figure 2.13 Nimrod Radar Precipitation Coverage for the Birmingham Region

Accurate precipitation observation data is critical for predicting extreme precipitation events. To capture true local precipitation, this study uses the Nimrod radar dataset developed and managed by the UK Met Office. Nimrod is an advanced radar network offering

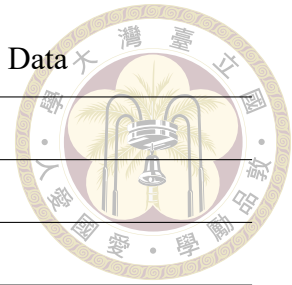
high spatial (1 km) and temporal (5-minute) resolution, capable of capturing precipitation event details and covering all of the UK, making it valuable for nationwide meteorological analysis. The study focuses on the Birmingham region, spanning from -2.017°W to -1.729°W and 52.381°N to 52.609°N (see Figure 2.7) to ensure detailed monitoring of precipitation.

The Nimrod dataset has been widely applied in numerous studies and practical weather forecasting operations. For instance, Roberts and Lean [31] utilized Nimrod data to validate rainfall accumulations in high-resolution forecasts of convective events, demonstrating its value in improving short-term precipitation forecast accuracy. Golding [32] discussed the use of Nimrod data in quantitative precipitation forecasting in the UK, highlighting its importance in flood warning and risk management.

In this study, Nimrod radar precipitation data (see Table 2.3 for details) will be used for downscaling hourly accumulated precipitation, allowing for a more accurate capture of localized precipitation results. By leveraging the high-resolution Nimrod data, this research aims to enhance the prediction capability of extreme precipitation events, providing a more precise scientific foundation for weather forecasting and disaster management.

Table 2.3 Overview of the Nimrod Radar Precipitation Data

Variables	Precipitation intensity
Time Period	2004-2020
Temporal Resolution	5 minutes
Spatial Coverage	Birmingham (-2.017°W to -1.729°W, 52.381°N to 52.609°N)
Spatial Resolution	1 km
Data Provider	UK Met Office



2.3 Extreme Precipitation Events



This study analyzed 157 precipitation events in the Birmingham region, selected based on criteria for intense precipitation. The selection process involved matching recorded flood occurrences with corresponding rainfall events to ensure each case represented a significant hydrometeorological interaction.

Rainfall data were sourced from the UK Met Office's C-band radar (1 km spatial resolution, 5-minute temporal resolution) covering a 34×30 km area in OSGB 1936/British National Grid coordinates. To maintain consistency, only flood records that spatiotemporally overlapped with storm events were included. Specifically, a flood record was assigned to a storm event if it occurred within the event's duration, reinforcing the link between precipitation intensity and observed flooding.

These events serve two primary purposes in this research: first, they enhance the accuracy and robustness of similarity metric calculations during the first phase of the study by providing a diverse and representative dataset for analog searches. Second, the events' temporal occurrence is used in the second phase to generate radar-based precipitation distribution maps, enabling further analysis of rainfall patterns and intensity.

While Table 2.4 in the main text presents a subset of the selected events, showcas-

ing the top five representative precipitation events, the complete list of all 157 events is included in Appendix A.1 for reference. This detailed listing provides a comprehensive overview of the dataset used in this study, offering additional context for the analysis.



Table 2.4 Extreme Precipitation Events (First Five)

Rank	ID	Start Datetime	End Datetime	Duration (hrs)
1	2464	2014-07-08 10:00:00	2014-07-08 20:00:00	10.0
2	632	2006-08-29 11:00:00	2006-08-29 14:00:00	3.0
3	1998	2012-07-18 05:00:00	2012-07-19 04:00:00	23.0
4	481	2005-12-31 20:00:00	2006-01-01 04:00:00	8.0
5	2879	2016-04-04 06:00:00	2016-04-04 16:00:00	10.0





Chapter 3 Methods

3.1 Framework of Two-Phase Weather Analog Method

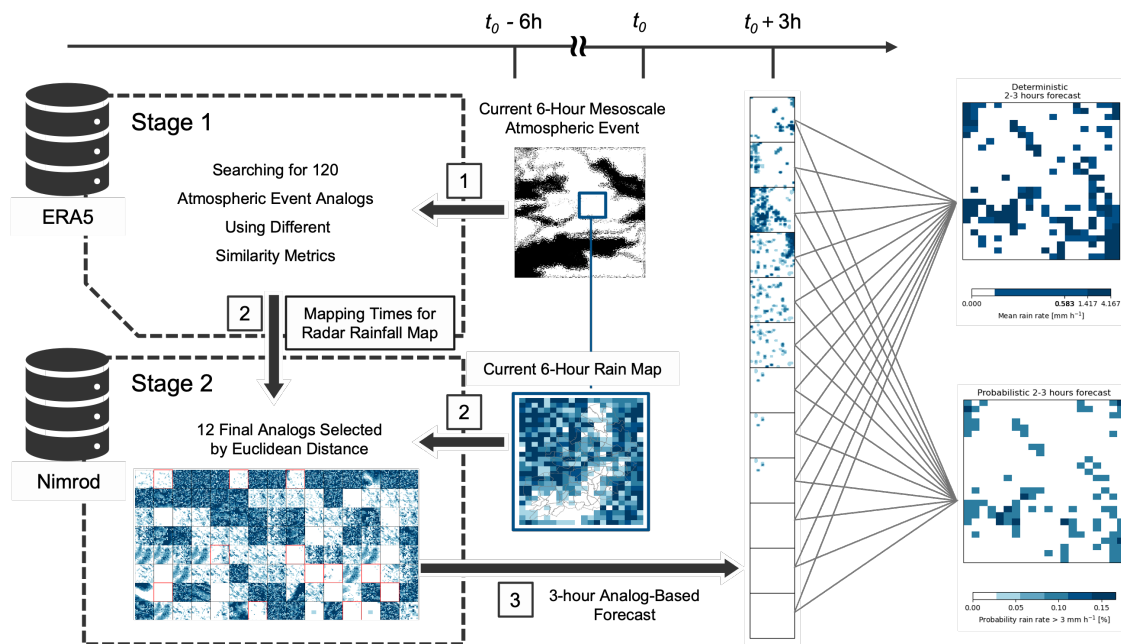


Figure 3.1 Framework of Two-Phase Weather Analog Method

The framework for the two-phase weather analog method includes three main parts: (1) Identify 120 analogs for the target event from the ERA5 historical dataset; (2) Extract radar rainfall maps for both the target event and the 120 analogs, then refine the selection to 12 final analogs based on euclidean similarity; (3) Generate rainfall forecasts for the next 2-3 hours using these 12 final analogs, producing both deterministic and probabilistic predictions.

The two-phase weather analog method aims to provide high-resolution predictions for extreme precipitation events by identifying atmospheric analogs similar to the cur-

rent state. In the first phase, historical ERA5 data is searched to find the 120 closest atmospheric states, capturing complex patterns across multiple variables like wind, temperature, and humidity. The second phase extracts radar rainfall maps for these analogs, narrowing them down to 12 final representatives using high-frequency Nimrod radar data, which aids downscaling.

This method demonstrates computational efficiency by balancing mesoscale ERA5 reanalysis (with a 1-hour resolution) for broad analog searches, and the 5-minute resolution of Nimrod data for localized refinement. Such an approach allows for 2-3 hour forecasts of 6-hour events, delivering both deterministic and probabilistic predictions.

Literature findings indicate this method significantly improves short-term extreme precipitation forecasts, particularly in complex terrains [21]. This demonstrates the two-phase analog framework's efficiency in computational resources, downscaling, and local feature capture, making it a reliable tool for extreme weather prediction.

Next is "Parameter Settings," crucial for searching analogs. The "time window" defines the period over which analogs are searched; a large window may introduce irrelevant data, while a narrow one could miss essential patterns [33].

For a detailed overview of the first phase, refer to Figure 3.2. The first phase of

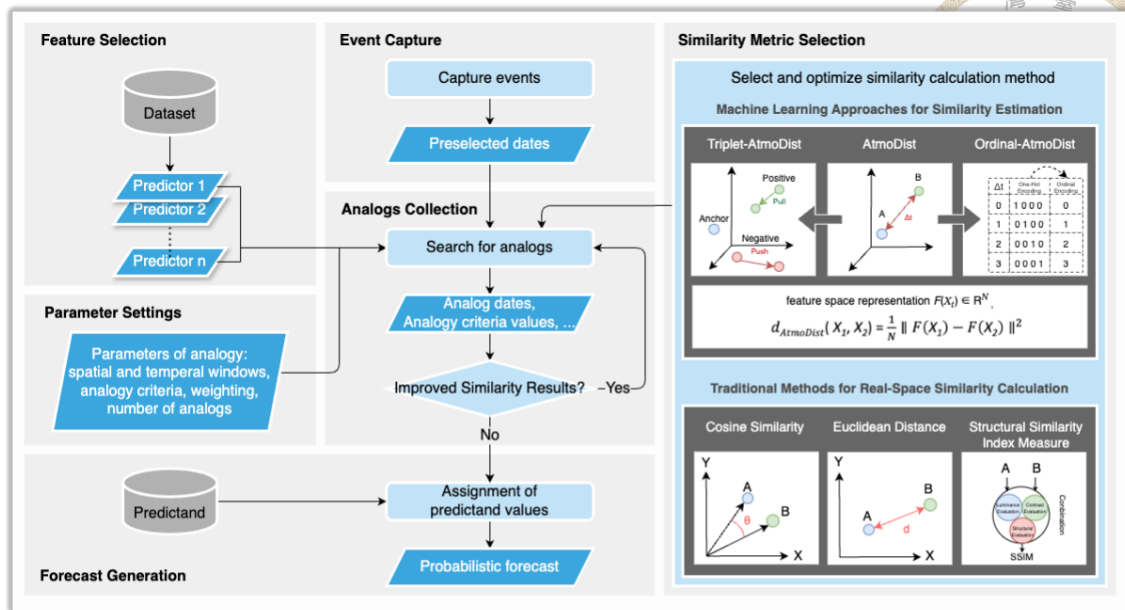


Figure 3.2 Workflow of Weather Analog Method (Adapted from [28])

the two-phase weather analog method consists of six key components. Initially, “Feature Selection” and “Event Capture” focus on identifying the meteorological variables and time ranges of target events, as outlined in the previous section. The variables, resolutions, lead time, event window, and number of analog members used in this study are summarized in Table 3.1. These settings provide the foundation for constructing a reliable analog-based forecasting framework.

Parameter Settings are crucial for searching analogs. The “time window” defines the period over which analogs are searched; a large window may introduce irrelevant data, while a narrow one could miss essential patterns [33]. Next is “spatial window” specifies the geographic range, tailored to the event’s scale. If too large, unnecessary spatial variability is introduced, while too small could overlook broader weather systems

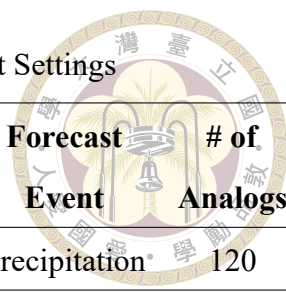
that impact local weather [34]. This study employs spatial weights to capture appropriate spatial information effectively [35].



Additional settings include “Weight Settings,” where the weights balance the contribution of each variable to similarity, enhancing the predictive value of the analogs. Lastly, the number of predictive variables is an essential aspect of ensuring that the final forecast reflects accurate weather conditions.

“Analog Collection” follows parameter settings. Once the parameters are defined, analogs are searched using specific similarity metrics to measure the resemblance between current and historical events. The choice of similarity metric is pivotal, as different metrics capture atmospheric dynamics differently [36]. Weights from parameter settings further refine the search, balancing variables’ influence on overall similarity [37]. This research will explore these considerations in-depth, including parameter choices and similarity measures. Finally, the selection and validation of analogs are iterative. By adjusting similarity metrics, weights, and parameters, continuous testing ensures optimal analog selection, improving the accuracy of the final forecast. This iterative process is vital to achieving precise and reliable weather predictions, which culminate in the generation of the required forecast based on the target predictive variables [38].

Table 3.1 Summary of Variables, Resolutions, and Forecast Settings



Variable(s)	Time Resolution	Spatial Resolution	Lead Time	Event Window	Forecast Event	# of Analog
<i>d2m, u, v, msl, r</i>	1-hourly	0.25°	3 hr	6 hr	Precipitation*	120

3.2 Traditional Methods for Real-Space Similarity Calculation



This chapter delves into traditional methods used for the calculation of real-space similarity in atmospheric data. It is structured into two main parts: the first part examines the role of event analogs and emphasizes the influence of variable weighting on similarity calculations for analog searches in weather events; the second part focuses on the specific similarity metrics and indicators used to quantify and identify these atmospheric analogs. Through a detailed exploration of both parts, the chapter aims to provide a comprehensive understanding of how similarity calculations drive the search for comparable weather events.

3.2.1 Weighting for Similarity Calculation

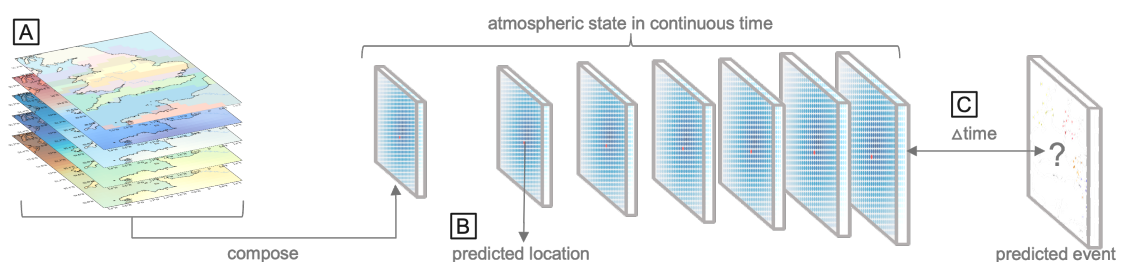


Figure 3.3 Weighted Multi-Factor Framework for Atmospheric Event Prediction

The figure illustrates key factors in weather event prediction: A. Various atmospheric variables are utilized as features, B. Spatial distribution highlights that atmospheric states closer to the target location are more influential, and C. Temporally, atmospheric states closer to the event time are more relevant to the prediction.

In this study, we develop a weighting scheme (see Figure 3.3) to effectively measure

the similarity between two atmospheric events, incorporating three key factors: variables, spatial grid points, and temporal dimensions. This approach builds on the methodologies presented by [20], which provided a foundation for developing our model. The following sections elaborate on the theoretical foundation and advantages of each weighting factor.

3.2.1.1 Formula-Driven Weight Calculation for Atmospheric Data

As referenced in [20], this study utilizes ERA5 data to analyze 157 precipitation events, treating them as a continuous time series for comprehensive evaluation. These weights are essential for capturing the relative importance of different variables and regions, ensuring a more accurate assessment of the events within the study framework.

Variable Weighting To determine the relative contribution of atmospheric variables to precipitation events, we calculate the correlation of each variable with precipitation. The variable weight is determined as follows:

$$w(j) = \frac{|R_j|}{\sum_{j=1}^v |R_j|} \quad (3.1)$$

where $w(j)$ is the weight of variable j , R_j is the correlation coefficient between variable j and precipitation, and v is the total number of variables. This ensures that variables

with the greatest impact on precipitation are assigned higher weights in the prediction model, enhancing its accuracy and interpretability [28].



By applying variable weighting, the model differentiates the impact of various variables on precipitation, allowing for a clearer understanding of their relative contributions to the event[37].

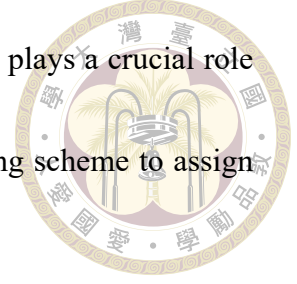
Spatial Weighting To capture the contribution of different geographic locations to the precipitation at a target location, we compute the correlation between the precipitation at each grid point and the target location:

$$G(i) = \frac{|r_i|}{\sum_{i=1}^n |r_i|} \quad (3.2)$$

Here, $G(i)$ is the weight of grid point i , r_i is the correlation between the precipitation at grid point i and the target location, and n is the total number of grid points. This design ensures that the model focuses on geographic areas with significant impacts on the precipitation at the target location.

Spatial weighting enhances the model's sensitivity to critical geographic regions, allowing for more precise identification of spatial features that are important for the overall atmospheric state[34].

Temporal Weighting Temporal evolution of the atmospheric state plays a crucial role in predicting precipitation events. We introduce a temporal weighting scheme to assign greater weight to time periods closer to the event being predicted:



$$\gamma_p(k) = \frac{k}{\sum_{i=l}^{l+t_w-1} i} \quad \text{for } k = l, l+1, \dots, l+t_w-1 \quad (3.3)$$

where $\gamma_p(k)$ is the weight for time point k , l is the starting lead time, and t_w is the time window. This allows the model to focus more on the recent atmospheric state changes, enhancing sensitivity to the timing of the event.

Temporal weighting significantly influences the model's predictive capabilities across various time scales, ensuring that both short-term and long-term atmospheric changes are adequately captured[36].

In summary, this similarity evaluation approach, incorporating variable, spatial, and temporal weighting, significantly enhances the model's interpretability and provides a solid theoretical foundation for accurate prediction of extreme precipitation events.

3.2.1.2 Analysis of Weight Calculations for Spatiotemporal and Variable Components

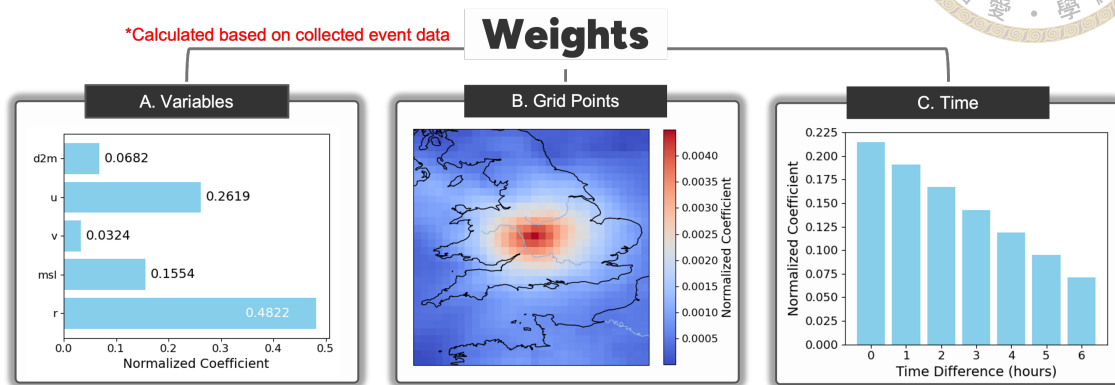


Figure 3.4 Weights Derived from Precipitation Event Data

A. Variables: Normalized correlation coefficients between meteorological variables and precipitation. B. Grid Points: Spatial distribution of correlation coefficients for precipitation. C. Time: Temporal weights based on time difference from precipitation events.

The variable weights are calculated based on the correlation between 157 precipitation events and each meteorological variable [35]. From the diagram, we observe that the variable r , which likely represents relative humidity or rainfall, has the highest correlation with precipitation events, reaching 0.4822. This indicates that r has the strongest association with precipitation predictions [39]. The next most significant variable is u , the zonal wind component, with a correlation of 0.2619, suggesting that horizontal wind plays a substantial role in the precipitation process, particularly in cases involving frontal activity or cyclone movement, where wind direction impacts the transport of moisture [40]. The variable msl , representing mean sea level pressure, has a correlation of 0.1554, indicating that pressure changes are somewhat related to precipitation, as low-pressure systems are often

associated with precipitation events. Other variables, such as v (meridional wind component) and $d2m$ (dew point temperature), show relatively lower correlations, suggesting that vertical wind components and dew point have a smaller influence on precipitation in these events [41]. Overall, humidity and horizontal wind components exhibit a strong influence, closely linked to moisture supply and transport processes [42].

The spatial weights are computed based on the correlation between precipitation at the target location and precipitation at all other locations [43]. The diagram shows that the highest weights are concentrated in central England, indicating that precipitation at this region has the strongest correlation with the target location [44]. As the distance from the center increases, the spatial weights gradually decrease, aligning with the regional nature of precipitation events. Locations farther from the center show lower correlation with the target area [45]. This may reflect the influence of local topography, the extent of regional weather systems, or the impact of frontal systems [46]. The highest weights in central England likely correspond to frequent influences from Atlantic systems in this region, as well as frontal activity concentrated in this area [47].

The time weights, as shown in the graph, increase as the time lag from the precipitation event grows. Notably, beyond eight hours, the weights rise significantly, ultimately reaching around 0.14. This trend reflects the assumed logic in the model that, as weather

systems approach a region, the probability of a precipitation event gradually increases [48]. Therefore, the time weights represent the hypothetical relationship between temporal proximity to a precipitation event and the likelihood of that event occurring: the closer the time is to the precipitation event, the more important the meteorological variables become in predicting precipitation [49].

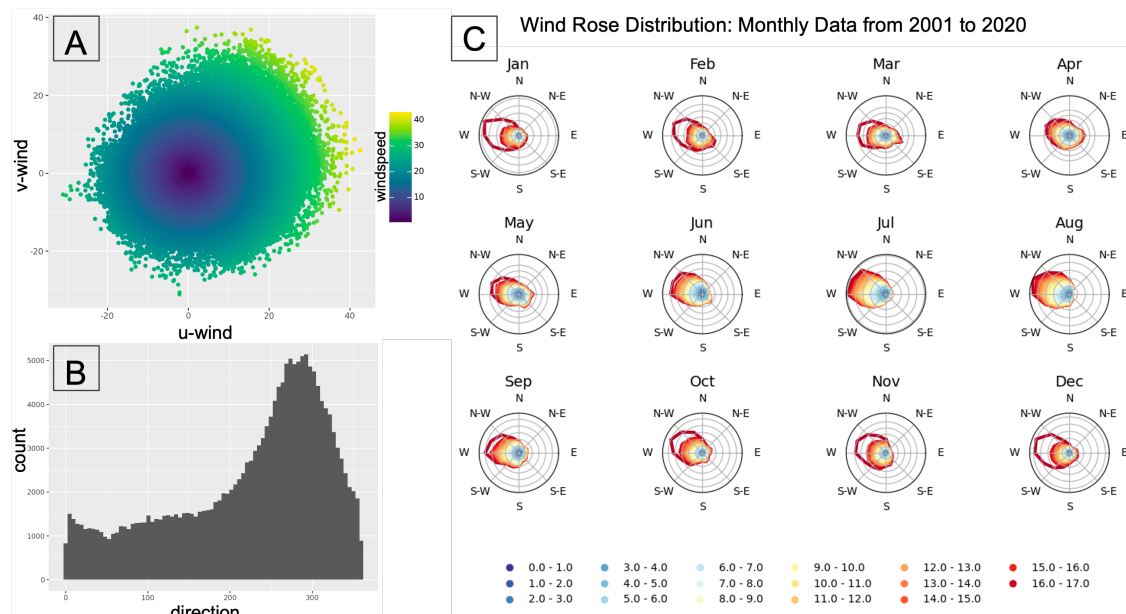


Figure 3.5 Wind Speed and Direction Distribution

A: Wind speed distribution plotted against the u-wind (zonal wind) and v-wind (meridional wind) components. B: Wind direction distribution histogram showing the frequency of wind directions. C: Monthly wind rose diagrams, displaying the predominant wind directions and speeds for each month from January to December.

To better understand the differences in correlation between the u-wind and v-wind components with precipitation, as well as the spatial distribution of precipitation weights, we analyzed the monthly wind patterns from 2001 to 2020 (refer to Figure 3.5). The relationship between the u-wind (zonal wind component) and v-wind (meridional wind com-

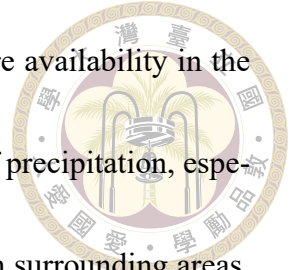
ponent) with precipitation is influenced by the role of wind direction in moisture transport.

When wind directions align with major moisture sources, such as the Atlantic Ocean or the North Sea, these wind components play a critical role in the precipitation formation process. The wind speed distribution reveals that stronger winds are primarily concentrated in the east-west direction, explaining the higher correlation of u-wind with precipitation in the variable weighting analysis. The u-wind facilitates the transport of moisture from the ocean to central England and surrounding regions, whereas the v-wind plays a relatively weaker role in this process.

Wind direction not only affects variable weights but also directly influences the spatial distribution of precipitation. Figure B and the wind rose diagrams for each month in Figure C show the seasonal variability in wind direction, which regulates the spatial manifestation of precipitation. Due to the consistently stronger westerly winds throughout the year, these regions receive direct moisture transport from source areas, resulting in a more pronounced horizontal distribution of precipitation events, as reflected in the spatial correlation.

In summary, these three types of weights together illustrate the multidimensional factors influencing precipitation prediction. The variable weights highlight the critical roles of relative humidity and horizontal wind speed, particularly the strong correlation between

humidity and precipitation, demonstrating the importance of moisture availability in the atmosphere. The spatial weights reveal the regional characteristics of precipitation, especially in central England, where precipitation correlates strongly with surrounding areas, reflecting the influence range of weather systems in this region. Finally, the time weights illustrate the hypothetical relationship between time and the occurrence of precipitation events in the model. Together, these three dimensions provide a comprehensive framework for improving the accuracy of precipitation predictions.



3.2.2 Comprehensive Similarity Calculation Formula for Weather Event Analogs

In atmospheric research, determining the similarity between atmospheric states at different time points is a critical analytical task, crucial for predicting weather patterns, studying climate change, and conducting environmental monitoring. This study examines methods for calculating atmospheric data similarity based on real-space values and compares three different metrics: Euclidean distance, Cosine similarity, and Structural Similarity Index (SSIM). These metrics are used to evaluate the similarity of weather data over time. The following sections introduce various similarity indices and the integrated formula ultimately used for finding analogs of atmospheric events.

3.2.2.1 Different Types of Similarity Metrics



Euclidean Similarity In this study, we utilize the Euclidean distance to quantify the similarity between weather data across different events. As an intuitive and computationally simple metric, Euclidean distance is applicable across domains like signal processing, data analysis, and machine learning [50] to effectively evaluate similarities between atmospheric states.

Atmospheric data typically consists of variables such as temperature, humidity, and wind speed, recorded at specific geographic coordinates. Each atmospheric state can thus be represented as a multidimensional vector, encompassing both these variables and their temporal evolution.

Görner et al. [51] introduced a multivariate, non-parametric Euclidean distance model to decompose daily climate data into hourly intervals. By selecting historical weather analogs based on similarity, their study included multiple climate variables (e.g., temperature, precipitation, and wind direction) and calculated Euclidean distances at various timescales to refine daily climate data into a higher temporal resolution. The application demonstrated that time-series-based Euclidean distances effectively capture differences in climate variables during the decomposition process.

Considering the temporal and spatial properties of atmospheric data, we aggregate each variable's data over a given time period and then calculate the Euclidean distance. This approach allows for efficient distance measurement and comparison of multi-channel, multidimensional weather data. The formulas are as follows:

$$\text{distance}_{\text{euclidean}} = \sqrt{\sum_{i=1}^n (G(i)A_i - G(i)B_i)^2} \quad (3.4)$$

where $G(i)$ represents the weight of each spatial grid point i , and A_i and B_i are the corresponding variables at each point. To convert this distance into a similarity score, we use an exponential function:

$$s_{\text{euclidean}} = \exp(-\text{distance}_{\text{euclidean}}) \quad (3.5)$$

This transformation maps the Euclidean distance into a similarity measure $s_{\text{euclidean}}$, where higher values indicate greater similarity.

Cosine Similarity It provides a method to measure the similarity between two points within a multi-dimensional dataset . Compared to other distance metrics, cosine similarity focuses on the direction rather than the magnitude of the data vectors, which makes it

particularly stable for measuring the similarity between vectors. This is highly beneficial given the complexity and multi-dimensionality of atmospheric data [52].



In meteorology and oceanography, cosine similarity has been employed to compare patterns between datasets, such as when analyzing the impact of turbulence on the refractivity of evaporation layers. By comparing vertical profiles of refractivity against observational data, researchers can more accurately assess model performance, thus improving forecast accuracy and reliability [53].

In this study, we apply weighted cosine similarity for each meteorological variable to compare multi-dimensional weather data over different time intervals. Cosine similarity evaluates the cosine of the angle between two vectors to assess their similarity. For weather data with four dimensions (time, variable, latitude, and longitude), we first aggregate the data for each variable over the given time period and then compute the cosine similarity. The formula is as follows:

$$\text{similarity}_{\cos} = \frac{\mathbf{A} \times \mathbf{B}}{\|\mathbf{A}\| \times \|\mathbf{B}\|} = \frac{\sum_{i=1}^n G(i)A_i G(i)B_i}{\sqrt{\sum_{i=1}^n [G(i)A_i]^2} \sqrt{\sum_{i=1}^n [G(i)B_i]^2}} \quad (3.6)$$

$$s_{\cos} = 1 - \frac{\cos^{-1}(\text{similarity}_{\cos})}{\pi} \quad (3.7)$$

SSIM (Structural Similarity Index Measure) Traditionally, image quality assessment methods rely heavily on error visibility models, such as Peak Signal-to-Noise Ratio (PSNR) and Mean Squared Error (MSE). However, these metrics often fail to effectively capture perceived image quality from a human visual perspective. The Structural Similarity Index (SSIM) proposed by Wang et al. evaluates image quality by comparing luminance, contrast, and structural information, thereby better aligning with human visual perception [54].

To expand SSIM's applications, Baker et al. [55] explored its use in floating-point data, particularly within climate datasets. By analyzing and comparing the different characteristics of floating-point data, they introduced an improved SSIM-based method to measure data similarity. This approach extends beyond image quality assessment to atmospheric data, enhancing the precision and reliability of climate models.

In our study, we apply the SSIM metric to assess the similarity of multi-dimensional atmospheric state data, adapting it to handle complex weather data with spatiotemporal characteristics. While the traditional SSIM model compares luminance, contrast, and structure between two-dimensional images, we extend these considerations to multi-dimensional atmospheric data. Each channel's SSIM is computed separately and then averaged to obtain an overall SSIM value, integrating these measures into the extended SSIM formula

for multi-variable weather data.



$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (3.8)$$

$$s_{\text{ssim}} = \text{SSIM}(\mathbf{A} \cdot \mathbf{G}, \mathbf{B} \cdot \mathbf{G}) \quad (3.9)$$

- μ_x and μ_y are the mean values of x and y , respectively.
- σ_x^2 and σ_y^2 are the variances of x and y , respectively.
- σ_{xy} is the covariance of x and y .
- C_1 and C_2 are two constants introduced to stabilize the denominator.

This approach enables a comprehensive and detailed analysis of the data structure, allowing us to identify and evaluate similarities and differences in atmospheric data across various time points. It facilitates a deeper understanding of how atmospheric conditions change over time and enhances the model's ability to predict and analyze these variations effectively.

3.2.2.2 Similarity Calculation Formula for Weather Events



The similarity assessment of atmospheric events is a complex task that requires robust mathematical approaches. The Euclidean distance provides a straightforward, intuitive measure of similarity by calculating the geometric distance between data points in multi-dimensional space. This is particularly helpful when dealing with direct differences in atmospheric variables across spatial and temporal dimensions. Cosine similarity, on the other hand, evaluates the directional alignment between vectors, focusing on the shape of the data rather than its magnitude. This allows for a more stable measurement when data amplitude varies significantly but maintains a consistent pattern or trend. Lastly, SSIM is a more complex metric that compares data based on structural properties, effectively capturing the perceived changes in data patterns. SSIM is well-suited for cases where spatial distribution and relationships between data variables are crucial, especially when considering temporal shifts and changes in atmospheric dynamics[54].

Each of these similarity metrics provides unique insights and complements the others in analyzing atmospheric data. Euclidean distance emphasizes absolute differences, Cosine similarity captures directional trends, and SSIM focuses on structural patterns in data. Together, they offer a comprehensive analysis of data similarity in atmospheric studies. The combination of these approaches forms a strong basis for understanding and compar-

ing different weather states over time and space [36].



The culmination of these similarity measures is presented in the integrated similarity formula. This formula is a summative approach that weighs each similarity metric based on variable weights, temporal dynamics, and spatial considerations to quantify the similarity between two atmospheric events comprehensively. By integrating these components, the formula allows for a nuanced comparison that is sensitive to the structure, distribution, and temporal patterns in atmospheric data, ensuring accurate similarity assessments for weather prediction and analysis.

The formula for calculating this comprehensive similarity score is as follows:

$$S_{w,\gamma_p,s}(\phi, \phi') = \sum_{j=1}^v w_j \sum_{k=0}^{t_w-1} \gamma_p(k+l) s(\phi_j^{k+l}, \phi_j'^{k+l}) \quad (3.10)$$

- $S_{w,\gamma_p,s}(\phi, \phi')$: The overall similarity measure between two atmospheric events.
- $\sum_{j=1}^v w_j$: The sum across variables j , where w_j represents the variable weights.
- $\sum_{k=0}^{t_w-1} \gamma_p(k+l)$: A summation across the time window, weighted by γ_p , which is the temporal weight function.
- $s(\phi_j^{k+l}, \phi_j'^{k+l})$: The similarity score for the weighted grid points between the two events at time k .



3.3 Machine Learning Approaches for Similarity Estimation

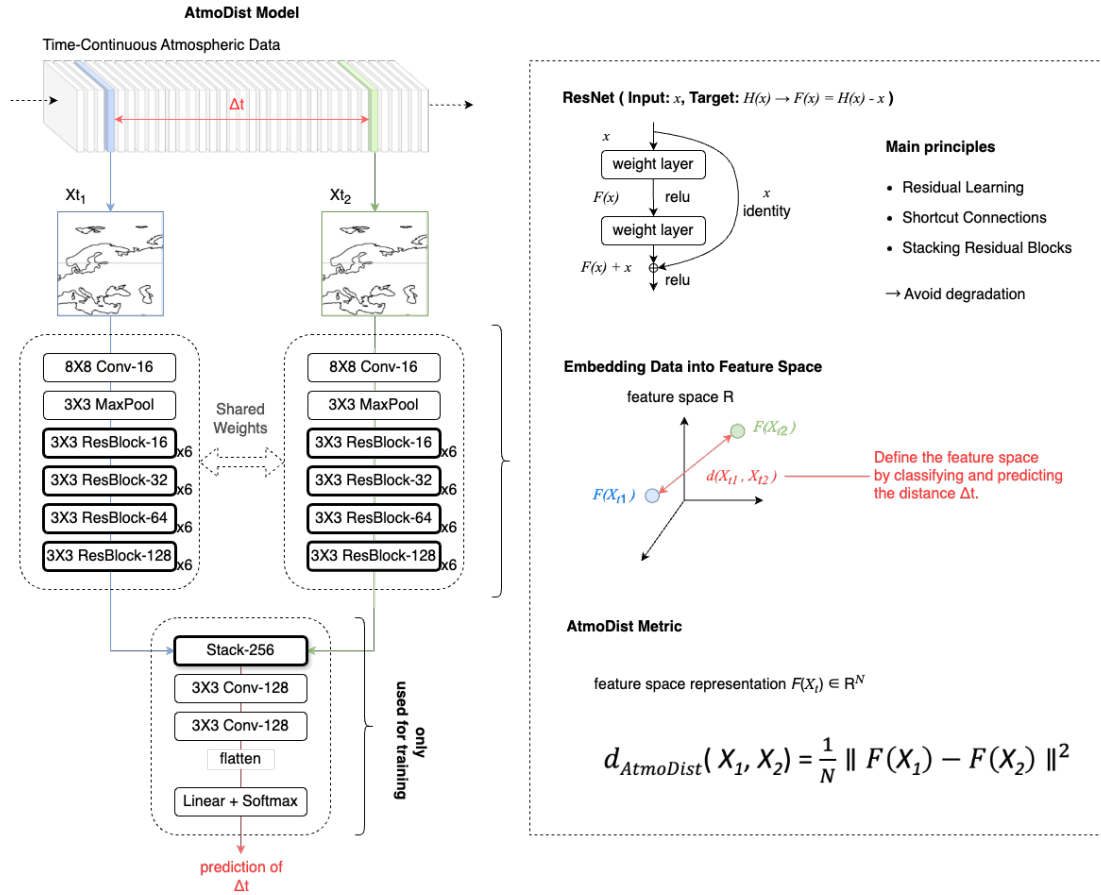


Figure 3.6 Atmodist Model Structure for Atmospheric Event Similarity Assessment

The Atmodist model consists of two main components: an upstream embedding of atmospheric data into feature space, and a downstream prediction of time difference categories between data pairs. Similarity assessment primarily relies on the upstream section, which utilizes a multi-layer ResNet structure to transform data into feature space. After embedding, the Atmodist index calculates similarity, providing precise judgments of atmospheric similarity. Source: [19]

In this study, I explore the capabilities of the Atmodist model (see Figure 3.6) in predicting localized extreme precipitation events, aiming to assess its potential for high-precision meteorological forecasting. The model evaluates the distance between atmo-

spheric events using a distance function defined as



$$D(\phi, \phi') = \sum_{k=0}^{t_w-1} d_{\text{AtmoDist}}(\phi_k, \phi'_k), \quad (3.11)$$

- t_w : Length of the time window.
- d_{AtmoDist} : Function that calculates the atmospheric state distance.
- ϕ_k : Atmospheric state of the first event at time step k .
- ϕ'_k : Atmospheric state of the second event at time step k .

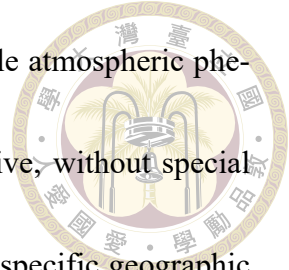
To effectively apply this model to our case, I have identified four key areas of focus:

1. Patch Size of Atmospheric States

In this research, I use a smaller patch size to better capture local atmospheric features. Smaller patches are more suited to represent localized weather variations, especially when forecasting extreme precipitation events, as they can more precisely reflect anomalies over smaller areas. In contrast, the original study employed a larger patch size, which might be more effective for detecting large-scale weather patterns but have limitations in fine-scale, localized forecasting.

2. Geographic Scope of Focus

The original research primarily focused on global or large-scale atmospheric phenomena, with data randomly sampled from a global perspective, without special attention to local variations. However, this study centers on a specific geographic region to enhance the model's accuracy in forecasting extreme events within that area.



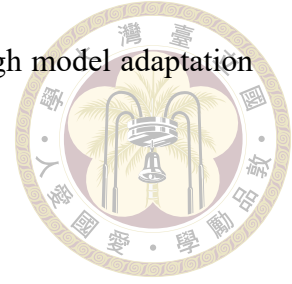
3. Variable Distribution Differences

In terms of variables, the data distribution in the original research included many extreme values, showing high skewness, and used logarithmic transformation for data preprocessing. In contrast, this study involves different variables with distinct distributions, requiring further analysis and potentially different preprocessing techniques to account for these differences.

4. Temporal Scale Differences

This study focuses on hourly-scale extreme event predictions, aiming to accurately forecast localized precipitation events over short time periods. The original study, however, likely emphasized longer timescale trend analysis, where the precision for short-term events was less critical. Here, the Atmodist model needs to be responsive to hourly-level changes to better capture the occurrence of extreme precipitation events.

In the following sections, I will address these challenges through model adaptation and data processing.



3.3.1 Reconstructed Model Structure

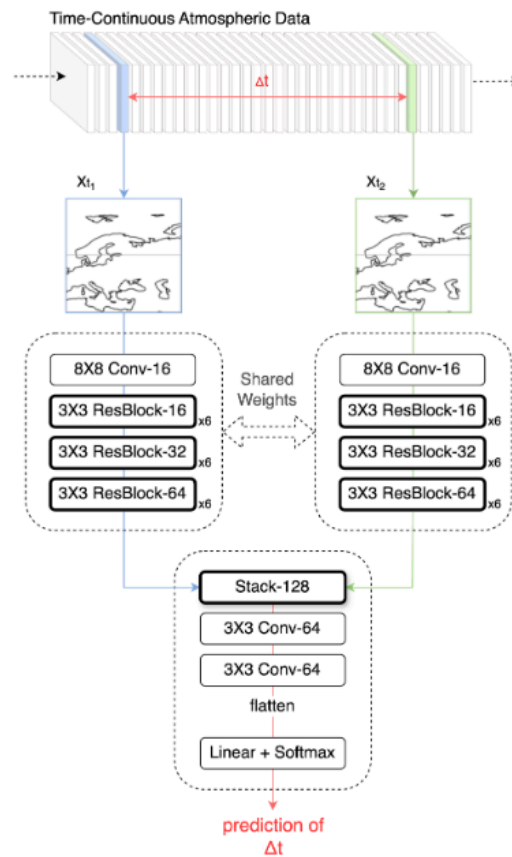


Figure 3.7 Revised Model Structure for Enhanced Small-Scale Patch Analysis

The original model was designed with a relatively large patch size (160x160) and included six layers of spatial downscaling to incrementally extract and compress broad-scale spatial features. However, when applied to the smaller patch size (32x32) required for this study, this design causes spatial features to converge prematurely to a 1x1 scale

(see Table 3.2). Such aggressive downscaling diminishes the model's ability to resolve fine details and results in a substantial loss of valuable spatial information, limiting its capacity to capture subtle, localized feature variations.



To enhance the model's adaptability in a small-scale setting, I systematically optimized the number and configuration of downscaling layers (see Appendix B.1). Specifically, I removed one MaxPooling layer and one ResBlock, allowing the model to retain more spatial information during feature extraction. This adjustment preserves richer detail within the 32x32 resolution, enabling the model to capture critical spatial changes over a smaller field of view and improve predictive accuracy.

The revised model structure, shown in Figure 3.7, retains spatial features while avoiding the resolution loss associated with excessive downscaling. This modification not only increases the model's flexibility for smaller patches but also enhances its accuracy and stability in capturing fine-scale local atmospheric features, demonstrating superior performance in this study's case applications.

Table 3.2 Layer-wise Comparison of Original Atmodist Model and Modified AtmoDist Model with Corresponding Shapes

Where C is the number of channels, H and W are the height and width of the data, respectively. The number of units in the final layer (e.g., 15) corresponds to the number of classification targets.

	Layer	Original Model	Modified Model
		Shape (C, H, W)	Shape (C, H, W)
Upstream	Input	(5, 32, 32)	(5, 32, 32)
	8×8 Conv	(16, 16, 16)	(16, 16, 16)
	3×3 MaxPool	(16, 8, 8)	-
	3×3 ResBlock - 16 ×6	(16, 4, 4)	(16, 8, 8)
	3×3 ResBlock - 32 ×6	(32, 2, 2)	(32, 4, 4)
	3×3 ResBlock - 64 ×6	(64, 1, 1)	(64, 2, 2)
	3×3 ResBlock - 128×6	(128, 1, 1)	-
Downstream	Stack - 256/128	(256, 1, 1)	(128, 2, 2)
	3×3 Conv - 128/64 ×2	(128, 1, 1)	(64, 1, 1)
	Flatten	(1, 128)	(1, 64)
	Linear + Softmax	(1, 15)	(1, 15)

3.3.2 Data Processing and Feature Analysis

Spatial Weighting To increase the model's sensitivity to local areas, I applied the spatial weighting method described in Section 3.2 and analyzed the results by comparing weighted and non-weighted outcomes. This approach aims to improve the model's responsiveness to meteorological changes within specific regions by focusing on localized

features. The spatial weighting adjustment is expected to help the model more accurately capture meteorological variations at particular geographic locations. Additionally, the non-weighted results are preserved as a baseline, allowing a comprehensive evaluation of the impact of spatial weighting on model performance.

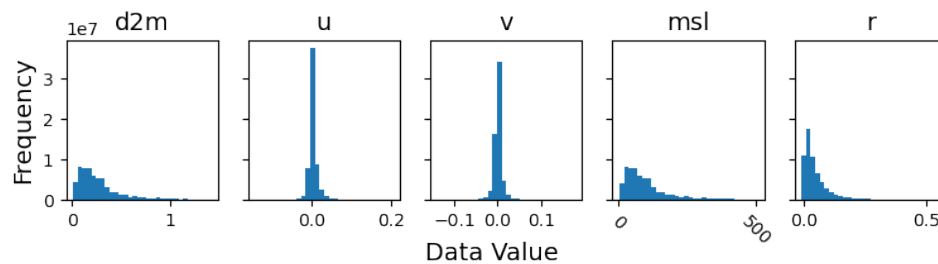


Figure 3.8 Distribution Characteristics of the Five Variables After Spatial Weighting

Preprocessing Methods To ensure variable comparability on a common scale and to enhance model stability, I employed targeted preprocessing techniques. For non-weighted variables, which often exhibit peaked distributions (see Figure 2.2), standardization was applied. This stabilizes the model's ability to capture inter-variable relationships and improves overall predictive accuracy. For spatially weighted data, given their more extreme distribution characteristics (see Figure 3.8), I utilized a logarithmic transformation [19], effectively mitigating the influence of outliers on model learning and enhancing the model's resolution in local feature analysis. Training results, detailed in the appendix B.2, further confirm the adaptability of these transformations to different variable types. In conclusion, selecting preprocessing methods suited to variable characteristics effectively enhances

overall model performance.

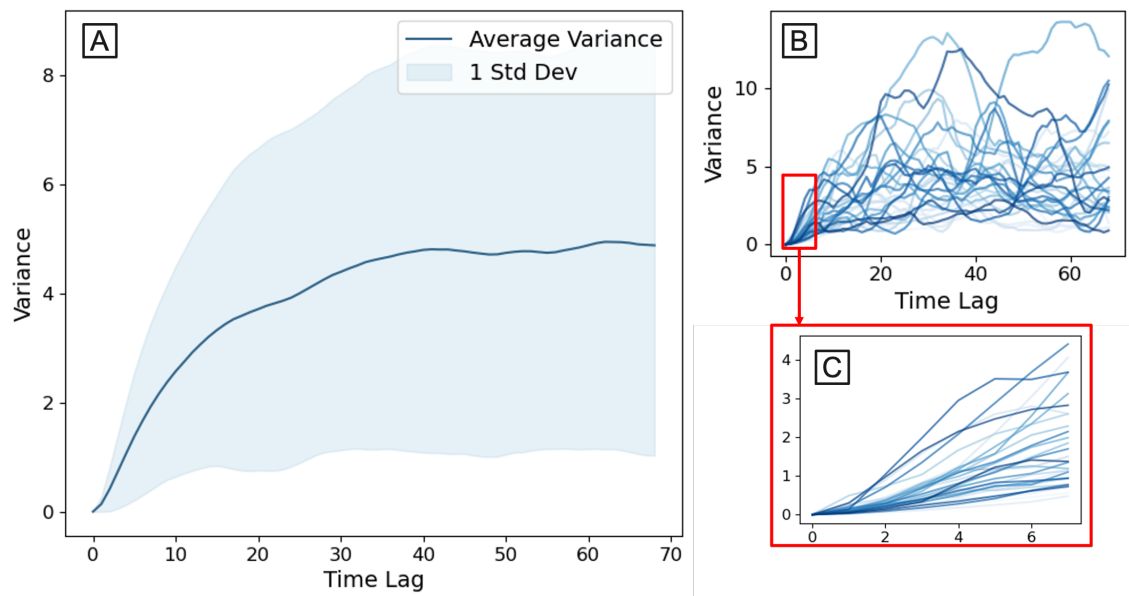


Figure 3.9 Variance Analysis of Temporal Lags using Dew Point Temperature ($d2m$)

(A) Shows the distribution of variance by randomly sampling 1000 times, calculating the variance between the data at the first time point and subsequent data points at different time lags (0-69 hours), with the average variance and one standard deviation shaded. (B) Displays 30 of these samples, with each line representing an individual variance curve. (C) Focuses on the first 7 hours of time lag for the same 30 samples, illustrating the trend of variance over shorter time intervals.

Temporal Interval and Scale Settings The Atmodist model segments the selected time

interval into distinct time lag categories based on appropriate temporal units, enabling the

model to differentiate between various atmospheric states effectively. As illustrated in

Figure 3.9, the variance of atmospheric variables (e.g., $d2m$) across increasing lag times

shows a gradual decrease in stability. Beyond 40 hours, variance values approach diver-

gence, making it increasingly difficult for the model to distinguish between atmospheric

states with longer lags.

Based on these observations and prior findings in the Atmodist framework [19], we set the temporal interval to **45 hours**, divided into **15 lag categories** at a **3-hour resolution**. This configuration balances the need for sufficient temporal detail with manageable computational complexity.

Through additional testing (see Appendix B.3), we found that higher temporal resolution improves prediction accuracy for the same number of predictive categories. However, higher resolution also increases computational demands. By adopting a 3-hour resolution, we optimized model performance while avoiding excessive training overhead.

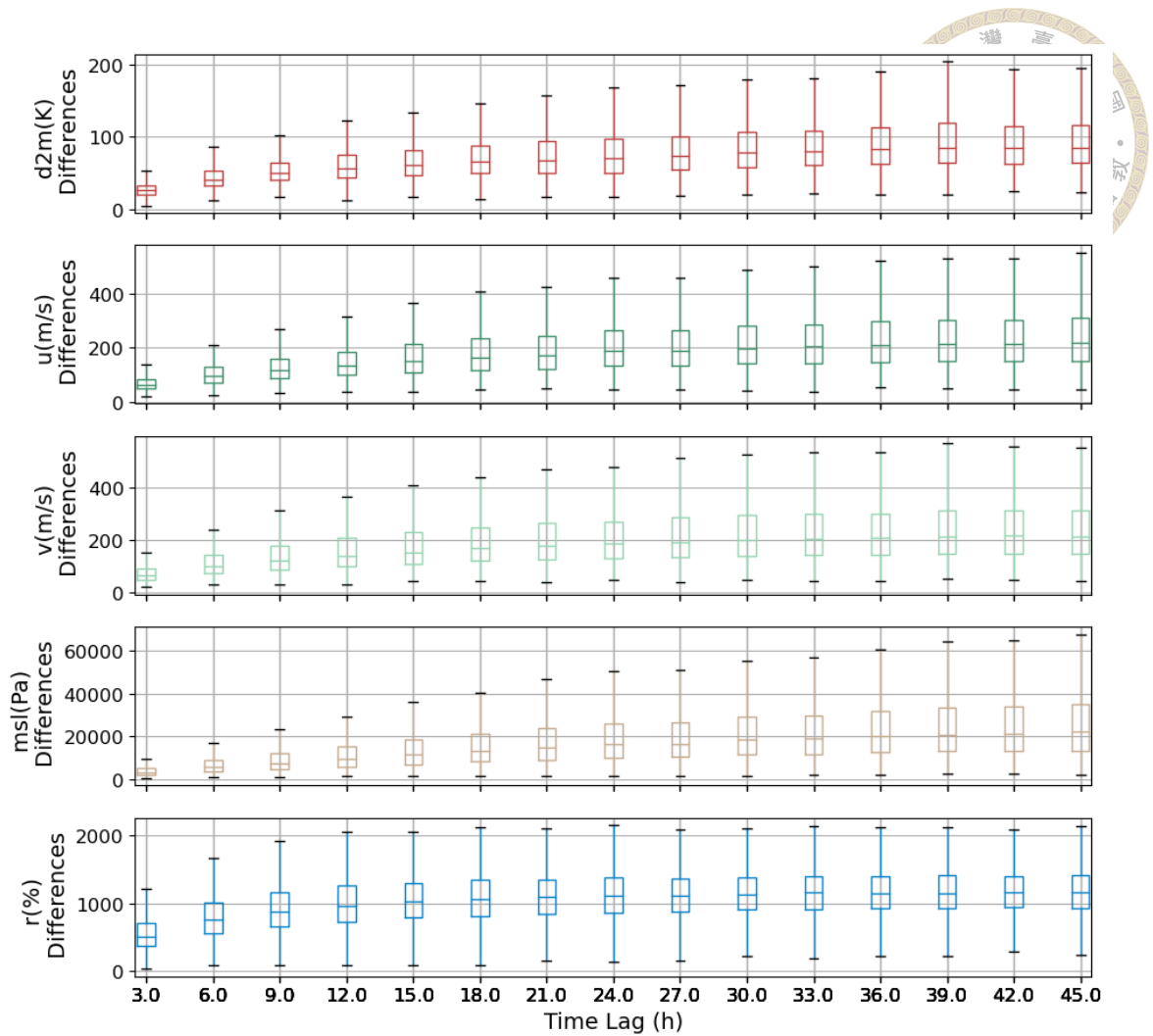


Figure 3.10 L2 Norm Difference of Atmospheric Variables Over Time Lag

Figure 3.10 shows the final distribution after sampling. The L2 norms of the five variables exhibit distinct trends as the time lag increases, reflecting varying sensitivities to lag time. Beyond a 40-hour lag, most variables struggle to maintain a consistently increasing trend across different unit scales, with some losing a clear pattern even at earlier stages. This behavior suggests that the model's ability to distinguish atmospheric states weakens over longer lags, potentially reducing its predictive accuracy. In particular, variables that lose an increasing trend early may fail to capture the cumulative effects of time progression, leading to less stable performance in extended forecasts.



3.3.3 Model Performance

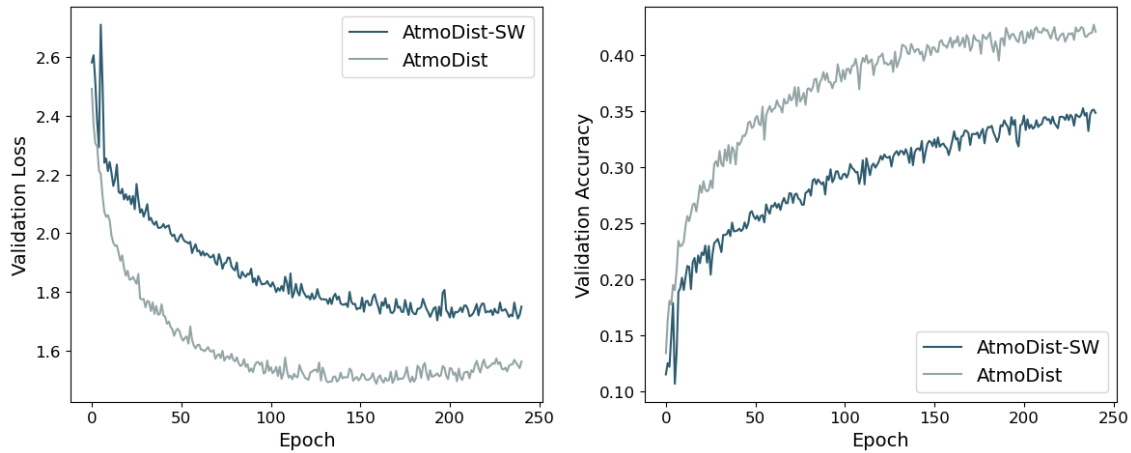


Figure 3.11 Validation Performance of Atmodist Model Using Weighted and Unweighted Data

Figure 3.11 compares the training performance of the Atmodist model with and without spatial weighting for time difference prediction tasks. The spatially weighted variant is referred to as AtmoDist-SW, where **SW** stands for **spatial weighting**. This approach applies weights based on spatial relevance, emphasizing key regions to enhance the model's sensitivity to localized patterns.

The objective of this task is to enable the model to learn data features and accurately classify time differences between actual and predicted states.

The left panel shows the trend of accuracy over epochs. The model trained with unweighted data consistently outperforms the one trained with weighted data throughout the training process. This indicates that unweighted data preserves the original feature struc-

ture more effectively, allowing the model to learn time difference classification patterns faster and more accurately. This effect is particularly evident in scenarios with short time differences, where the model converges more quickly and achieves a steady improvement in accuracy.

The right panel illustrates the loss over epochs, providing further insights into the impact of weighted and unweighted data on model training. The model trained with unweighted data experiences a faster decline in loss and stabilizes at a lower value in later epochs, demonstrating superior convergence. In contrast, the model trained with weighted data exhibits a slower loss reduction and maintains a higher loss level in the later training stages, suggesting that the characteristics of the weighted data increase the difficulty of learning.

Overall, these results indicate that unweighted data is more effective in helping the model learn time difference features, especially for short time difference prediction tasks. While weighted data may emphasize specific spatial features, it also introduces additional challenges, particularly during early training stages and in scenarios involving long time differences. This highlights the need to carefully balance feature emphasis and learning difficulty when designing data weighting strategies to better suit the requirements of specific prediction tasks.

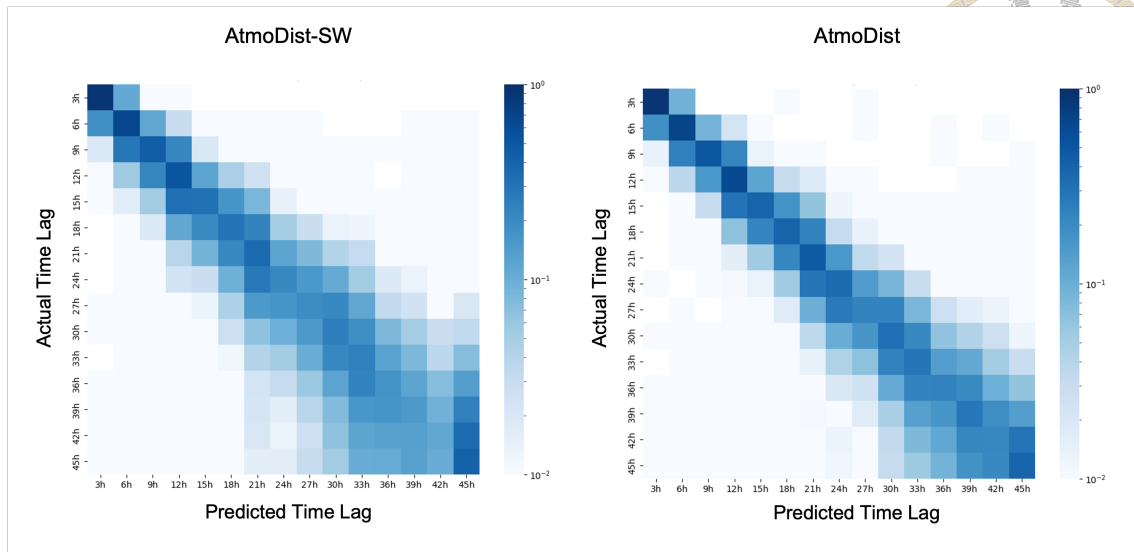


Figure 3.12 Confusion Matrices of Unweighted and Weighted Models for Time Difference Prediction

To comprehensively evaluate the performance of the model during training, we framed its task as a standard time classification problem. Specifically, the actual and predicted time differences were divided into 15 categories, spanning a range from 3 hours to 45 hours. A multi-class confusion matrix (Figure 3.12) was employed to visually represent the model's prediction accuracy across these time difference categories. This matrix effectively highlights the model's performance under varying time differences, allowing us to observe how its prediction capability evolves as the time difference increases.

The confusion matrix reveals a noticeable decline in prediction accuracy as the time difference increases. This decline is particularly pronounced under longer time differences when weighting is applied. Such degradation in accuracy could be attributed to the excessive compression of spatial information during the weighting process, which hinders the model's ability to capture localized feature variations, thereby affecting prediction per-

formance. This indicates that spatial weighting requires higher data resolution to maintain accuracy and enable the model to fully leverage spatial information for prediction.



Despite the limited overall accuracy, the concentration of predictions remains satisfactory. Since the primary objective of the model is to map data into a feature space to assess atmospheric state similarity, achieving perfect accuracy for every time difference is less critical. Instead, the focus lies on ensuring stability within the feature space and capturing the overall trends.

In summary, the model demonstrates higher accuracy within shorter time differences and a gradual, expected decline as the time difference increases. This suggests that the model is proficient in handling short-term variations, while its ability to manage longer time differences, though challenging, remains reasonably stable. These results further validate the model's capability to capture atmospheric state similarity within the feature space, serving as a potential basis for future applications.

3.4 Framework and Metrics for Assessing Method Performance

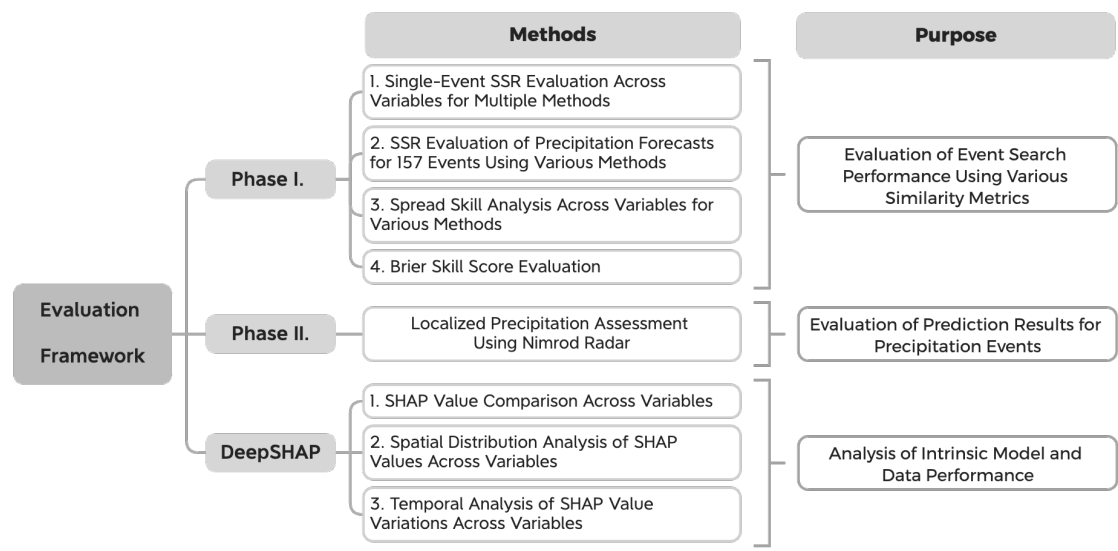
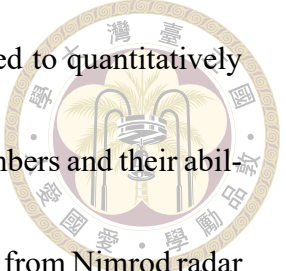


Figure 3.13 Evaluation Framework of Similarity Metrics and Precipitation Forecasts

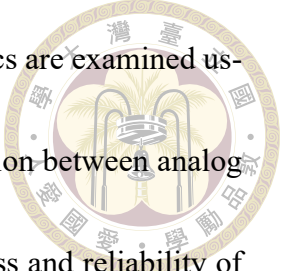
The evaluation framework of this study, as shown in 3.13, is divided into three stages to assess the similarity calculation methods, precipitation prediction results, and the intrinsic performance of the model and data. In particular, this study evaluates five similarity calculation methods: AtmoDist, AtmoDist-SW, Euclidean, Cosine, and SSIM, providing a comprehensive comparison to determine their effectiveness in identifying atmospheric state similarities. In the first phase, the event search capabilities of different similarity calculation methods are analyzed mainly through the SSR evaluation of single events and multiple events. It also includes the Spread Skill evaluation of different methods on variables to gain an in-depth understanding of the differences and performance between meth-



ods. At the end of this phase, the Brier Skill Score (BSS) is employed to quantitatively assess the effectiveness of different methods in selecting analogue members and their ability to predict precipitation. In the second phase, local precipitation data from Nimrod radar and high-resolution precipitation data from regional stations are used to further verify the prediction results of precipitation events. Subsequently, the DeepSHAP (Deep Learning SHapley Additive exPlanations) method is applied to analyze the model's performance from an internal perspective. By comparing weighted and unweighted data, evaluating the spatial distribution of variable SHAP values, and analyzing temporal changes, the internal characteristics of the model and data are comprehensively examined.

3.4.1 Phase I: Evaluation Methods for the Capability of Different Similarity Metrics in Precipitation Event Search

The evaluation process begins with the first step, which focuses on identifying the appropriate number of analog members and examining their general characteristics using SSR as the evaluation metric. This step provides an initial understanding of the relationship between the observed data and the analog members. The second step directly evaluates the predictive capability of different similarity metrics by assessing the precipitation prediction results across 157 events. Unlike the first step, this stage aggregates the SSR results over time and space to provide a comprehensive measure of prediction accuracy.



Finally, the stability and consistency of different similarity metrics are examined using binned spread-skill analysis. This approach evaluates the correlation between analog spread and prediction skill, offering deeper insights into the robustness and reliability of each similarity metric. To further quantify the performance of different similarity metrics in both analog member selection and precipitation prediction, the Brier Skill Score (BSS) is employed as a final evaluation measure. By comparing the BS values of different methods, BSS provides a relative assessment of their effectiveness, ensuring a standardized metric for evaluating forecasting skill. Together, these analyses form a systematic framework for assessing the effectiveness of similarity metrics in analog-based weather forecasting.

3.4.1.1 Methodology for Evaluating Similarity Metrics Across Variables in a Single Event

A precipitation event in the Birmingham area on August 7, 2011, at 1 PM was randomly selected as the target event. Data from the 6-hour period preceding the event was used for analysis. The Sum of Squared Residuals (SSR) is employed to evaluate the similarity between the observed data and the averaged values derived from analog members. This metric treats each spatial grid point independently, allowing for a detailed temporal and spatial analysis of prediction accuracy. Let $\mathbf{X}_0(t, c, h, w)$ represent the observed data

at time t , channel c , and grid location (h, w) , where $t \in [1, T]$, $c \in [1, C]$, and (h, w) span the spatial dimensions (H, W) . Similarly, $\mathbf{X}_i(t, c, h, w)$ denotes the data from the i th analog member, with $i \in [1, n]$ and n as the total number of analog members.

For each time step t , channel c , and grid point (h, w) , the average value across n analog members is computed as:

$$\mathbf{M}(t, c, h, w) = \frac{1}{n} \sum_{i=1}^n \mathbf{X}_i(t, c, h, w) \quad (3.12)$$

where $\mathbf{M}(t, c, h, w)$ represents the mean prediction at a specific location and time based on the analog members. Using these averaged values, the SSR for each grid point is calculated by summing the squared residuals between the observed data and the mean of the analog members across all time steps:

$$\text{SSR}(c, h, w) = \sum_{t=1}^T (\mathbf{X}_0(t, c, h, w) - \mathbf{M}(t, c, h, w))^2 \quad (3.13)$$

This formula quantifies the cumulative squared error for each grid point (h, w) in channel c , providing a spatially resolved metric of prediction accuracy.

The final output of the SSR calculation is a spatial grid for each channel with dimensions (H, W) , where each value reflects the total residuals at a specific location. This method offers a robust framework for assessing analog performance, as it dynamically ad-

just to variations in the number of analog members and provides localized insights into prediction quality. By independently evaluating each grid point, this approach highlights areas with significant discrepancies, thus supporting a comprehensive spatial and temporal analysis of prediction errors.

- T : The total number of time steps considered in the analysis.
- C : The number of channels (variables) in the dataset.
- H, W : The height and width of the spatial grid, respectively.

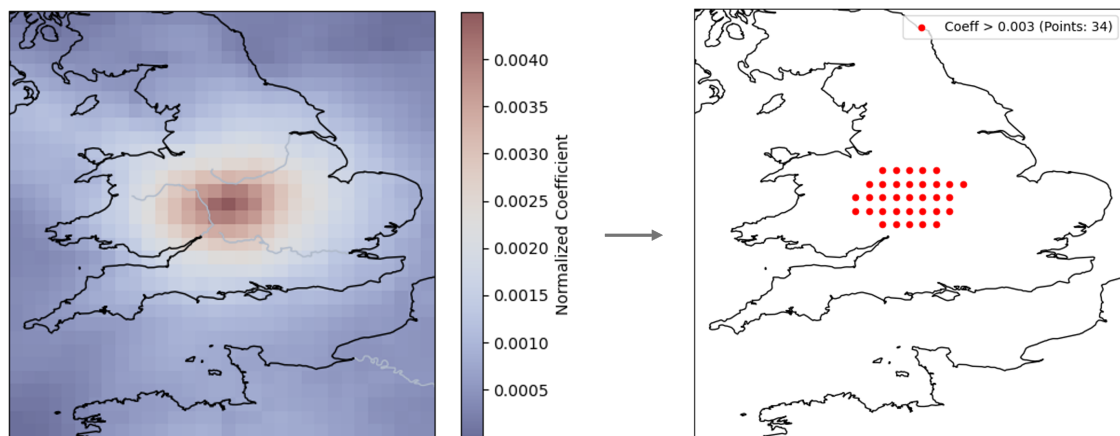


Figure 3.14 Selected Local Data Points Based on Correlation Coefficients

Selected 34 data points (about 3%) with a normalized correlation coefficient above 0.003. The left panel shows the spatial distribution of the normalized correlation coefficient across the region, and the right panel highlights the selected points with coefficients exceeding the threshold.

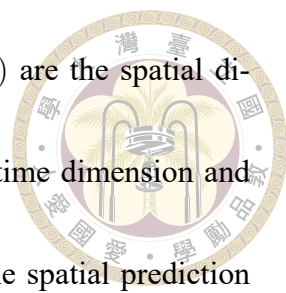
In addition to the global analysis, a localized analysis will be conducted to provide a more detailed understanding of the model's behavior in specific regions. The analysis will focus on the selected 34 points within this region, which represent about 3 percent of

the total data (Figure 3.14). These points were chosen based on their normalized correlation coefficient exceeding 0.003. By examining the results within this subset, we aim to evaluate the localized response to different similarity calculation methods. This approach highlights the attention mechanisms of each method at a finer spatial scale, offering insights into their performance and interpretability in capturing local atmospheric patterns.

3.4.1.2 Evaluation Methodology for Similarity Search Results Across 157 Precipitation Events

After determining the optimal number of analog members from the first step, the second step evaluates precipitation prediction using this fixed number of analog members. Unlike the first step, which focuses on individual grid points and channels, this step calculates the Sum of Squared Residuals (SSR) across all spatial dimensions and averages it over the temporal dimension for 157 precipitation events. Let $\mathbf{X}_0^{(e)}(t, h, w)$ represent the observed precipitation data for event e at time t and grid location (h, w) , and let $\mathbf{M}^{(e)}(t, h, w)$ denote the mean prediction from the selected number of analog members for the same event. The average SSR for event e is defined as:

$$\text{SSR}^{(e)} = \frac{1}{T} \sum_{t=1}^T \sum_{h=1}^H \sum_{w=1}^W \left(\mathbf{X}_0^{(e)}(t, h, w) - \mathbf{M}^{(e)}(t, h, w) \right)^2 \quad (3.14)$$



Here, T represents the total number of time steps, and (H, W) are the spatial dimensions of the grid. By averaging the squared residuals over the time dimension and summing over all grid points, the SSR for each event aggregates the spatial prediction errors into a single scalar value, normalized by the event duration. This approach enables direct comparisons of prediction accuracy across the 157 events, providing a robust evaluation of the analog members' performance in representing precipitation patterns across diverse spatial scales.

3.4.1.3 Computational Approach for Binned Spread Skill Assessment Using Different Similarity Metrics

To further analyze the predictive performance, **binned spread skill** is introduced by comparing the spread of analog members with the skill calculated using SSR. In this approach, **spread** is defined as the SSR between each analog member and the ensemble mean at each grid point, while the **skill** is derived from the SSR results discussed in the previous sections. This method highlights how different methods perform in capturing the relationship between spread and skill for various variables, offering insights into the consistency between the ensemble spread and the accuracy of the predictions.

For each variable c , its spread and skill are calculated as:

$$\text{Spread}_c = \sum_{i=1}^n (\mathbf{X}_i(t, c, h, w) - \mathbf{M}(t, c, h, w))^2 \quad (3.15)$$

$$\text{Skill}_c = \sum_{i=1}^n (\mathbf{X}_i(t, c, h, w) - \mathbf{X}_0(t, c, h, w))^2 \quad (3.16)$$

where $\mathbf{X}_i(t, c, h, w)$ represents the value of the i -th member, n is the total number of members, and $\mathbf{M}(t, c, h, w)$ is the mean value across all members.

This formulation ensures that each variable c has its own independent measure of spread, calculated based on the squared deviations of all members $\mathbf{X}_i(t, c, h, w)$ from the mean value $\mathbf{M}(t, c, h, w)$. This allows for a detailed assessment of variability for each variable over time and space.

The relationship between spread and skill is assessed by binning the spread values into predefined intervals based on their numerical ranges, not time. The calculation of skill follows the same method described earlier, using the Sum of Squared Residuals (SSR) as the metric. Specifically, for a bin interval $[a, b)$, the binned spread skill is given as:

$$\text{Skill}_{[a,b)} = \frac{\sum_{(t,c,h,w) \in [a,b)} \text{SSR}(t, c, h, w)}{\sum_{(t,c,h,w) \in [a,b)} 1}$$



The summation is performed over all grid points, variables, and time steps whose spread values fall within the interval $[a, b)$.



This approach enables the evaluation of prediction skill across different ranges of spread, providing insights into the relationship between the spread of analog members and their predictive accuracy.

To further analyze this relationship, the coefficient of determination (R^2) between the binned spread and skill values is calculated. The R^2 value quantifies the proportion of variance in skill that can be explained by the spread, serving as a measure of the strength and direction of their relationship. The formula for R^2 is given by:

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

where y_i represents the observed skill values for each bin, $\hat{y}_i$ represents the predicted skill values based on a linear fit, $\bar{y}$ is the mean of the observed skill values, and n is the total number of bins. This calculation provides a quantitative measure to assess the relationship between spread and skill across the bins.

This methodology provides insights into how different methods perform in capturing the relationship between ensemble spread and forecast skill for various variables. High consistency between spread and skill indicates that the spread is an effective measure of

forecast uncertainty, supporting the reliability of the analog prediction approach.

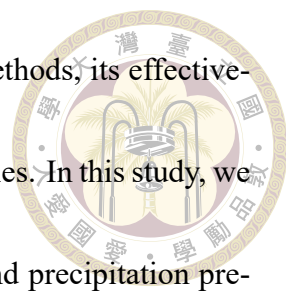


3.4.1.4 Brier Score and Brier Skill Score Computation

The Brier Score (BS) is a widely used metric for evaluating the accuracy of probabilistic forecasts, particularly in meteorology and climate science [56]. It quantifies the mean squared difference between forecast probabilities and observed binary outcomes, providing a measure of reliability and resolution in forecasting performance. A lower BS indicates better forecast accuracy. However, BS alone does not account for the baseline skill of a reference forecast, making it less suitable for comparative evaluations across different models or methods. To address this limitation, the Brier Skill Score (BSS) extends BS by comparing a given forecasting method to a reference model, often climatology or a simpler analogue method [57]:

$$BSS = 1 - \frac{BS_{\text{method}}}{BS_{\text{reference}}} \quad (3.17)$$

where BS_{method} represents the Brier Score of the evaluated forecasting method, and $BS_{\text{reference}}$ corresponds to the baseline reference. A positive BSS indicates that the evaluated method outperforms the reference, whereas a negative BSS suggests that the reference model provides better predictions.



While the BSS provides a relative comparison of forecasting methods, its effectiveness depends on how BS is computed for different atmospheric variables. In this study, we evaluate BS by considering both individual atmospheric variables and precipitation prediction across different thresholds. The computation of BS is performed independently for each time step and spatial location, ensuring that each result is derived separately before averaging over the entire dataset. Specifically, we evaluate the performance of AtmoDist-SW and Cosine similarity by computing BS for their respective analogue sets, allowing for a direct comparison of their effectiveness in capturing atmospheric similarities and predicting precipitation.

For evaluating the similarity of analogue set members to observed atmospheric variables, BS is computed separately for each variable. Given an observed field O_v for variable v and an ensemble of analogue members A_v , the probability of similarity is determined based on deviations within the observed variability:

$$P_v = \frac{1}{N} \sum_{i=1}^N \mathbb{I}(|A_{v,i} - O_v| \leq \delta_v), \quad (3.18)$$

where δ_v is the standard deviation of the observed field for variable v , used as a threshold to determine whether an analogue member is sufficiently close to the observed

state.



Since all observed values are considered relevant in this evaluation, the reference field O_v is assigned a constant value of 1:

$$O_v = 1 \quad (3.19)$$

This assumption ensures that the forecast probability P_v is evaluated against a reference where all observations are treated equally. The BS for each variable is then computed independently at each grid point in time and space, with the final BS obtained by averaging over all time steps and spatial locations:

$$BS_v = \frac{1}{TXY} \sum (P_v - O_v)^2 \quad (3.20)$$

For precipitation evaluation, BS is computed across different precipitation thresholds T , taking into account multiple independent events. The observed precipitation field is binarized based on whether it exceeds a given threshold:

$$O_T = \mathbb{I}(O \geq T) \quad (3.21)$$

The probability of exceeding the threshold is estimated from the analogue ensemble:



$$P_T = \frac{1}{N} \sum_{i=1}^N \mathbb{I}(A_i \geq T) \quad (3.22)$$

The BS for each precipitation threshold is computed independently at each grid point in time and space, considering multiple events. The final BS is obtained by averaging over all time steps, spatial locations, and events:

$$BS_T = \frac{1}{\sum_e T_e XY} \sum_e \sum (P_T - O_T)^2 \quad (3.23)$$

Finally, BSS is computed by comparing the BS values of AtmoDist-SW and Cosine similarity methods across different precipitation thresholds, with the results plotted to assess performance improvements.

3.4.2 Phase II: Evaluation Methods for Localized Precipitation Forecast Results

To evaluate the effectiveness of the method in generating localized precipitation forecasts, this study utilizes the Nimrod radar precipitation data. This dataset provides spatially distributed observations of precipitation across the target region, enabling an assess-

ment of the model' s ability to capture localized precipitation characteristics and reflect regional dynamics.



Nimrod radar precipitation data serves as a critical resource for refining the ensemble members identified during Phase I. The initial ensemble, consisting of 120 members derived from ERA5 reanalysis data, represents a diverse range of atmospheric conditions that are potentially analogous to the target event. To enhance the relevance of these ensemble members to localized precipitation features, the radar precipitation observations are employed as a higher-resolution reference dataset.

For each event, the precipitation fields of the ensemble members are accumulated over the final hour preceding the event. The **L2 norm** is then computed between these accumulated precipitation values and the radar observations, providing a quantitative measure of similarity. Based on these similarity metrics, the **12 ensemble members** with the smallest **L2 norm** distances are selected as the refined subset most representative of the observed conditions.

The selected members are subsequently utilized to generate localized precipitation forecasts. These forecasts include **deterministic predictions**, which provide specific precipitation estimates, and **probabilistic predictions**, which quantify the uncertainty associated with the forecast. Both types of forecasts are generated for a lead time of **3**

hours, enabling an in-depth assessment of the methods' capability to capture the spatial and temporal dynamics of localized precipitation events. This process establishes a robust framework for evaluating both the reliability and uncertainty of the forecasting approaches under study.



3.4.3 Methods for Explaining the Intrinsic Performance of Models Using DeepSHAP

In this study, we utilize four-dimensional weather data (time, variables, latitude, longitude) to predict changes in atmospheric states across different periods. To better interpret our machine learning model's performance, we introduce the DeepSHAP method. DeepSHAP is an extension of the SHAP (SHapley Additive exPlanations) framework, combining the DeepLIFT (Deep Learning Important FeaTures) algorithm with SHAP values. Specifically designed for deep learning models, DeepSHAP effectively computes SHAP values for complex models, thereby offering insights into model predictions.

3.4.3.1 Theoretical Foundation of DeepSHAP

DeepSHAP is an algorithm for interpreting predictions of deep learning models by combining the principles of DeepLIFT (Deep Learning Important FeaTures) and SHAP (Shapley Additive exPlanations) values. Below, we elaborate on both components and

their application in DeepSHAP.



Shapley Value The Shapley value is used to fairly distribute the “payout” (or contribution) among a set of features[58]. It ensures that each feature’s contribution to the overall model prediction is considered.

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!} [f(S \cup \{i\}) - f(S)] \quad (3.24)$$

- N is the total number of features.
- S is a subset of features.
- $f(S)$ is the model output when only the subset S of features is present.

This formula calculates the marginal contribution of feature i over all possible subsets S of features.

DeepLIFT Contribution Calculation DeepLIFT provides a method to compute feature attributions by comparing a reference input and the actual input [59].

$$C_{\Delta x \rightarrow \Delta y} = \frac{\Delta y}{\Delta x} \quad (3.25)$$



- Δx is the difference between the input and the reference value.
- Δy is the difference between the output and the reference value.

DeepLIFT computes how much each input contributes to the difference between the output and a “reference” output.

DeepSHAP is an algorithm that combines the concepts of Shapley values and DeepLIFT [60]. It leverages the efficiency of DeepLIFT to compute Shapley values for deep learning models by using backpropagation to efficiently distribute the importance of each feature across the network. DeepSHAP provides a principled way to explain model predictions by attributing contributions to individual features.

3.4.3.2 Processing Flow of DeepSHAP

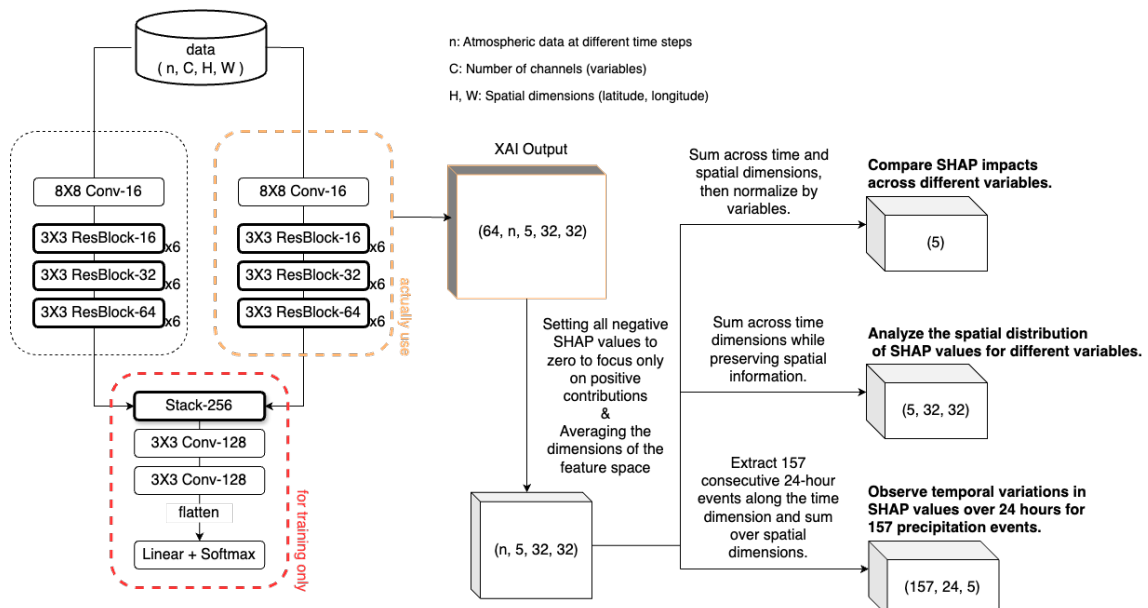


Figure 3.15 Processing Flowchart of DeepSHAP

In this analysis, we applied DeepSHAP interpretation to the Encoder layer of the model rather than the prediction layer [58]. The Encoder layer is responsible for mapping atmospheric states into feature space and estimating distances, which is critical for understanding relationships between different atmospheric states [61]. In contrast, the prediction layer functions solely as a training objective and does not directly contribute to distance calculations. Therefore, the DeepSHAP analysis was focused on the more relevant intermediate stage of the model.

After obtaining the DeepSHAP outputs, we observed that the distributions of positive and negative values were largely symmetrical. This indicates that features contributing positively to the formation of the feature space could also produce equivalent negative impacts. To simplify interpretation, we set negative values to zero and averaged across the feature dimensions. This approach provides a more intuitive representation of feature importance within the spatiotemporal domain.

At the variable level, the model input consists of five atmospheric variables, and the contribution of each variable to the prediction outcome was evaluated using SHAP values. By calculating SHAP values across different time steps and spatial locations, we assessed the relative influence of each variable. The SHAP values were averaged [62] and normalized to highlight their relative importance. This process facilitates the rapid

identification of key variables driving the prediction while revealing interactions and relationships among variables.

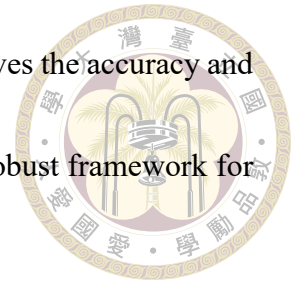


At the spatial level, each grid point in the input data is treated as an independent feature, enabling us to evaluate its contribution to the overall feature space [63]. This fine-grained approach allows us to examine how atmospheric parameters at specific geographic locations influence model predictions and explains the significance of each location to the forecast results. Such analysis is particularly effective in identifying regions that have a more pronounced impact on the overall atmospheric state.

At the temporal level, we analyzed 157 precipitation events to track SHAP value trends over 24 hours following each event. This time-series analysis helps distinguish the model's sensitivity to short-term variations and evaluates how it captures atmospheric dynamics at different temporal scales. By examining the evolution of SHAP values hourly, we gain insights into how the model integrates temporal features and accounts for cumulative effects during event development.

This multi-level approach reveals deeper relationships between features, time, and space, providing a more complete and nuanced interpretation of atmospheric states. Such interpretability is particularly useful for studying extreme weather events, as it helps uncover key patterns in atmospheric dynamics and enhances the model's ability to capture

features across different scales [64]. Furthermore, this process improves the accuracy and interpretability of atmospheric state distance estimation, offering a robust framework for analyzing and predicting complex meteorological patterns.







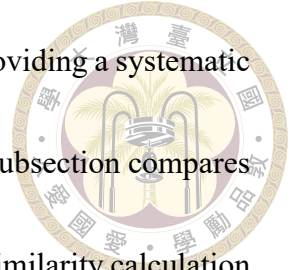
Chapter 4 Results and Discussion

The following results section is divided into three main chapters, which sequentially analyze weather event similarity calculation and prediction, the application of analog methods to local precipitation forecasting, and the interpretation of model prediction results. This systematic presentation highlights the analytical process and key findings, providing a comprehensive evaluation of the performance and predictive capabilities of weather analog methods.

4.1 Phase I: Comparison of Weather Event Similarity Calculation and Prediction

This subsection focuses on the calculation and prediction of weather event similarity, exploring the performance of analog methods from multiple perspectives. First, the analysis begins with a single-event evaluation using the Sum of Squared Residuals (SSR), which is combined with further analysis to determine the optimal number of ensemble

members. The evaluation then extends to 157 precipitation events, providing a systematic assessment of the overall predictive performance. Additionally, this subsection compares the Spread Skill of different methods, highlighting the differences in similarity calculation and prediction capabilities among the various analog approaches.



4.1.1 Sum of Squared Differences Analysis of Calculation Results for a Single Event

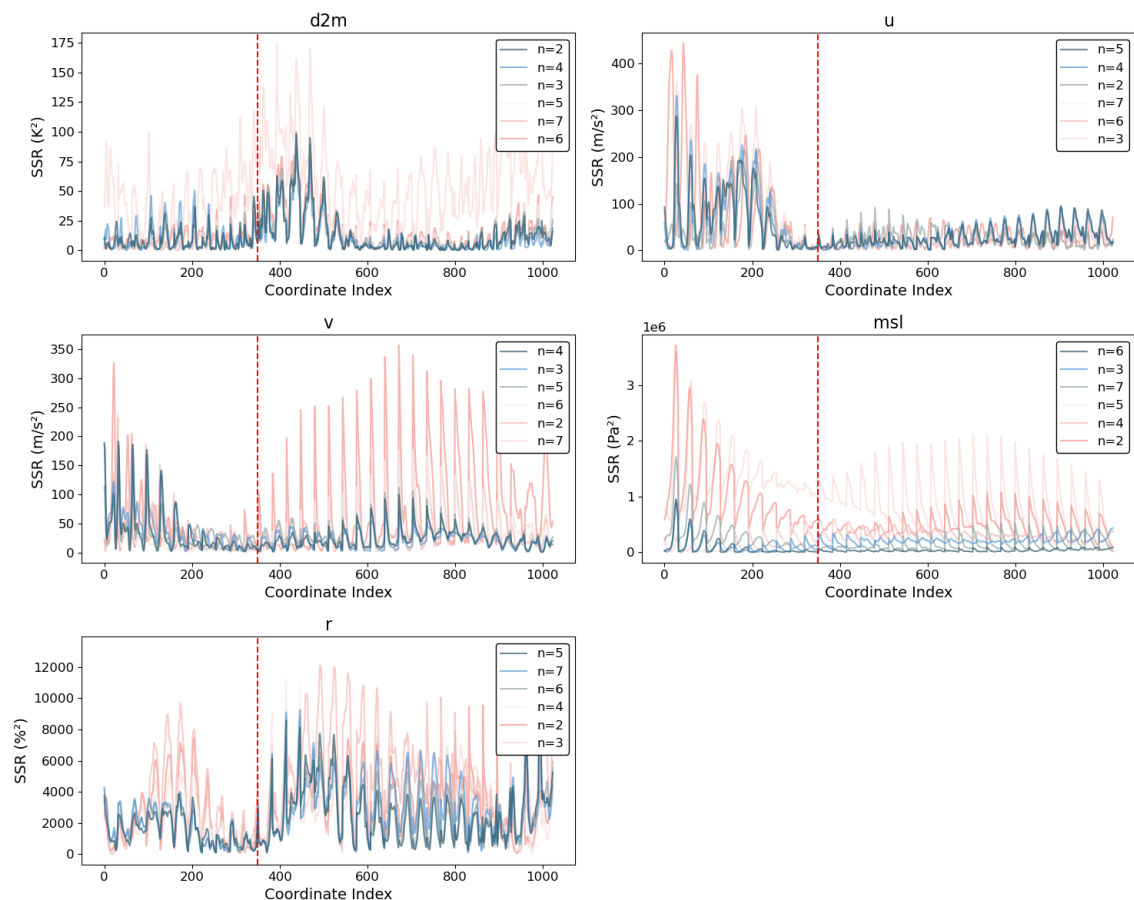


Figure 4.1 Sum of Squared Residuals Analysis for Five Atmospheric Variables across Different Analogues Based on AtmoDist Results

Figure 4.1 illustrates the performance of the AtmoDist-SW similarity method after identifying analog members under varying numbers of members and different variables. The analog members were ranked based on their performance, where better performance corresponds to lower Sum of Squared Residuals (SSR). The analysis reveals that each variable achieves its optimal performance at different numbers of analog members, indicating varying demands across variables regarding the number of analogs.

Furthermore, the performance of the five variables is not necessarily optimal in the target region, suggesting that the search for analog members still relies on the overall spatial structure and patterns. However, the spatial distribution characteristics of the variables exhibit significant differences, which may be related to the inherent properties of the variables, physical processes, or regional climate conditions.

Ultimately, by referencing the performance of different methods (Appendix A.1), six analog members were chosen as the standard for subsequent evaluations. This selection balances computational efficiency while ensuring the performance across methods and accounting for the spatial distribution and characteristics of the variables in the analog analysis.

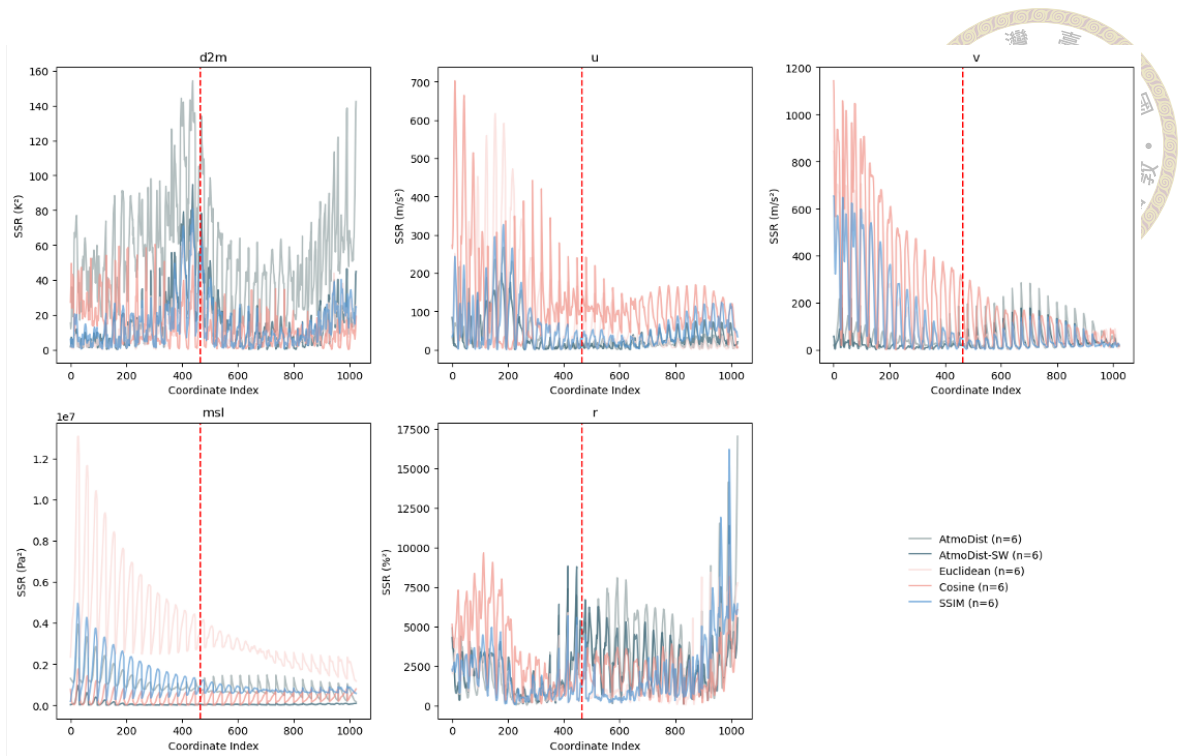


Figure 4.2 Sum of Squared Residuals for Five Variables across Five Methods

Figure 4.2 presents the results of analog member searches for different atmospheric variables using various methods. In this case study, the spatial distribution of the SSR for each variable exhibits a certain degree of regularity. The results across the five methods show similar trend distributions for multiple atmospheric variables. Although some fluctuations are observed in localized regions, the overall differences remain minimal. This indicates that the analog members identified by the five methods have a relatively consistent level of similarity, demonstrating the stability of these methods in calculating atmospheric state similarity.

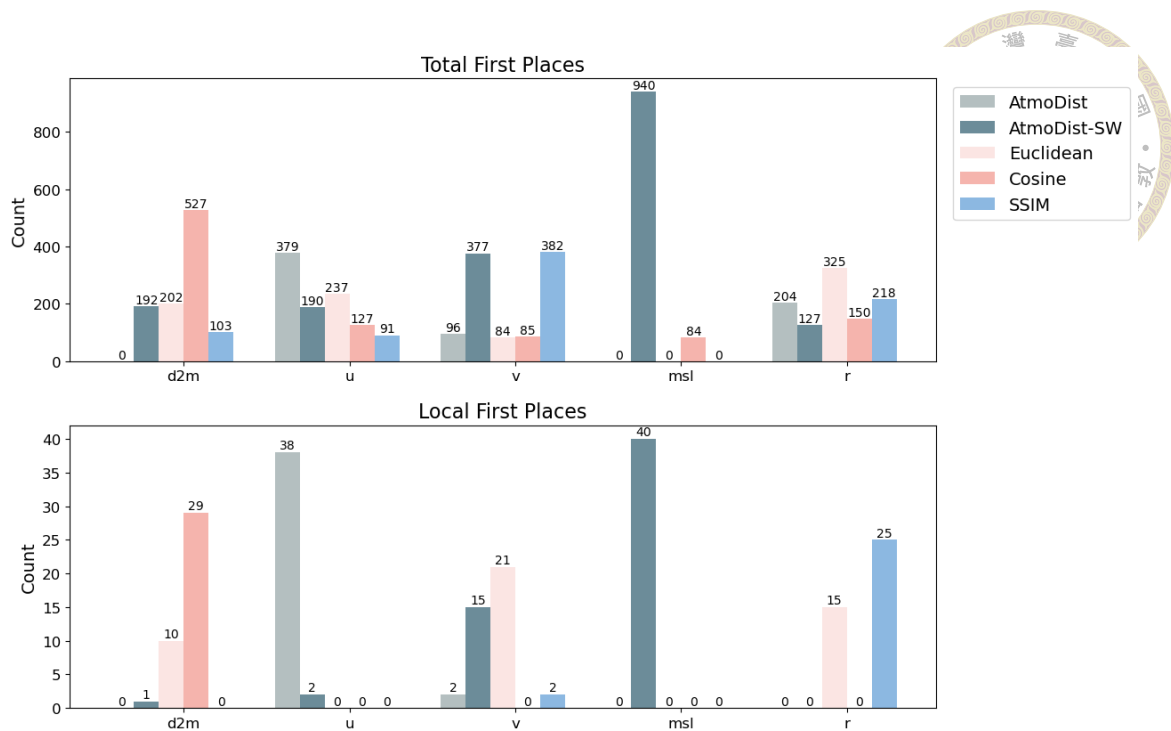


Figure 4.3 Comparison of First-Place Counts for Different Methods Across Atmospheric Variables

To facilitate a more intuitive comparison of method performance, we counted and visualized the number of instances where each method achieved the best performance, defined as the lowest SSR, across different data points (Figure 4.3). The results indicate that there is no universally superior method; each method demonstrates the best performance for specific variables, highlighting the differences in their ability to handle different meteorological variables.

Notably, in the local analysis, the non-weighted AtmoDist method outperforms others for the zonal wind variable (u). This result is particularly striking because traditional methods have specifically emphasized and assigned higher weights to relative humid-

ity (the second highest weighted variable among all). Yet, the non-weighted AtmoDist method still achieves superior performance, suggesting its potential advantage in capturing zonal wind features compared to other methods. This finding provides insight for further refinement and application of the method.

Overall, the weighted AtmoDist method and the Cosine method demonstrate relatively stable performance, frequently achieving the best results across various variables.

4.1.2 Sum of Squared Differences Analysis of Prediction Results for 157 Precipitation Events

To comprehensively evaluate the predictive capabilities of different methods for precipitation events, the following analysis focuses on 157 precipitation events. Compared to single-event assessments, a multi-event comparison provides a more accurate reflection of each method's performance across varying scenarios. It is important to note that the numerical values alone do not fully represent the superiority or inferiority of predictive abilities. Therefore, the analysis will focus on the overall stability of the predictions and the differences between methods, offering deeper insights into their practical applicability and potential.

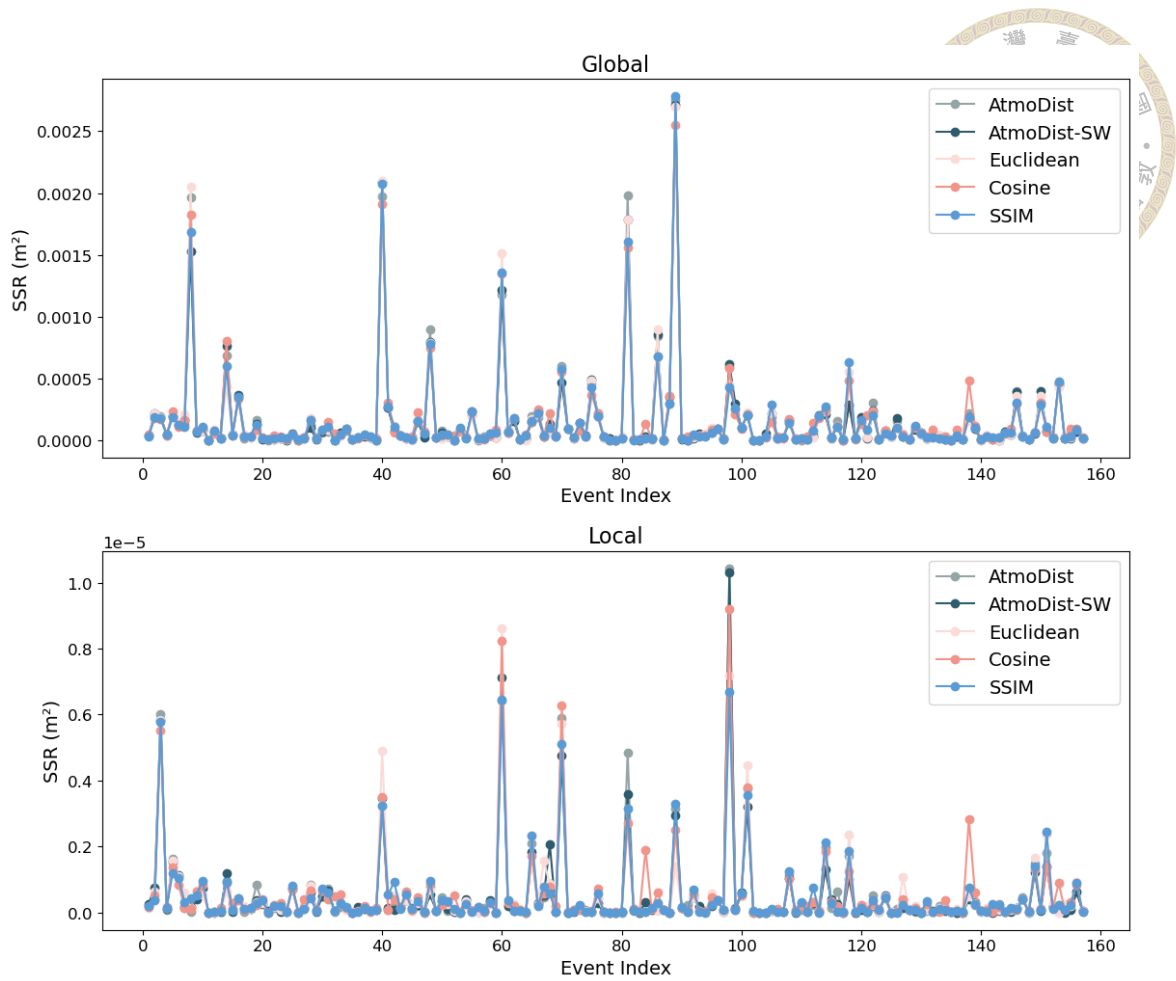


Figure 4.4 Comparison of Sum of Squared Residuals Distribution across Five Methods for 157 Precipitation Events

As observed in Figure 4.4, the SSR results for different precipitation events exhibit similar trends in both global and local analyses. This indicates a high degree of consistency in how the complexity or intensity of precipitation events is captured across different similarity calculation methods. In the global analysis, while minor fluctuations are observed for certain events, the primary peak positions and trends remain largely consistent, demonstrating the stability of all five methods in capturing the overall characteristics of precipitation events.

In the local analysis, the positions of SSR peaks align closely with those observed in the global analysis, suggesting good coordination between local features and global patterns. This consistency enhances the reliability of the results and highlights the stability and generalizability of these methods in the analysis of atmospheric precipitation events.

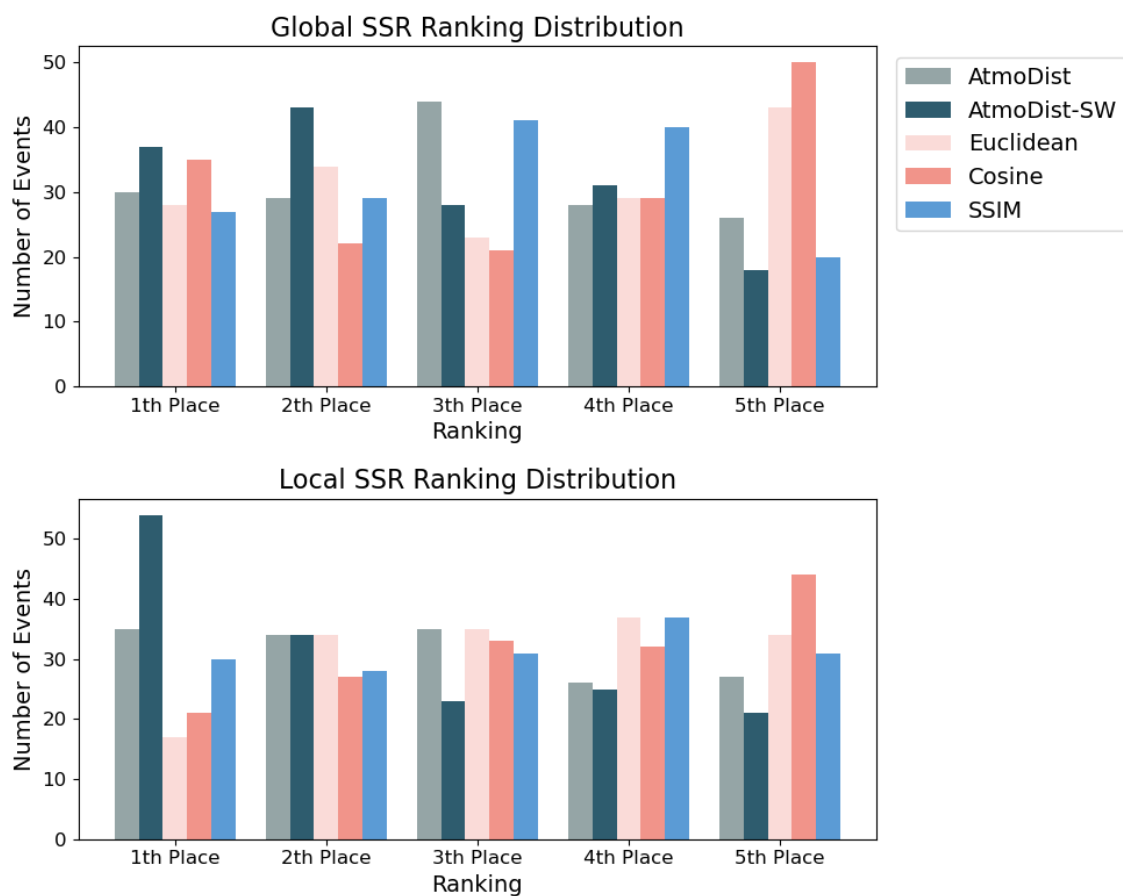


Figure 4.5 Global and Local SSR Ranking Distribution Across Methods

To provide a clearer evaluation of the predictive performance, we directly analyzed the ranking of methods based on their results for each event (Figure 4-5). From the ranking distribution, it can be observed that the weighted AtmoDist method achieves the highest number of first and second places in both global and local analyses, demonstrating its

overall advantage in precipitation forecasting.



Notably, even though the non-weighted AtmoDist method does not apply spatial weighting during data processing, its performance in the local analysis still surpasses that of traditional methods. This highlights the potential of its algorithmic structure in capturing local features effectively.

Among the traditional methods, SSIM shows the most consistent performance, as it rarely falls into the lowest rankings compared to other traditional approaches. This reflects its robustness in precipitation event predictions.

In summary, the comparison of SSR analyses and method ranking distributions verifies the superiority of AtmoDist-SW while also revealing the differences in performance among various methods in both global and local analyses. These findings provide valuable insights for further optimization and application of these methods.

4.1.3 Spread Skill Evaluation of Models and Methods

In this section, we examine the Spread Skill performance of different methods, focusing on their stability and accuracy in searching for analog members for precipitation prediction across various variables.

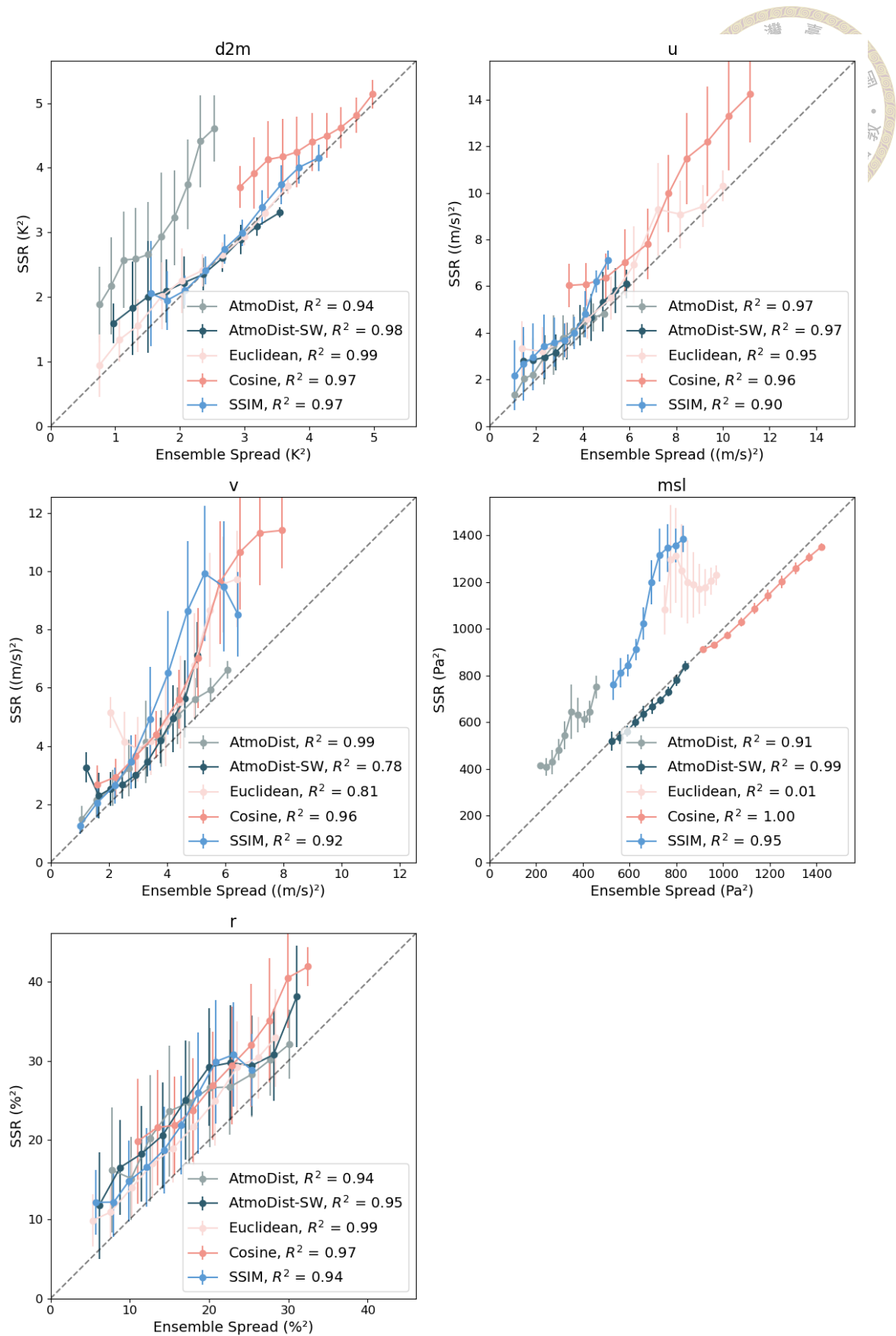


Figure 4.6 Binned Spread Skill Comparison for Five Variables Using Different Methods

Figure 4.6 presents the analysis results for a single event, evaluating the relationship between Ensemble Spread and SSR across five variables for different methods. Overall, the AtmoDist-SW method demonstrates superior performance, with its Spread and Skill distributions aligning more closely, indicating relatively stable results. Additionally, all variables show good correlations, except for the variable *msl*, which exhibits notably poorer performance.

For *msl*, only the AtmoDist-SW and Cosine methods show distributions that align well with the diagonal, while the Euclidean method performs poorly, with a correlation as low as 0.01, failing to effectively capture the characteristics of this variable.

In summary, this analysis highlights the overall stability and performance advantage of the AtmoDist-SW method for a single event. Furthermore, the Cosine method also demonstrates consistent performance across most variables, particularly in *msl*, where its Spread and Skill distributions are second only to the AtmoDist-SW method.

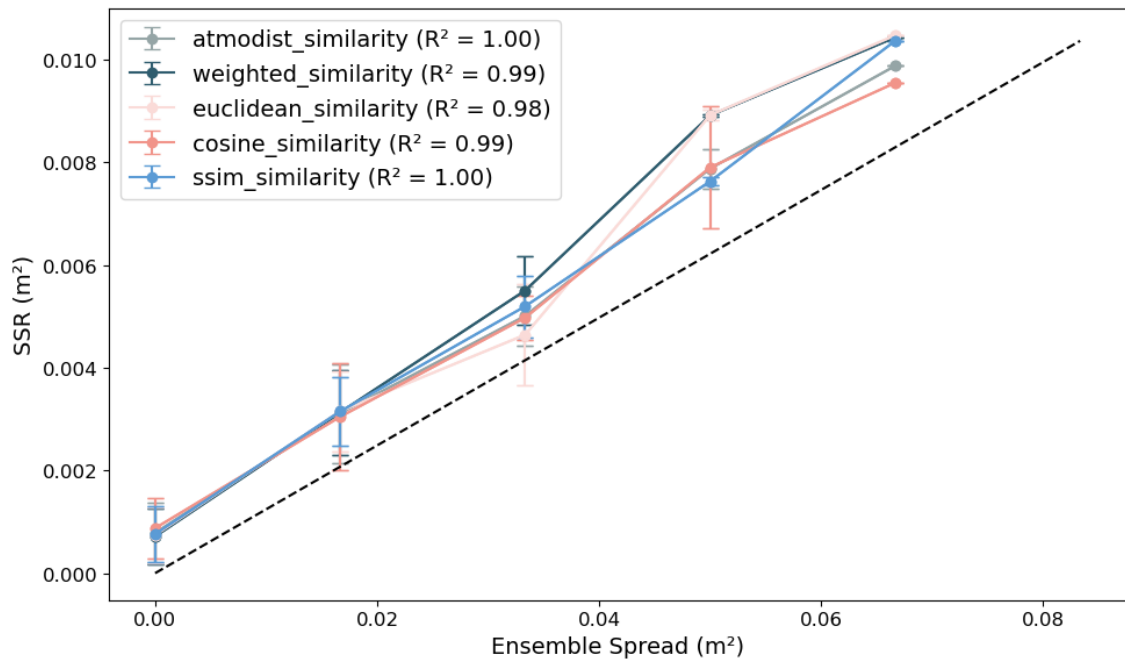


Figure 4.7 Binned Spread-Skill Relationship for Precipitation

Figure 4.7 presents the analog analysis results for all 157 precipitation events, focusing on the relationship between Ensemble Spread and SSR. The results indicate that the SSR of all methods increases linearly with the Ensemble Spread, demonstrating a positive correlation between model prediction uncertainty and error. Furthermore, the distributions of all methods closely align with the diagonal, exhibiting strong linearity and confirming that the Ensemble Spread effectively reflects prediction uncertainty.

Based on the combined results of similarity analysis, precipitation prediction, Spread-Skill evaluation AtmoDist-SW and Cosine methods have been selected as the primary focus for the second phase of comparison and analysis. These two methods consistently exhibit relatively stable and favorable performance across multiple metrics, making them

suitable for further application and evaluation.



4.1.4 Brier Skill Score Evaluation

In this section, we assess the performance of the AtmoDist-SW and Cosine methods by comparing their analogue set members in terms of variable similarity and precipitation prediction accuracy. The evaluation is conducted using the Brier Skill Score (BSS), which quantifies the relative improvement of AtmoDist-SW over the Cosine method. By analyzing BSS across different variables and precipitation thresholds, we aim to understand how each method captures atmospheric similarity and its effectiveness in forecasting precipitation events.

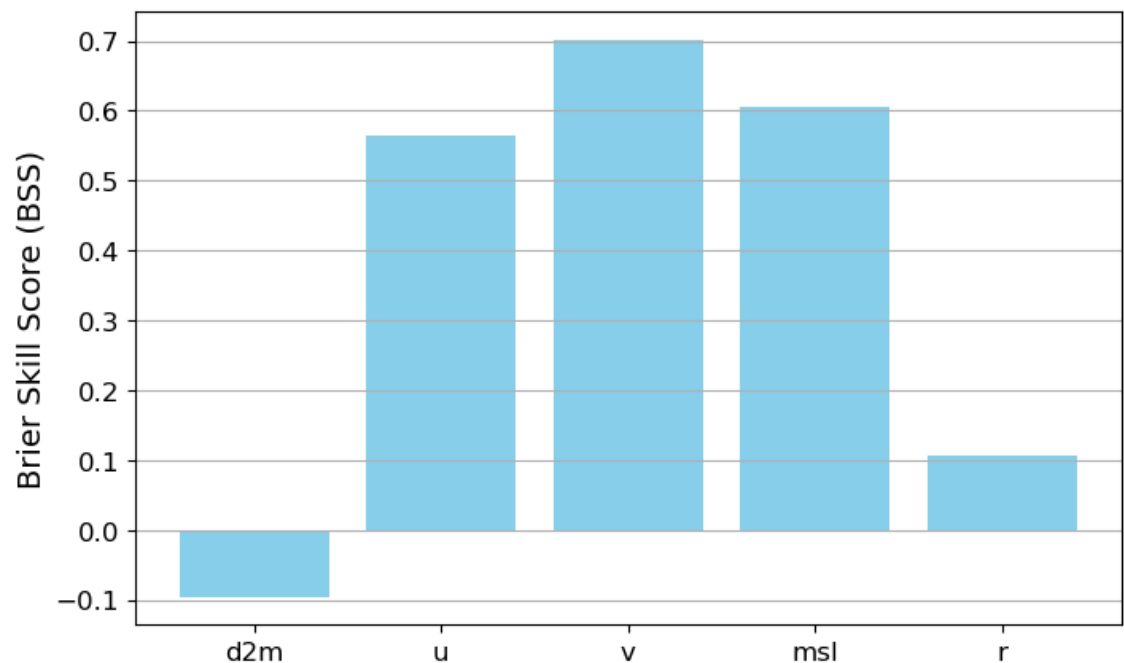


Figure 4.8 Brier Skill Score for Different Variables

The Brier Skill Score (BSS) analysis highlights the varying effectiveness of AtmoDist-

SW and the Cosine method across different atmospheric variables. Figure 4.8 provides a visual representation of these findings, illustrating the model's varying performance across different variables. Notably, AtmoDist-SW achieves significantly higher BSS values in zonal (u) and meridional (v) wind components, reaching approximately 0.6 and 0.7, respectively, indicating its strong capability in representing wind field variations. Similarly, sea level pressure (msl) also shows a high BSS, suggesting that AtmoDist-SW effectively captures large-scale circulation patterns. In contrast, relative humidity (r) exhibits only a moderate improvement, which may be attributed to its high spatial and temporal variability that limits the benefits of analogue-based selection. Dew point temperature (d2m) stands out as an exception, with a negative BSS, indicating that the Cosine method provides a better representation of near-surface moisture. These trends are quantitatively summarized in Table ??, further validating the contrasting performances of the two methods.

Variable	AtmoDist-SW (BS)	Cosine (BS)	BSS
d2m	0.493	0.450	-0.0957
u	0.167	0.382	0.564
v	0.080	0.269	0.702
msl	0.190	0.483	0.606
r	0.199	0.222	0.106

Table 4.1 Comparison of BS and BSS values for different variables between AtmoDist-SW and Cosine methods.

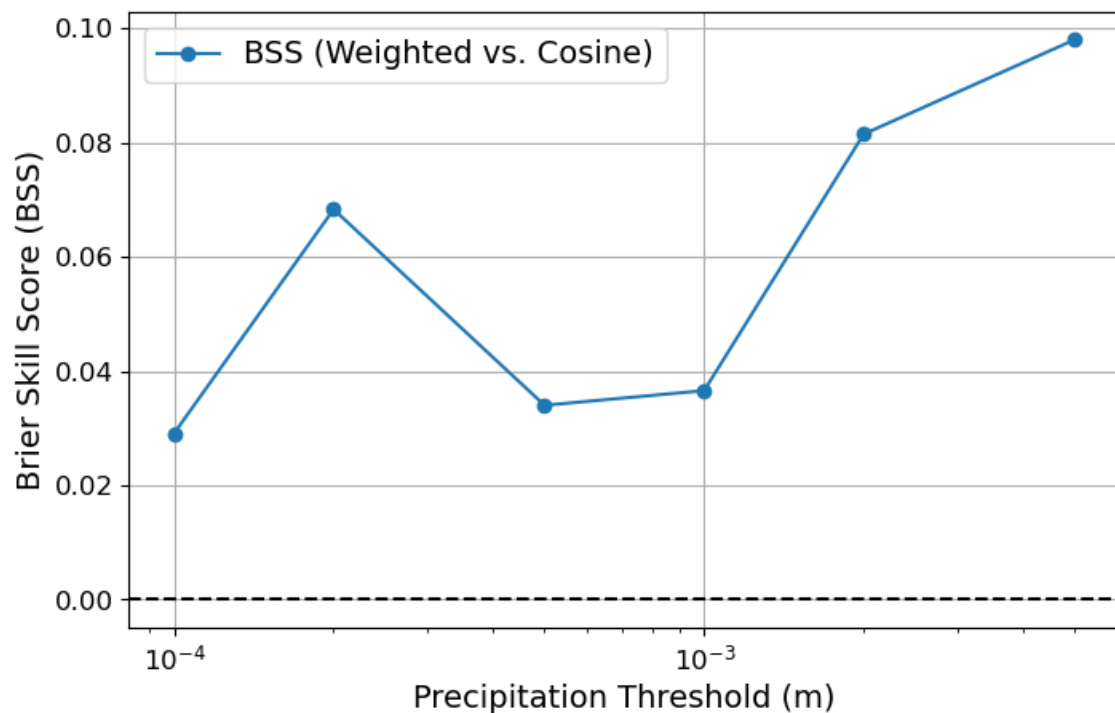


Figure 4.9 Brier Skill Score for Precipitation Prediction

Figure 4.9 presents the Brier Skill Score (BSS) as a function of precipitation threshold, comparing the performance of AtmoDist-SW and the Cosine method. The results indicate that AtmoDist-SW consistently outperforms the Cosine method across all precipitation thresholds, with BSS values remaining positive throughout. At lower thresholds (10^{-4} to 10^{-3} m), BSS exhibits fluctuations, potentially due to the higher uncertainty associated with low-intensity precipitation events. A local maximum of approximately 0.07 is observed near $10^{-4.5}$ m, suggesting that AtmoDist-SW effectively captures precipitation similarities within this range. However, BSS decreases slightly before stabilizing around 10^{-3} m. As the precipitation threshold increases, BSS shows a clear upward trend, with a notable rise near 10^{-2} m, reaching approximately 0.1 at the highest threshold.

The numerical values of BSS across different precipitation thresholds are summarized in Table ???. Consistent with Figure 4.9, AtmoDist-SW achieves a BSS of 0.098 at the 0.005 m threshold, highlighting its substantial improvement over the Cosine method in higher precipitation ranges. At lower thresholds, such as 0.0001 m and 0.0002 m, the BSS values remain positive but relatively lower, with fluctuations that align with the observed variability in Figure 4.9. This further confirms that AtmoDist-SW provides more reliable analogue selection, particularly for extreme precipitation cases where its performance improvement is most pronounced.

Precipitation Threshold (m)	AtmoDist-SW (BS)	Cosine (BS)	BSS
0.0001	0.213	0.220	0.029
0.0002	0.160	0.172	0.068
0.0005	0.107	0.110	0.034
0.0010	0.060	0.062	0.037
0.0020	0.021	0.023	0.081
0.0050	0.001	0.001	0.098

Table 4.2 Comparison of BS and BSS values for different precipitation thresholds between AtmoDist-SW and Cosine methods.

4.2 Phase II: Localized Precipitation Forecasting



This section presents an evaluation of the proposed method for generating localized precipitation forecasts using Nimrod radar data. Figure 4.10 illustrates the radar imagery of a precipitation event observed at 13:00 on August 7, 2011, at its initial stage, along with precipitation projections for the subsequent 2–3 hours. The following results compare the forecasts produced by the analog members generated in the first phase with the observed values for this event.

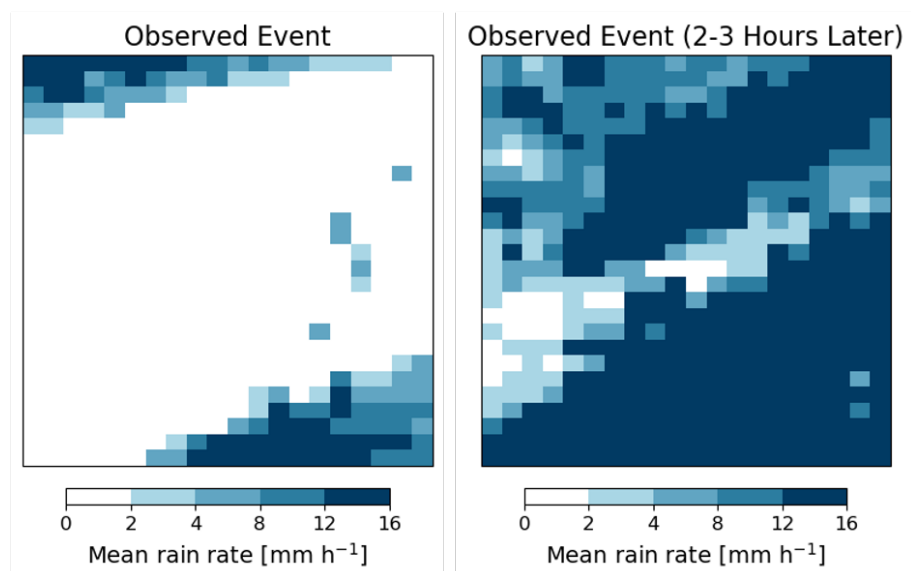


Figure 4.10 Observed Precipitation Patterns

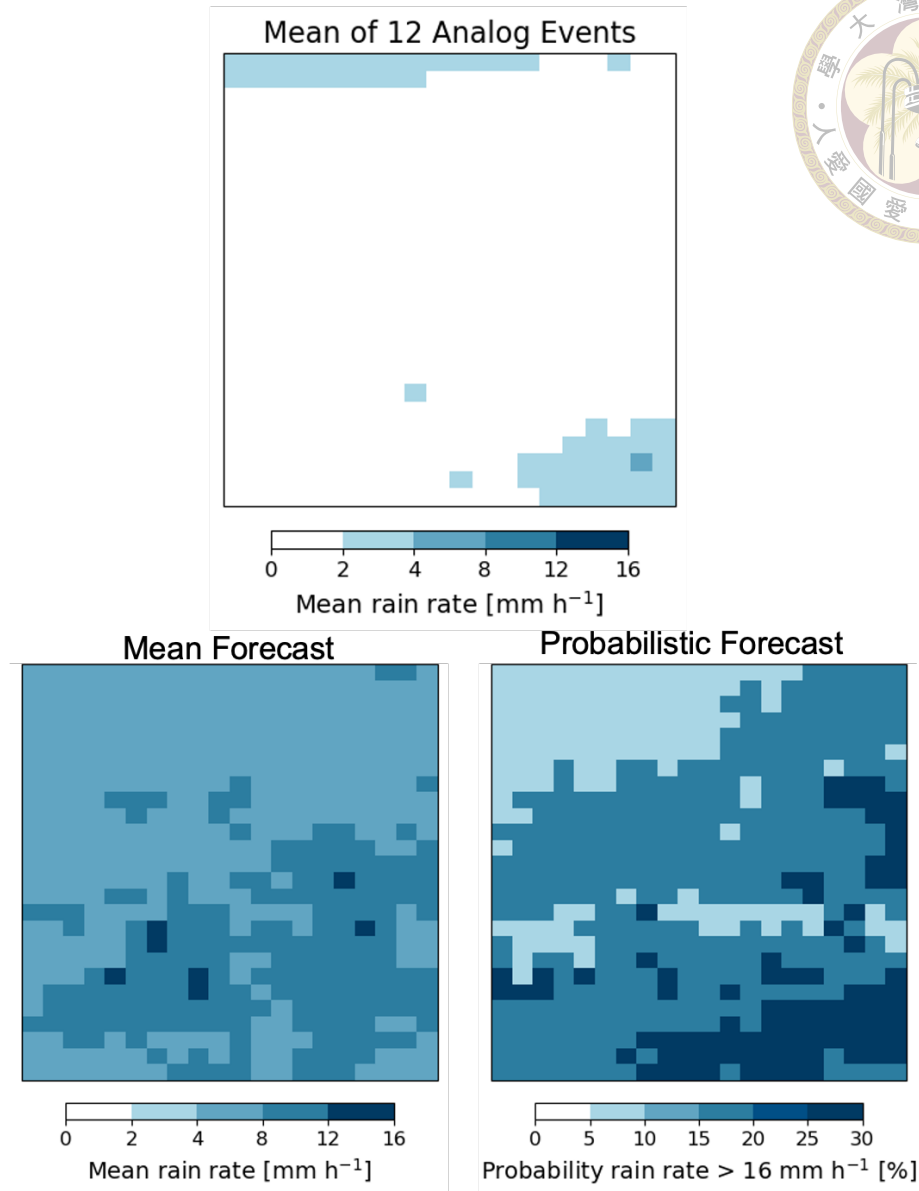


Figure 4.11 Localized precipitation forecasts using AtmoDist-SW. The top panel shows the mean precipitation rate derived from 12 analog events. The bottom panels present the deterministic mean forecast (left) and the probabilistic precipitation forecast (right), both representing the expected precipitation over the next 2–3 hours.

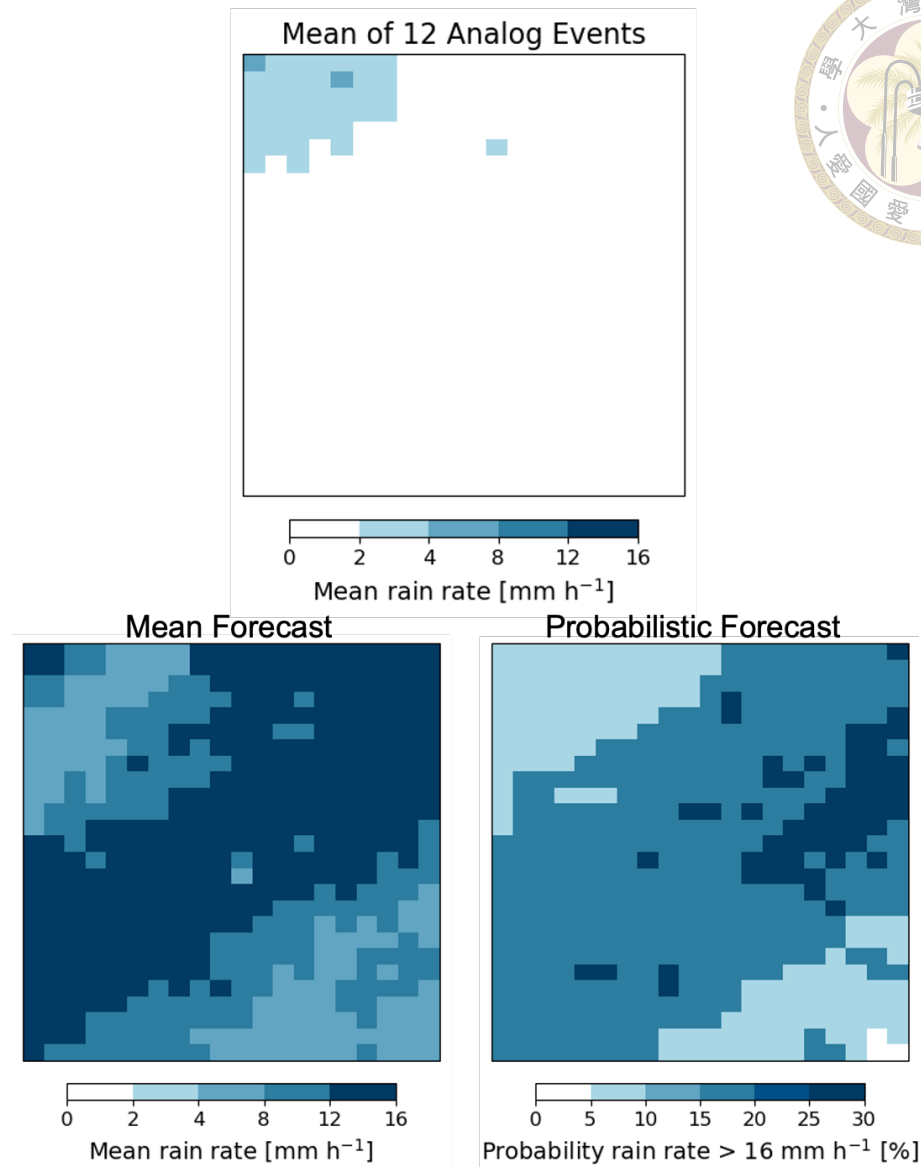


Figure 4.12 Localized Precipitation Forecasts Using Cosine Similarity. The top panel shows the mean precipitation rate derived from 12 analog events. The bottom panels present the deterministic mean forecast (left) and the probabilistic precipitation forecast (right), both representing the expected precipitation over the next 2–3 hours.

In light of the analysis presented in Figures 4.10, 4.11, and 4.12, the ensemble members selected by the AtmoDist-SW method exhibit spatial and intensity distributions that closely match observed rainfall events. This approach effectively captures the overall precipitation trends, though it tends to underestimate localized heavy rainfall, particularly in

the core regions of intense precipitation. The mean absolute error (MAE) further supports this observation, with AtmoDist-SW achieving a lower MAE of 2.125 compared to 2.201 for the Cosine method, indicating its superior ability to approximate observed precipitation patterns. In contrast, the members chosen by the Cosine method display more pronounced discrepancies with observations in both spatial distribution and precipitation characteristics, indicating weaker representation of local features.

Regarding the analog ensemble mean forecast, the AtmoDist-SW method captures large-scale precipitation patterns but underestimates the intensity within the heavy rainfall core areas. Although the Cosine method is more responsive to localized heavy rainfall, its resulting distributions are relatively fragmented. In some regions, rainfall intensity deviates from observations, thereby reducing forecast reliability.

For probabilistic forecasts, the AtmoDist-SW method concentrates the probability of low-intensity rainfall over the core precipitation regions, reflecting a higher degree of stability. However, it shows low probability estimates for high-intensity rainfall and covers only a limited spatial extent, pointing to insufficient sensitivity to extreme rainfall events. On the other hand, the Cosine method demonstrates a more dispersed probability distribution: it casts a wider net for lower-intensity rainfall probabilities but assigns lower probabilities in heavy rainfall regions. Moreover, while the high-intensity rainfall prob-

ability covers a larger area, the predicted structure differs markedly from observations, revealing instability and randomness.

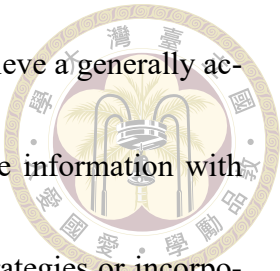


These outcomes are closely linked to the use of the L2 Norm as the criterion for final member selection. Since the L2 Norm emphasizes minimizing overall errors, both methods tend to select members with smoother distributions and coherent large-scale consistency during the analog-based screening. Consequently, the distinctive features of heavy rainfall core regions are attenuated. Furthermore, the L2 Norm's inherent low sensitivity to extreme precipitation may treat localized heavy rainfall deviations as comparable to widespread weaker rainfall deviations, further constraining its ability to represent extreme events.

Overall, the AtmoDist-SW method yields more robust predictions and is suitable for describing precipitation trends and large-scale structures, though it adopts a conservative stance on heavy rainfall events. Although the Cosine method is more responsive to localized intense precipitation, the forecast patterns are scattered, with potential biases that undermine the overall reliability.

In conclusion, using large-scale data for small-scale precipitation forecasting is inherently challenging, and the high-resolution, unstable nature of radar precipitation data further underscores its complexity. This study focuses on comparing different methods

and demonstrates that, despite some localized discrepancies, they achieve a generally acceptable level of predictive skill and effectively integrate large-scale information with radar echo characteristics. Future efforts aimed at refining analog strategies or incorporating additional data sources could further enhance sensitivity and accuracy in forecasting localized heavy rainfall, thus expanding opportunities for applying large-scale data to small-scale predictive endeavors.



4.3 Explaining Prediction Results Using DeepSHAP

To further understand the differences in performance between AtmoDist-SW and AtmoDist, the following analysis will compare the results of DeepSHAP calculations. The objective is to reveal differences in the proportional contributions of features across various variables, as well as their spatial and temporal distributions.

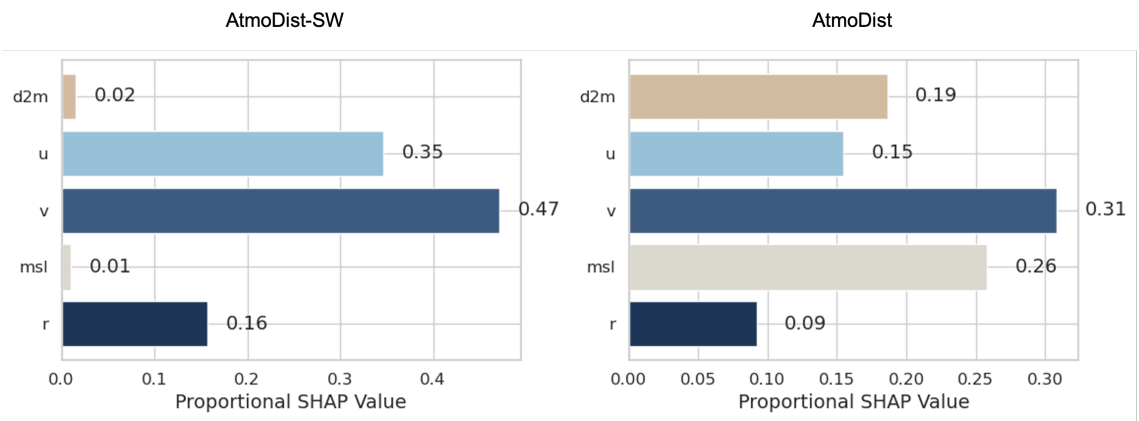


Figure 4.13 Comparison of Proportional SHAP Values Between AtmoDist-SW and AtmoDist

Figure 4.13 compares the importance of atmospheric variables between the AtmoDist-SW and AtmoDist methods, with the contributions of each variable to the prediction results measured using SHAP values. As shown in the figure, the AtmoDist-SW method (left panel) primarily relies on wind-related variables, where the v wind component (0.47) and u wind component (0.35) dominate, indicating a heightened sensitivity to local spatial variations in wind field features.

In contrast, the AtmoDist method (right panel) shows a more balanced distribution of variable contributions. While the v wind component (0.31) remains significant, the contributions of $d2m$ (0.19) and msl (0.26) increase notably, reflecting a more comprehensive consideration of multiple atmospheric variables.

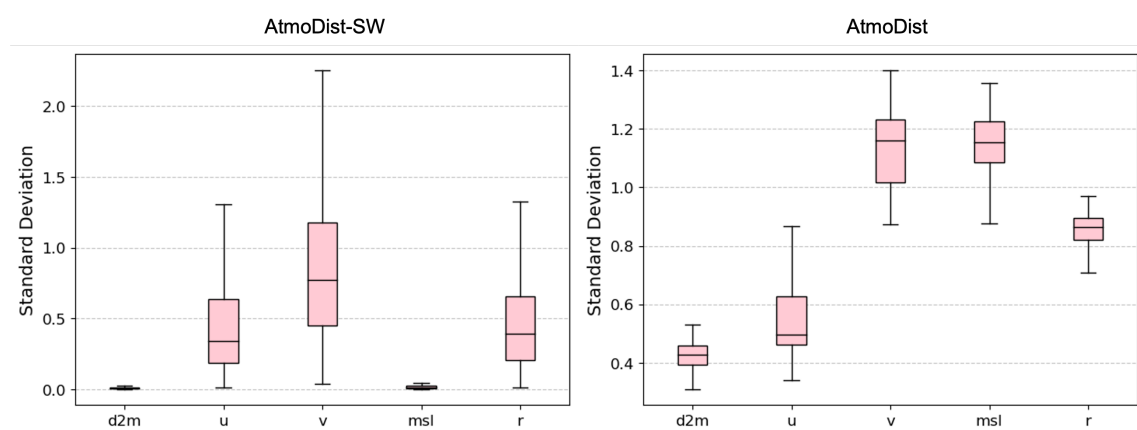


Figure 4.14 Comparison of Standard Deviation Distributions Under Two Data Processing Methods

Each of the 32x32 grid points in the study area was analyzed using ERA5 time series data from 2001 to 2020. Standard deviations were calculated for each point, and the results are summarized in boxplots. The left panel (AtmoDist-SW) applies spatial weighting and logarithmic transformation, while the right panel (AtmoDist) uses standardization, providing a comparison of variability across different variables.

Figure 4.14 compares the standard deviation distributions of variables processed using two distinct methods. In the left panel, the standard deviation distributions for variables $d2m$ and msl exhibit relatively low variability, with values concentrated near zero. This indicates that, following the logarithmic transformation, these two variables undergo smaller variations and maintain more stable values compared to other variables, which show higher variability.

In contrast, the right panel (AtmoDist) demonstrates a more compact and consistent distribution of standard deviations across variables after standardization. This highlights balanced variability and emphasizes relative differences among variables.

Examining the characteristics of SHAP values alongside standard deviation distributions reveals a strong alignment in variability patterns. Variables with lower standard deviations correspond to lower SHAP importance values, whereas those with higher variability tend to exhibit higher SHAP values.

Although the variable r displays higher variability compared to u , its distribution characteristics (refer to Figure 2.2) show skewness and greater dispersion. This lack of a centralized trend may reduce the model's sensitivity and interpretability regarding this feature, thereby impacting its importance evaluation within the model.

Overall, the results obtained from these two data processing methods exhibit a notable correspondence with SHAP value distributions. This alignment suggests a strong association between variable importance and spatiotemporal variability characteristics.

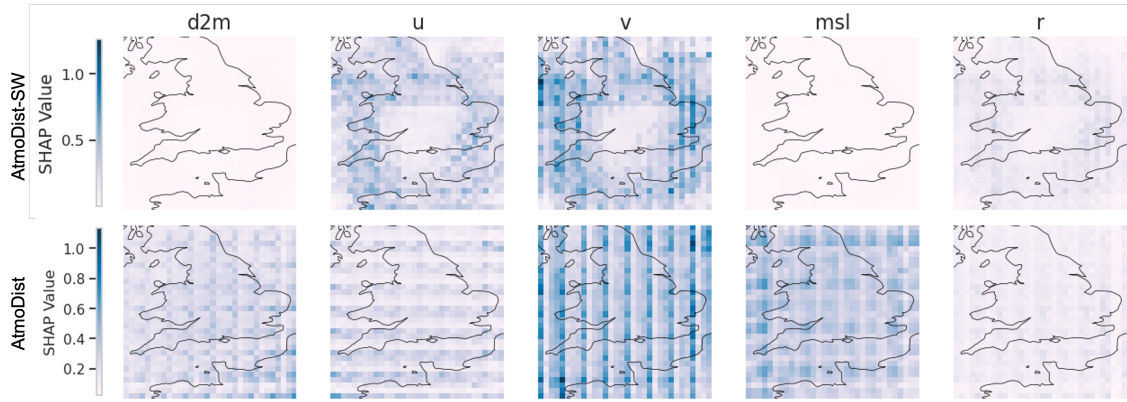


Figure 4.15 Spatial Distribution of SHAP Values for AtmoDist-SW and AtmoDist Across Variables

Figure 4.15 illustrates the spatial distribution of SHAP values for different atmospheric variables using the AtmoDist-SW and AtmoDist methods. The results show that the SHAP values from the AtmoDist-SW method are more localized, particularly for wind-related variables (u and v), where higher local importance is observed. In contrast, the AtmoDist method exhibits a more uniform SHAP value distribution. Besides the wind components, variables such as $d2m$ and msl also display notable increases in SHAP values, reflecting a more comprehensive consideration of multiple variables and global features.

Interestingly, the AtmoDist-SW method shows lower SHAP values in regions where higher spatial weights are applied. This could be because the model avoids over-relying on

the weight-enhanced information in these regions, instead maximizing the use of overall spatial information to achieve more robust predictions. Moreover, due to the high spatial continuity of atmospheric variables, the model can effectively capture features in the surrounding areas of high-weight regions, achieving satisfactory prediction results while reducing dependency on the high-weight regions themselves. This phenomenon, however, requires further verification to confirm the underlying reasons.

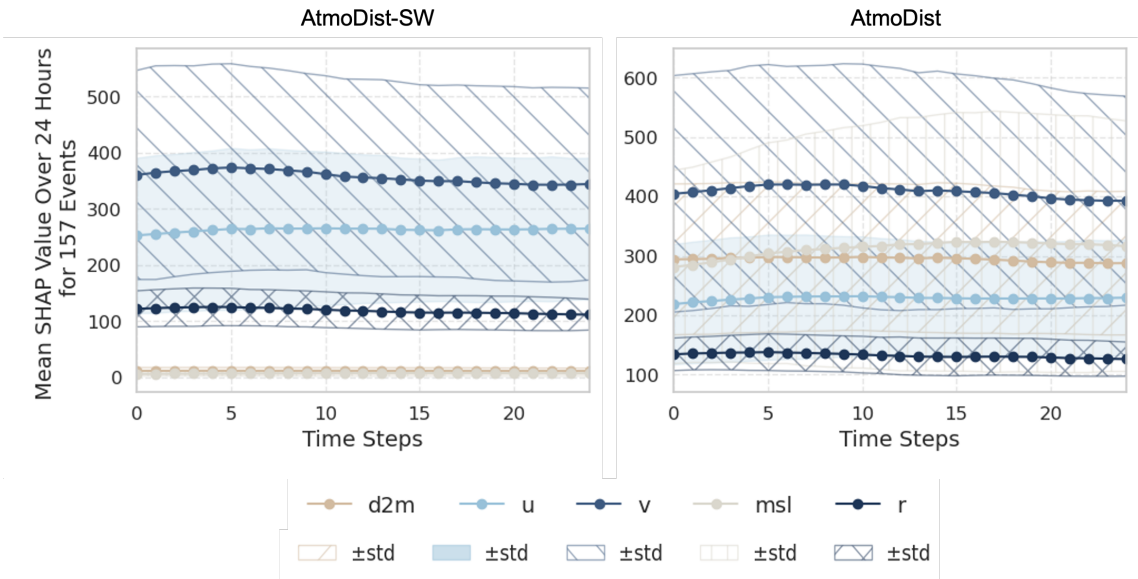


Figure 4.16 Mean SHAP Value Over Time Steps for AtmoDist-SW and AtmoDist Across 157 Events

Figure 4.16 presents the average SHAP value variations over 24 consecutive hours for 157 precipitation events using the AtmoDist-SW (left) and AtmoDist (right) methods. The results show that variables with larger SHAP values tend to exhibit greater standard deviations. This is because variables contributing more significantly to the prediction results are more sensitive to feature variations across different events and time steps, leading

to larger fluctuations in SHAP values. Conversely, variables with smaller contributions, such as $d2m$ and msl , exhibit lower SHAP values and standard deviations, indicating higher stability.



In the AtmoDist-SW method, the SHAP value variations across time steps are smoother compared to the AtmoDist method, with narrower standard deviation ranges. This suggests more stable feature contributions over time.

In summary, the SHAP value analyses across the three figures highlight the differences in feature contributions and temporal variations between the AtmoDist-SW and AtmoDist methods. The distribution and fluctuation trends of SHAP values reflect the differences in how these models learn and utilize features. These results provide valuable insights into the feature utilization and stability of the models.



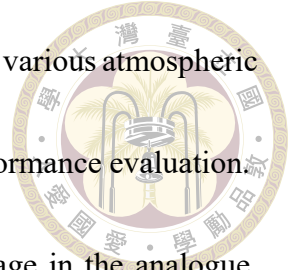


Chapter 5 Conclusion and Outlook

5.1 Conclusion

In this study, we utilized the AtmoDist machine learning model to calculate atmospheric state similarity and predict precipitation events. To optimize its performance, we modified the model by removing certain structural components while preserving its core framework, enhancing its ability to focus on spatial features. Additionally, spatial weighting and a two-phase weather analog method were applied to further improve its sensitivity to localized regions.

Building on these enhancements, The performance of AtmoDist was systematically compared with traditional methods, providing a comprehensive evaluation of its predictive capabilities. Furthermore, explainable AI (XAI) techniques were employed to analyze the model's mechanisms, offering insights into feature contributions and the underlying decision-making process. The main findings of this study are summarized as follows:



1. Different similarity calculation methods exhibit strengths across various atmospheric variables, highlighting the importance of variable-specific performance evaluation. For instance, AtmoDist-SW demonstrates a dominant advantage in the analogue results for msl, whereas the Cosine method performs significantly better in d2m compared to other approaches.
2. Spatially weighted data processing effectively enhances the model's ability to predict localized precipitation events. Based on SSR evaluation of precipitation, AtmoDist-SW ranks in the top two positions significantly more often across 157 events, indicating superior predictive capability.
3. The study reveals that the accuracy of a model does not necessarily reflect its ability to predict weather events effectively. While AtmoDist achieves over 40% accuracy, compared to just over 30% for AtmoDist-SW, the latter demonstrates a clear advantage in precipitation prediction performance.
4. The time-based AtmoDist model demonstrates superior performance in precipitation prediction compared to traditional methods. In addition to SSR indicating better performance for machine learning-based methods, BSS evaluation further confirms that AtmoDist-SW outperforms the Cosine method, which represents a conventional similarity approach, across nearly all variables, consistently yielding positive val-

ues. For different precipitation thresholds, the BSS values remain within the range of 0.03 to 0.1, further demonstrating the advantage of AtmoDist-SW. In the second phase, radar-based precipitation evaluation also shows that AtmoDist-SW achieves lower MAE.

5. Through the application of DeepSHAP, it is observed that the AtmoDist model places greater focus on atmospheric variables with high spatial-temporal variability.

These findings emphasize the necessity of dynamically adjusting the selection of variables and data processing strategies for different weather events. Moreover, the training objectives of a model should be determined by downstream tasks to optimize task performance, ensuring that the methods and metrics align closely with the intended application. This approach provides a pathway for improving the robustness and adaptability of models in atmospheric prediction tasks.

5.2 Future Work and Outlook

This study presents a comprehensive prediction framework that integrates atmospheric similarity calculations, precipitation event prediction, spatial weighting, a two-phase weather analog method, and explainable AI (XAI) techniques. By systematically refining the model structure and optimizing its spatial sensitivity, we have developed an approach

that not only improves precipitation forecasting accuracy but also enhances interpretability. The findings from this study provide a strong foundation for future advancements in atmospheric modeling, particularly in adapting machine learning techniques to dynamic weather systems.



Looking ahead, we aim to further refine this framework to address two key challenges: (1) improving the model's ability to adaptively weigh atmospheric variables and data processing strategies, and (2) mitigating periodicity-related issues in atmospheric similarity calculations. These improvements will allow the framework to become more adaptive, interpretable, and resilient to complex atmospheric patterns, ultimately contributing to more reliable precipitation forecasting and risk assessment applications.

Reinforcement Learning for Adaptive Decision-Making Building on the findings of this study, reinforcement learning (RL) presents a promising direction to further enhance atmospheric similarity calculations and precipitation event prediction. Unlike traditional methods, RL provides an adaptive decision-making framework, enabling the model to dynamically adjust the importance of different atmospheric variables and data processing strategies based on real-time feedback.

By leveraging RL's ability to maximize long-term cumulative rewards, models can

optimize their predictions by learning from sequential atmospheric states, aligning well with the temporal and spatial dynamics of weather events. For instance, RL-based approaches could adaptively fine-tune spatial weighting schemes or optimize ensemble member selection to improve localized prediction accuracy. Additionally, incorporating explainable RL methods would offer deeper insights into the decision-making process, ensuring the framework remains interpretable, robust, and adaptable to real-world atmospheric conditions.

Overall, reinforcement learning offers a pathway toward making weather analog methods more flexible, stable, and accurate, paving the way for more effective and data-driven atmospheric event forecasting in complex, dynamic environments.

Addressing Periodicity Issues in Atmospheric Similarity Calculations In addition to enhancing adaptability, this study also identifies a periodicity issue within the data, where the assumption that greater time intervals correspond to greater atmospheric distances does not always hold. Such periodic behaviors introduce challenges in accurately capturing the temporal relationships between atmospheric states, potentially affecting the model's predictive reliability.

To address this, the triplet loss method presents a viable solution by optimizing fea-



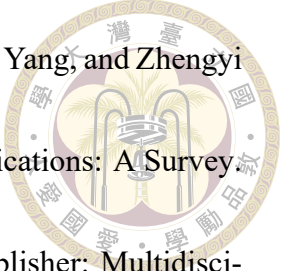
ture space distances, ensuring that the learned representations of atmospheric states remain consistent and meaningful. By dynamically adjusting feature embeddings, triplet loss can enforce a more structured representation of atmospheric similarities and differences, improving the model's ability to generalize and enhancing training accuracy. This is particularly valuable in scenarios where periodicity disrupts the expected temporal-distance relationship.

Towards a More Robust Atmospheric Prediction Framework By integrating reinforcement learning for adaptive decision-making and triplet loss optimization for more robust feature learning, this study lays the groundwork for a more resilient and precise atmospheric similarity modeling framework. These advancements will provide greater flexibility in handling complex weather patterns, improve the model's ability to generalize across diverse atmospheric conditions, and ultimately enhance the accuracy and reliability of precipitation event forecasting.



References

- [1] Yukiko Hirabayashi, Haireti Alifu, Dai Yamazaki, Yukiko Imada, Hideo Shiogama, and Yuki Kimura. Anthropogenic climate change has changed frequency of past flood during 2010-2013. Progress in Earth and Planetary Science, 8(1):1–9, December 2021. Number: 1 Publisher: SpringerOpen.
- [2] Luca Delle Monache, F. Anthony Eckel, Daran L. Rife, Badrinath Nagarajan, and Keith Searight. Probabilistic Weather Prediction with an Analog Ensemble. Monthly Weather Review, 141(10):3498–3516, October 2013. Publisher: American Meteorological Society Section: Monthly Weather Review.
- [3] Pascal Horton. AtmoSwing: Analog Technique Model for Statistical Weather forecastING and downscaling (v2.1.0). Geoscientific Model Development, 12(7):2915–2940, July 2019. Publisher: Copernicus GmbH.
- [4] Ashesh Chattopadhyay, Ebrahim Nabizadeh, and Pedram Hassanzadeh. Analog forecasting of extreme-causing weather patterns using deep learning, July 2019.

- 
- [5] Liuyi Chen, Bocheng Han, Xuesong Wang, Jiazhen Zhao, Wenke Yang, and Zhengyi Yang. Machine Learning Methods in Weather and Climate Applications: A Survey. Applied Sciences, 13(21):12019, January 2023. Number: 21 Publisher: Multidisciplinary Digital Publishing Institute.
- [6] Maria J. Molina, Travis A. O'Brien, Gemma Anderson, Moetasim Ashfaq, Katrina E. Bennett, William D. Collins, Katherine Dagon, Juan M. Restrepo, and Paul A. Ullrich. A Review of Recent and Emerging Machine Learning Applications for Climate Variability and Weather Phenomena. September 2023. Section: Artificial Intelligence for the Earth Systems.
- [7] Catherine O. de Burgh-Day and Tennessee Leeuwenburg. Machine learning for numerical weather and climate modelling: a review. Geoscientific Model Development, 16(22):6433–6477, November 2023. Publisher: Copernicus GmbH.
- [8] Bogdan Bochenek and Zbigniew Ustrnul. Machine Learning in Weather Prediction and Climate Analyses—Applications and Perspectives. Atmosphere, 13(2):180, February 2022. Number: 2 Publisher: Multidisciplinary Digital Publishing Institute.
- [9] Amy McGovern, Ryan Lagerquist, David John Gagne, G. Eli Jergensen, Kimberly L. Elmore, Cameron R. Homeyer, and Travis Smith. Making the Black Box

More Transparent: Understanding the Physical Implications of Machine Learning.

November 2019. Section: Bulletin of the American Meteorological Society.



- [10] Leonardo Olivetti and Gabriele Messori. Advances and prospects of deep learning for medium-range extreme weather forecasting. Geoscientific Model Development, 17(6):2347–2358, March 2024. Publisher: Copernicus GmbH.
- [11] Kevin J. Gutzwiller and Kimberly M. Serno. Using the risk of spatial extrapolation by machine-learning models to assess the reliability of model predictions for conservation. Landscape Ecology, 38(6):1363–1372, June 2023.
- [12] Peter AG Watson. Machine learning applications for weather and climate need greater focus on extremes, September 2022. arXiv:2207.07390.
- [13] Xiaoli Ren, Xiaoyong Li, Kaijun Ren, Junqiang Song, Zichen Xu, Kefeng Deng, and Xiang Wang. Deep Learning-Based Weather Prediction: A Survey. Big Data Research, 23:100178, February 2021.
- [14] Weiming Hu, Guido Cervone, George Young, and Luca Delle Monache. Machine Learning Weather Analogs for Near-Surface Variables. Boundary-Layer Meteorology, 186(3):711–735, March 2023. Company: Springer Distributor:

Springer Institution: Springer Label: Springer Number: 3 Publisher: Springer
Netherlands.



- [15] Kimberly L. Elmore and Michael B. Richman. Euclidean Distance as a Similarity Metric for Principal Component Analysis. Monthly Weather Review, 129(3):540–549, March 2001. Publisher: American Meteorological Society Section: Monthly Weather Review.
- [16] Baiquan Zhou and Panmao Zhai. A New Forecast Model Based on the Analog Method for Persistent Extreme Precipitation. Weather and Forecasting, 31(4):1325–1341, August 2016. Publisher: American Meteorological Society Section: Weather and Forecasting.
- [17] Peter Hoffmann, Jascha Lehmann, Bijan Fallah, and Fred F. Hattermann. Atmosphere similarity patterns in boreal summer show an increase of persistent weather conditions connected to hydro-climatic risks. Scientific Reports, 11(1):22893, November 2021. Publisher: Nature Publishing Group.
- [18] Weiming Hu, Guido Cervone, George Young, and Luca Delle Monache. Weather Analogs with a Machine Learning Similarity Metric for Renewable Resource Forecasting, March 2021. arXiv:2103.04530.

[19] Sebastian Hoffmann and Christian Lessig. AtmoDist: Self-supervised Representation Learning for Atmospheric Dynamics, August 2022. arXiv:2202.01897 [physics].



[20] C. Matulla, X. Zhang, X. L. Wang, J. Wang, E. Zorita, S. Wagner, and H. von Storch. Influence of similarity measures on the performance of the analog method for down-scaling daily precipitation. Climate Dynamics, 30(2):133–144, February 2008. Company: Springer Distributor: Springer Institution: Springer Label: Springer Number: 2 Publisher: Springer-Verlag.

[21] L. Panziera, U. Germann, M. Gabella, and P. V. Mandapaka. NORA – Now-casting of Orographic Rainfall by means of Analogues. Quarterly Journal of the Royal Meteorological Society, 137(661):2106–2123, 2011. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/qj.878>.

[22] Vimal Mishra, Francina Dominguez, and Dennis Lettenmaier. Urban precipitation extremes: How reliable are regional climate models? Geophysical Research Letters, 39:3407, February 2012.

[23] Michael D. Warner, Clifford F. Mass, and Eric P. Salathé. Wintertime Extreme Precipitation Events along the Pacific Northwest Coast: Climatology and Synop-

tic Evolution. Monthly Weather Review, 140(7):2021–2043, July 2012. Publisher:
American Meteorological Society Section: Monthly Weather Review.



[24] Paul Gregory, Frederic Vitart, Rabi Rivett, Andrew Brown, and Yuriy Kuleshov.

Subseasonal Forecasts of Tropical Cyclones in the Southern Hemisphere Using a
Dynamical Multimodel Ensemble. Weather and Forecasting, 35(5):1817–1829, Oc-
tober 2020. Publisher: American Meteorological Society Section: Weather and Fore-
casting.

[25] David A. Lavers and Gabriele Villarini. The nexus between atmospheric rivers and
extreme precipitation across Europe.

[26] Shih-Yu Wang, Jin-Ho Yoon, Robert R. Gillies, and Changrae Cho. What Caused the
Winter Drought in Western Nepal during Recent Years? November 2013. Section:
Journal of Climate.

[27] Hongyu Chen, Tim Li, and Jing Cui. The Reexamination of the Moisture – Vortex
and Baroclinic Instabilities in the South Asian Monsoon. Atmosphere, 15(2):147,
January 2024.

[28] Jean-Philippe Baudouin, Michael Herzog, and Cameron A. Petrie. Contribution of
Cross-Barrier Moisture Transport to Precipitation in the Upper Indus River Basin.

Monthly Weather Review, 148(7):2801–2818, July 2020. Publisher: American Meteorological Society Section: Monthly Weather Review.



[29] Ashesh Chattopadhyay, Pedram Hassanzadeh, and Saba Pasha. Predicting clustered weather patterns: A test case for applications of convolutional neural networks to spatio-temporal climate data. Scientific Reports, 10(1):1317, January 2020. Publisher: Nature Publishing Group.

[30] Peter D. Dueben and Peter Bauer. Challenges and design choices for global weather and climate models based on machine learning. Geoscientific Model Development, 11(10):3999–4009, October 2018. Publisher: Copernicus GmbH.

[31] Nigel M. Roberts and Humphrey W. Lean. Scale-Selective Verification of Rainfall Accumulations from High-Resolution Forecasts of Convective Events. January 2008. Section: Monthly Weather Review.

[32] Quantitative precipitation forecasting in the UK. Journal of Hydrology, 239(1-4):286–305, December 2000. Publisher: Elsevier.

[33] H. M. Van Den Dool. Searching for analogues, how long must we wait? Tellus A, 46(3):314–324, 1994. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1034/j.1600-0870.1994.t01-2-00006.x>.

[34] S. Radanovics, J.-P. Vidal, E. Sauquet, A. Ben Daoud, and G. Bontron. Optimising predictor domains for spatially coherent precipitation downscaling. Hydrology and Earth System Sciences, 17(10):4189–4208, October 2013. Publisher: Copernicus GmbH.

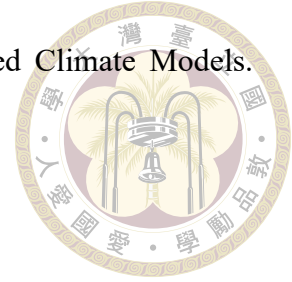
[35] Charles Obled, Guillaume Bontron, and Rémy Garçon. Quantitative precipitation forecasts: a statistical adaptation of model outputs through an analogues sorting approach. Atmospheric Research, 63(3):303–324, August 2002.

[36] Thomas M. Hamill and Jeffrey S. Whitaker. Probabilistic Quantitative Precipitation Forecasts Based on Reforecast Analogs: Theory and Application. November 2006. Section: Monthly Weather Review.

[37] Renaud Marty, Isabella Zin, Charles Obled, Guillaume Bontron, and Abdelatif Djerboua. Toward Real-Time Daily PQPF by an Analog Sorting Approach: Application to Flash-Flood Catchments. March 2012. Section: Journal of Applied Meteorology and Climatology.

[38] Luca Delle Monache, Thomas Nipen, Yubao Liu, Gregory Roux, and Roland Stull. Kalman Filter and Analog Schemes to Postprocess Numerical Weather Predictions. Monthly Weather Review, 139(11):3554–3570, November 2011. Publisher: American Meteorological Society Section: Monthly Weather Review.

- [39] Aiguo Dai. Precipitation Characteristics in Eighteen Coupled Climate Models. September 2006. Section: Journal of Climate.

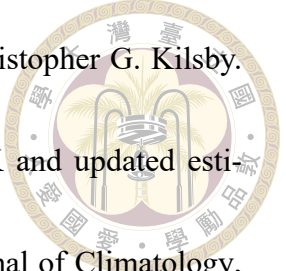


- [40] K. E. Trenberth and C. J. Guillemot. Evaluation of the atmospheric moisture and hydrological cycle in the NCEP/NCAR reanalyses. Climate Dynamics, 14(3):213–231, March 1998.

- [41] Yong Zhu and Reginald E. Newell. A Proposed Algorithm for Moisture Fluxes from Atmospheric Rivers. March 1998. Section: Monthly Weather Review.

- [42] Luis Gimeno, Francina Dominguez, Raquel Nieto, Ricardo Trigo, Anita Drumond, Chris J. C. Reason, Andréa S. Taschetto, Alexandre M. Ramos, Ramesh Kumar, and José Marengo. Major Mechanisms of Atmospheric Moisture Transport and Their Role in Extreme Precipitation Events. Annual Review of Environment and Resources, 41(Volume 41, 2016):117–141, October 2016. Publisher: Annual Reviews.

- [43] Aurélien Ben Daoud, Eric Sauquet, Guillaume Bontron, Charles Obled, and Michel Lang. Daily quantitative precipitation forecasts based on the analogue method: Improvements and application to a French large river basin. Atmospheric Research, 169:147–159, March 2016. ADS Bibcode: 2016AtmRe.169..147B.

- 
- [44] Mari R. Jones, Stephen Blenkinsop, Hayley J. Fowler, and Christopher G. Kilsby. Objective classification of extreme rainfall regions for the UK and updated estimates of trends in regional extreme rainfall. International Journal of Climatology, 34(3):751–765, 2014. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/joc.3720>.
- [45] C. Berndt and U. Haberlandt. Spatial interpolation of climate variables in Northern Germany—Influence of temporal resolution and network density. Journal of Hydrology: Regional Studies, 15:184–202, February 2018.
- [46] Ronald B. Smith. The Influence of Mountains on the Atmosphere. In Barry Saltzman, editor, Advances in Geophysics, volume 21, pages 87–230. Elsevier, January 1979.
- [47] Brian J. Hoskins and Kevin I. Hodges. New Perspectives on the Northern Hemisphere Winter Storm Tracks. March 2002. Section: Journal of the Atmospheric Sciences.
- [48] Clifford F. Mass, Mark Albright, David Ovens, Richard Steed, Mark Maciver, Eric Gritmit, Tony Eckel, Brian Lamb, Joseph Vaughan, Kenneth Westrick, Pascal Storck, Brad Colman, Chris Hill, Naydene Maykut, Mike Gilroy, Sue A. Ferguson, Joseph Yetter, John M. Sierchio, Clint Bowman, Richard Stender, Robert Wilson, and

William Brown. Regional Environmental Prediction Over the Pacific Northwest.
October 2003. Section: Bulletin of the American Meteorological Society.



[49] Luca Delle Monache, Joshua P. Hacker, Yongmei Zhou, Xingxiu Deng, and Roland B. Stull. Probabilistic aspects of meteorological and ozone regional ensemble forecasts. Journal of Geophysical Research: Atmospheres, 111(D24), 2006. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1029/2005JD006917>.

[50] Ivan Dokmanic, Reza Parhizkar, Juri Ranieri, and Martin Vetterli. Euclidean Distance Matrices: Essential theory, algorithms, and applications. IEEE Signal Processing Magazine, 32(6):12–30, January 2015. Conference Name: IEEE Signal Processing Magazine.

[51] Christina Görner, Johannes Franke, Rico Kronenberg, and Olaf Hellmuth. Multivariate non-parametric Euclidean distance model for hourly disaggregation of daily climate data. Theoretical and Applied Climatology, 143, January 2021.

[52] Charu C. Aggarwal, Alexander Hinneburg, and Daniel A. Keim. On the Surprising Behavior of Distance Metrics in High Dimensional Space. In Jan Van den Bussche and Victor Vianu, editors, Database Theory —ICDT 2001, pages 420–434, Berlin, Heidelberg, 2001. Springer.

[53] Zihao Chen. Study on Evaporation Duct Model A Considering Turbulence Based on Cosine Similarity. In 2022 IEEE 10th Joint International Information Technology and Artificial Intelligence Conference (ITAIC), volume 10, pages 2538–2541, June 2022. ISSN: 2693-2865.



[54] Zhou Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing, 13(4):600–612, April 2004. Conference Name: IEEE Transactions on Image Processing.

[55] Allison H. Baker, Alexander Pinard, and Dorit M. Hammerling. On a Structural Similarity Index Approach for Floating-Point Data. IEEE Transactions on Visualization and Computer Graphics, pages 1–13, 2023. Conference Name: IEEE Transactions on Visualization and Computer Graphics.

[56] Glenn W. Brier. VERIFICATION OF FORECASTS EXPRESSED IN TERMS OF PROBABILITY. January 1950. Section: Monthly Weather Review.

[57] Statistical Methods in the Atmospheric Sciences. May 2011.

[58] Scott M Lundberg and Su-In Lee. A Unified Approach to Interpreting Model Predic-

tions. In Advances in Neural Information Processing Systems, volume 30. Curran Associates, Inc., 2017.



- [59] Avanti Shrikumar, Peyton Greenside, and Anshul Kundaje. Learning Important Features Through Propagating Activation Differences. In Proceedings of the 34th International Conference on Machine Learning, pages 3145–3153. PMLR, July 2017. ISSN: 2640-3498.
- [60] Scott M. Lundberg, Gabriel G. Erion, and Su-In Lee. Consistent Individualized Feature Attribution for Tree Ensembles, March 2019. arXiv:1802.03888.
- [61] Matthew D. Zeiler and Rob Fergus. Visualizing and Understanding Convolutional Networks, November 2013. arXiv:1311.2901.
- [62] Marco Ancona, Enea Ceolini, Cengiz Öztireli, and Markus Gross. Towards better understanding of gradient-based attribution methods for Deep Neural Networks, March 2018. arXiv:1711.06104.
- [63] Sungyong Seo, Jing Huang, Hao Yang, and Yan Liu. Interpretable Convolutional Neural Networks with Dual Local and Global Attention for Review Rating Prediction. In Proceedings of the Eleventh ACM Conference on Recommender Systems,

RecSys '17, pages 297–305, New York, NY, USA, August 2017. Association for Computing Machinery.



- [64] Benjamin A. Toms, Elizabeth A. Barnes, and Imme Ebert-Uphoff. Physically Interpretable Neural Networks for the Geosciences: Applications to Earth System Variability. Journal of Advances in Modeling Earth Systems, 12(9):e2019MS002002, 2020. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1029/2019MS002002>.

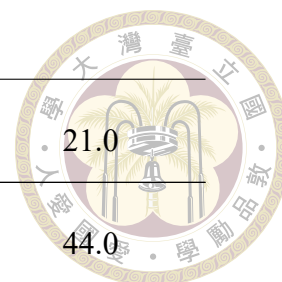


Appendix A — 157 Extreme Precipitation Events

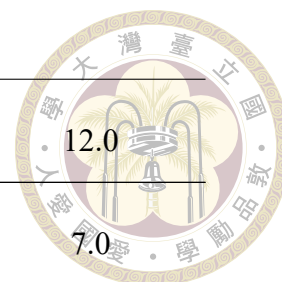
A.1 Complete List of Precipitation Events

Rank	ID	Start Datetime	End Datetime	Duration (hrs)
1	2464	2014-07-08 10:00:00	2014-07-08 20:00:00	10.0
2	632	2006-08-29 11:00:00	2006-08-29 14:00:00	3.0
3	1998	2012-07-18 05:00:00	2012-07-19 04:00:00	23.0
4	481	2005-12-31 20:00:00	2006-01-01 04:00:00	8.0
5	2879	2016-04-04 06:00:00	2016-04-04 16:00:00	10.0
6	2026	2012-08-25 12:00:00	2012-08-25 22:00:00	10.0
7	2371	2014-02-07 22:00:00	2014-02-08 05:00:00	7.0
8	1564	2010-09-06 15:00:00	2010-09-07 00:00:00	9.0
9	393	2005-08-18 16:00:00	2005-08-18 22:00:00	6.0

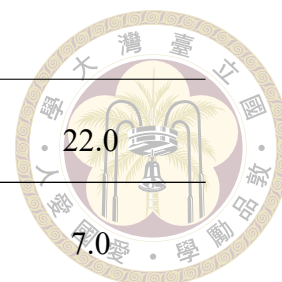
10	2533	2014-11-03 03:00:00	2014-11-04 00:00:00	21.0
11	2539	2014-11-10 15:00:00	2014-11-12 11:00:00	44.0
12	2004	2012-07-30 23:00:00	2012-07-31 07:00:00	8.0
13	1377	2009-11-12 14:00:00	2009-11-14 16:00:00	50.0
14	1433	2010-01-16 03:00:00	2010-01-16 10:00:00	7.0
15	1367	2009-11-02 22:00:00	2009-11-03 15:00:00	17.0
16	828	2007-07-03 21:00:00	2007-07-04 02:00:00	5.0
17	2459	2014-06-28 09:00:00	2014-06-28 17:00:00	8.0
18	2367	2014-02-01 13:00:00	2014-02-02 00:00:00	11.0
19	2926	2016-06-19 15:00:00	2016-06-19 22:00:00	7.0
20	2536	2014-11-06 14:00:00	2014-11-08 13:00:00	47.0
21	2099	2012-11-24 05:00:00	2012-11-25 05:00:00	24.0
22	1983	2012-06-29 20:00:00	2012-06-30 21:00:00	25.0
23	554	2006-04-19 06:00:00	2006-04-20 17:00:00	35.0
24	1080	2008-07-31 09:00:00	2008-08-01 12:00:00	27.0
25	1773	2011-08-07 13:00:00	2011-08-07 23:00:00	10.0
26	1106	2008-09-09 00:00:00	2008-09-09 18:00:00	18.0

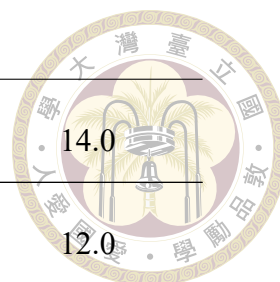


27	1513	2010-06-13 14:00:00	2010-06-14 02:00:00	12.0
28	1827	2011-11-01 01:00:00	2011-11-01 08:00:00	7.0
29	1547	2010-08-09 17:00:00	2010-08-10 07:00:00	14.0
30	1434	2010-01-16 17:00:00	2010-01-17 09:00:00	16.0
31	1287	2009-06-15 10:00:00	2009-06-15 16:00:00	6.0
32	2034	2012-09-10 03:00:00	2012-09-11 03:00:00	24.0
33	2003	2012-07-30 09:00:00	2012-07-30 13:00:00	4.0
34	2235	2013-07-24 22:00:00	2013-07-25 05:00:00	7.0
35	2918	2016-06-08 14:00:00	2016-06-08 17:00:00	3.0
36	870	2007-09-23 15:00:00	2007-09-24 07:00:00	16.0
37	2475	2014-08-01 13:00:00	2014-08-02 15:00:00	26.0
38	1253	2009-04-28 13:00:00	2009-04-29 01:00:00	12.0
39	493	2006-01-16 03:00:00	2006-01-16 23:00:00	20.0
40	2454	2014-06-09 19:00:00	2014-06-10 00:00:00	5.0
41	2919	2016-06-10 16:00:00	2016-06-10 21:00:00	5.0
42	2791	2015-12-03 12:00:00	2015-12-03 22:00:00	10.0
43	840	2007-07-19 19:00:00	2007-07-21 23:00:00	52.0

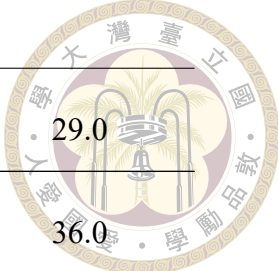


44	740	2007-01-09 10:00:00	2007-01-10 08:00:00	22.0
45	1097	2008-08-20 14:00:00	2008-08-20 21:00:00	7.0
46	1832	2011-11-04 12:00:00	2011-11-05 02:00:00	14.0
47	2122	2012-12-22 01:00:00	2012-12-23 05:00:00	28.0
48	644	2006-09-22 10:00:00	2006-09-22 15:00:00	5.0
49	1550	2010-08-13 08:00:00	2010-08-14 16:00:00	32.0
50	1732	2011-05-26 07:00:00	2011-05-26 14:00:00	7.0
51	1880	2012-01-03 01:00:00	2012-01-03 14:00:00	13.0
52	1103	2008-09-01 09:00:00	2008-09-03 02:00:00	41.0
53	931	2007-12-08 07:00:00	2007-12-08 20:00:00	13.0
54	2496	2014-08-29 10:00:00	2014-08-30 03:00:00	17.0
55	2370	2014-02-06 12:00:00	2014-02-07 08:00:00	20.0
56	3190	2017-07-11 04:00:00	2017-07-12 00:00:00	20.0
57	2790	2015-12-02 18:00:00	2015-12-03 05:00:00	11.0
58	1921	2012-03-04 04:00:00	2012-03-04 17:00:00	13.0
59	2753	2015-10-05 03:00:00	2015-10-06 18:00:00	39.0
60	2984	2016-09-16 07:00:00	2016-09-16 09:00:00	2.0



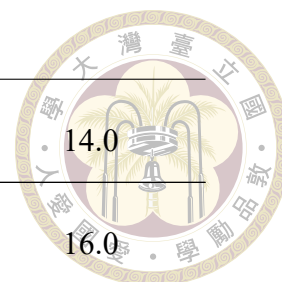


61	2994	2016-10-01 07:00:00	2016-10-01 21:00:00	14.0
62	2025	2012-08-24 14:00:00	2012-08-25 02:00:00	12.0
63	1280	2009-06-05 13:00:00	2009-06-07 15:00:00	50.0
64	1067	2008-07-04 23:00:00	2008-07-07 21:00:00	70.0
65	809	2007-05-30 06:00:00	2007-05-30 15:00:00	9.0
66	376	2005-07-04 03:00:00	2005-07-04 20:00:00	17.0
67	1252	2009-04-27 03:00:00	2009-04-28 01:00:00	22.0
68	1078	2008-07-29 10:00:00	2008-07-29 18:00:00	8.0
69	1510	2010-06-07 18:00:00	2010-06-09 08:00:00	38.0
70	2022	2012-08-21 16:00:00	2012-08-21 17:00:00	1.0
71	2739	2015-08-31 01:00:00	2015-09-02 00:00:00	47.0
72	2889	2016-04-13 20:00:00	2016-04-15 00:00:00	28.0
73	2868	2016-03-09 00:00:00	2016-03-09 15:00:00	15.0
74	1267	2009-05-14 20:00:00	2009-05-16 21:00:00	49.0
75	1556	2010-08-22 22:00:00	2010-08-23 16:00:00	18.0
76	795	2007-05-08 07:00:00	2007-05-08 09:00:00	2.0
77	833	2007-07-09 09:00:00	2007-07-10 00:00:00	15.0

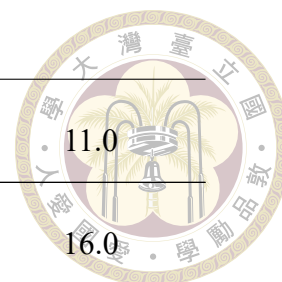


78	1144	2008-11-01 16:00:00	2008-11-02 21:00:00	29.0
79	826	2007-06-30 01:00:00	2007-07-01 13:00:00	36.0
80	696	2006-11-24 08:00:00	2006-11-25 10:00:00	26.0
81	646	2006-09-24 07:00:00	2006-09-24 08:00:00	1.0
82	2846	2016-02-05 20:00:00	2016-02-08 17:00:00	69.0
83	1996	2012-07-16 03:00:00	2012-07-16 19:00:00	16.0
84	2449	2014-06-02 19:00:00	2014-06-03 05:00:00	10.0
85	2155	2013-02-05 08:00:00	2013-02-06 08:00:00	24.0
86	2699	2015-07-05 13:00:00	2015-07-05 18:00:00	5.0
87	573	2006-05-21 02:00:00	2006-05-22 21:00:00	43.0
88	2043	2012-09-23 11:00:00	2012-09-24 12:00:00	25.0
89	2698	2015-07-03 23:00:00	2015-07-04 01:00:00	2.0
90	2095	2012-11-19 13:00:00	2012-11-22 04:00:00	63.0
91	1109	2008-09-12 08:00:00	2008-09-12 21:00:00	13.0
92	1120	2008-10-04 09:00:00	2008-10-05 11:00:00	26.0
93	2445	2014-05-23 10:00:00	2014-05-24 18:00:00	32.0
94	3030	2016-11-21 02:00:00	2016-11-22 09:00:00	31.0

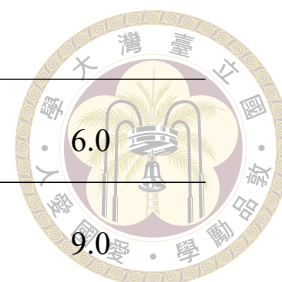
95	1986	2012-07-06 03:00:00	2012-07-06 17:00:00	14.0
96	424	2005-10-12 09:00:00	2005-10-13 01:00:00	16.0
97	2241	2013-08-02 06:00:00	2013-08-03 03:00:00	21.0
98	892	2007-10-28 07:00:00	2007-10-28 13:00:00	6.0
99	869	2007-09-21 14:00:00	2007-09-22 12:00:00	22.0
100	377	2005-07-05 04:00:00	2005-07-06 05:00:00	25.0
101	712	2006-12-11 02:00:00	2006-12-11 10:00:00	8.0
102	1342	2009-09-02 03:00:00	2009-09-03 04:00:00	25.0
103	2281	2013-10-11 02:00:00	2013-10-14 17:00:00	87.0
104	2378	2014-02-14 09:00:00	2014-02-15 17:00:00	32.0
105	3028	2016-11-19 21:00:00	2016-11-20 08:00:00	11.0
106	958	2008-01-13 18:00:00	2008-01-16 04:00:00	58.0
107	2288	2013-10-19 15:00:00	2013-10-23 08:00:00	89.0
108	2482	2014-08-11 13:00:00	2014-08-11 17:00:00	4.0
109	1493	2010-04-29 11:00:00	2010-05-01 02:00:00	39.0
110	1177	2008-12-12 18:00:00	2008-12-14 03:00:00	33.0
111	628	2006-08-25 15:00:00	2006-08-26 05:00:00	14.0



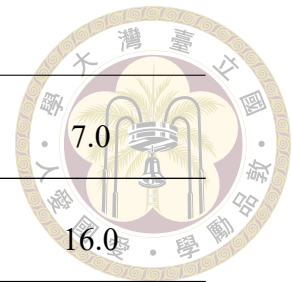
112	1065	2008-07-01 17:00:00	2008-07-02 04:00:00	11.0
113	641	2006-09-13 20:00:00	2006-09-14 12:00:00	16.0
114	871	2007-09-24 19:00:00	2007-09-24 21:00:00	2.0
115	1322	2009-07-31 17:00:00	2009-08-01 14:00:00	21.0
116	1511	2010-06-09 15:00:00	2010-06-09 22:00:00	7.0
117	1296	2009-06-30 03:00:00	2009-06-30 13:00:00	10.0
118	442	2005-11-08 12:00:00	2005-11-08 21:00:00	9.0
119	2481	2014-08-09 21:00:00	2014-08-11 01:00:00	28.0
120	2922	2016-06-12 07:00:00	2016-06-12 15:00:00	8.0
121	2234	2013-07-23 01:00:00	2013-07-23 08:00:00	7.0
122	2983	2016-09-13 12:00:00	2016-09-13 17:00:00	5.0
123	2450	2014-06-03 14:00:00	2014-06-04 21:00:00	31.0
124	1981	2012-06-28 00:00:00	2012-06-28 12:00:00	12.0
125	1994	2012-07-12 15:00:00	2012-07-14 06:00:00	39.0
126	2916	2016-06-01 07:00:00	2016-06-01 08:00:00	1.0
127	2154	2013-02-04 21:00:00	2013-02-05 01:00:00	4.0
128	1970	2012-06-06 10:00:00	2012-06-09 00:00:00	62.0



129	2453	2014-06-09 01:00:00	2014-06-09 07:00:00	6.0
130	810	2007-05-31 10:00:00	2007-05-31 19:00:00	9.0
131	1316	2009-07-23 03:00:00	2009-07-23 17:00:00	14.0
132	1291	2009-06-26 07:00:00	2009-06-27 02:00:00	19.0
133	1077	2008-07-28 15:00:00	2008-07-29 02:00:00	11.0
134	1151	2008-11-08 04:00:00	2008-11-09 23:00:00	43.0
135	1052	2008-05-26 06:00:00	2008-05-30 09:00:00	99.0
136	1284	2009-06-10 09:00:00	2009-06-10 23:00:00	14.0
137	853	2007-08-14 00:00:00	2007-08-15 14:00:00	38.0
138	2238	2013-07-29 11:00:00	2013-07-29 16:00:00	5.0
139	2100	2012-11-25 15:00:00	2012-11-26 01:00:00	10.0
140	827	2007-07-02 01:00:00	2007-07-03 12:00:00	35.0
141	821	2007-06-22 04:00:00	2007-06-23 01:00:00	21.0
142	1295	2009-06-29 15:00:00	2009-06-29 18:00:00	3.0
143	1081	2008-08-01 20:00:00	2008-08-02 10:00:00	14.0
144	2425	2014-04-25 09:00:00	2014-04-26 15:00:00	30.0
145	1494	2010-05-01 17:00:00	2010-05-02 05:00:00	12.0



146	2927	2016-06-20 05:00:00	2016-06-20 12:00:00	7.0
147	2471	2014-07-18 23:00:00	2014-07-19 15:00:00	16.0
148	2924	2016-06-14 11:00:00	2016-06-17 15:00:00	76.0
149	1932	2012-04-03 11:00:00	2012-04-05 00:00:00	37.0
150	2016	2012-08-15 08:00:00	2012-08-15 15:00:00	7.0
151	391	2005-08-13 07:00:00	2005-08-13 19:00:00	12.0
152	686	2006-11-17 01:00:00	2006-11-17 15:00:00	14.0
153	2224	2013-06-20 17:00:00	2013-06-20 20:00:00	3.0
154	3161	2017-05-17 02:00:00	2017-05-17 17:00:00	15.0
155	479	2005-12-30 02:00:00	2005-12-30 17:00:00	15.0
156	909	2007-11-18 01:00:00	2007-11-19 14:00:00	37.0
157	1090	2008-08-12 01:00:00	2008-08-14 07:00:00	54.0





Appendix B — Model Optimization

Process and Accuracy Comparison

The dataset used in these experiments focuses on three meteorological variables: $d2m$, u , and v . Data samples were taken at a 3-hour temporal resolution over the period 2001 – 2020. To accelerate the experiments, a smaller data volume and fewer variables were intentionally selected. Each experiment was conducted over 60 epochs to evaluate performance and refine the architecture and preprocessing steps.

B.1 Model Architectures

The first optimization step explored the impact of removing MaxPooling layers from the network structure. As shown in Figure B.1, eliminating MaxPooling layers led to a lower validation loss and a higher validation accuracy compared to the original model. While the results of different removal configurations were similar, the configuration without MaxPooling exhibited more stable performance throughout the training process. This

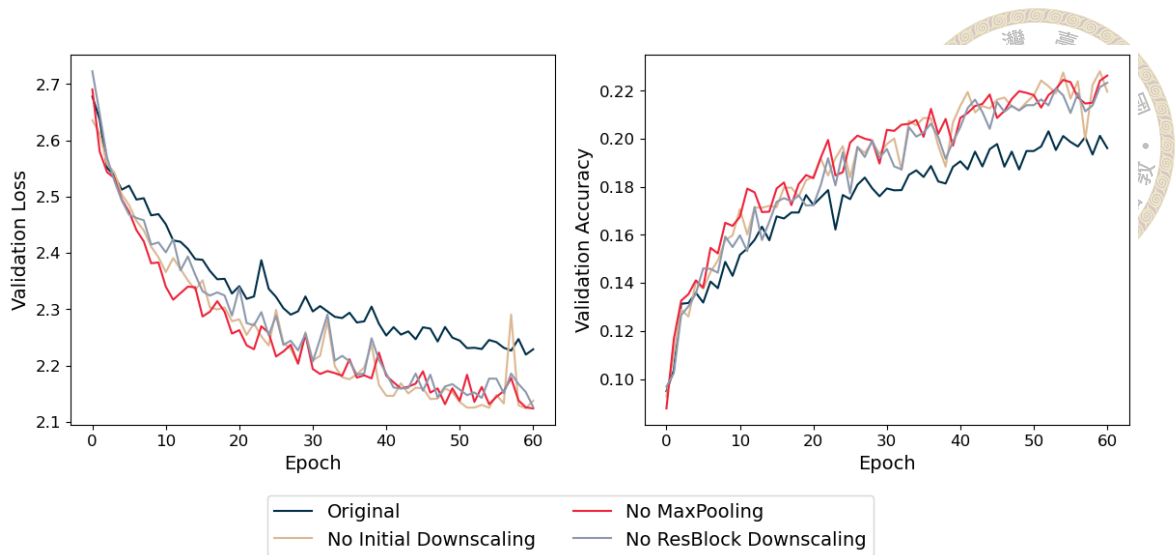


Figure B.1 Validation Loss and Accuracy Comparisons for Different Structural Modifications

result suggests that removing MaxPooling preserved spatial details, which are critical for capturing complex patterns in meteorological data. Based on these findings, the decision was made to remove MaxPooling layers in the subsequent model design.

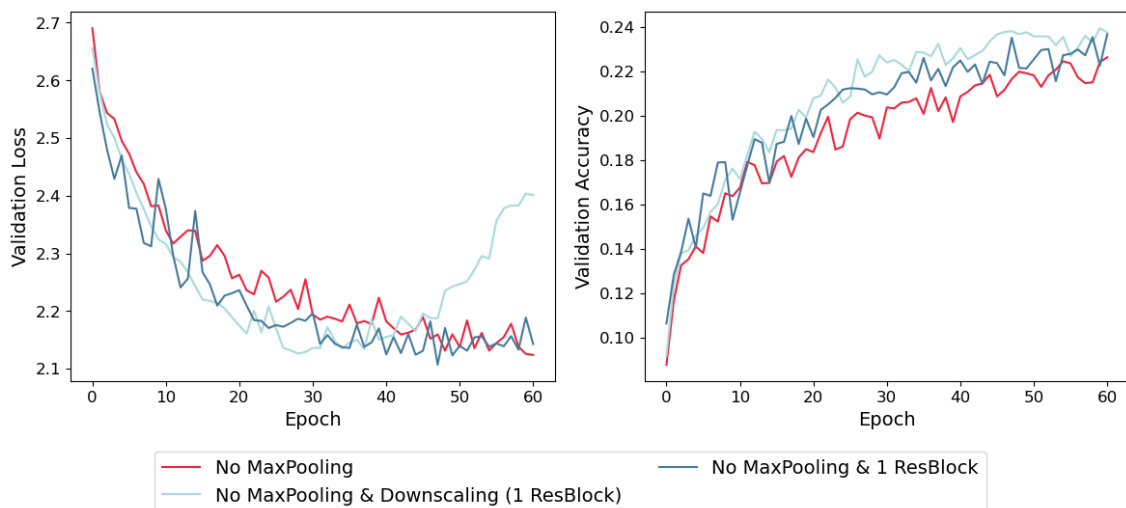


Figure B.2 Validation Loss and Accuracy Comparisons for MaxPooling and ResBlock Modifications

Following the removal of MaxPooling layers, the next step tested the effect of re-

moving downscaling within one ResBlock. Figure B.2 shows that this additional change further reduced validation loss and improved validation accuracy, especially in the later epochs.



The two configurations differ in that one completely removes a ResBlock, while the other only removes the downscaling operation within a ResBlock.

The results indicate that limiting downscaling enhances the model's ability to retain spatial information while still learning hierarchical features. Consequently, the final structure excludes both MaxPooling and one ResBlock downscaling.

B.2 Different Data Preprocessing Methods

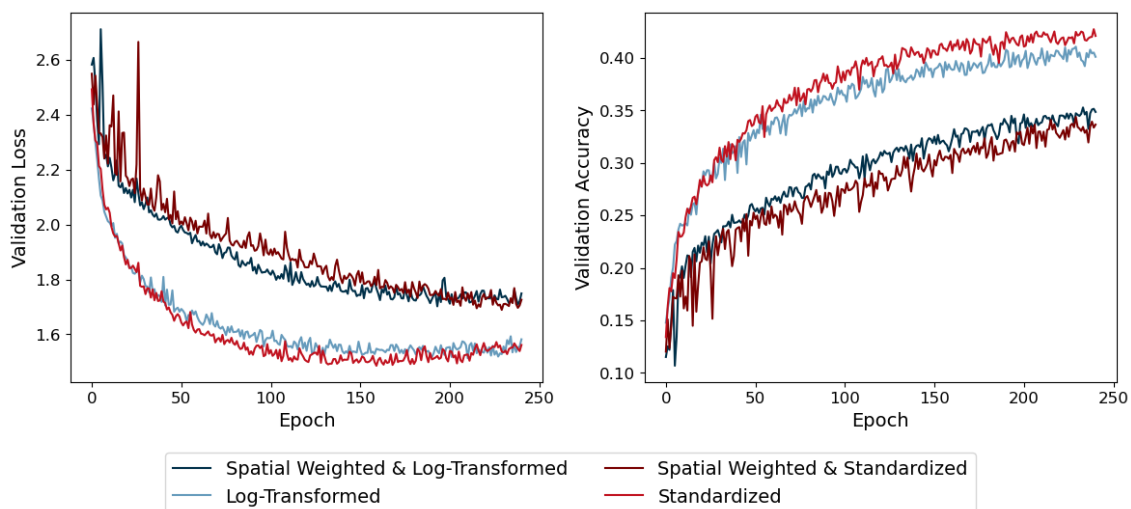


Figure B.3 Validation Loss and Accuracy Comparisons for Different Data Preprocessing Methods

In the data preprocessing stage, the effects of log transformation and standardization were tested with and without spatial weighting. As shown in Figure B.3, applying log transformation with spatial weighting provided better performance in terms of validation accuracy and loss reduction compared to other configurations. However, for non-weighted data, standardization performed better than log transformation. Based on these results, the preprocessing method was finalized as:

- **Spatial weighted data:** Log transformation.
- **Non-weighted data:** Standardization.

B.3 Effect of Temporal Resolution

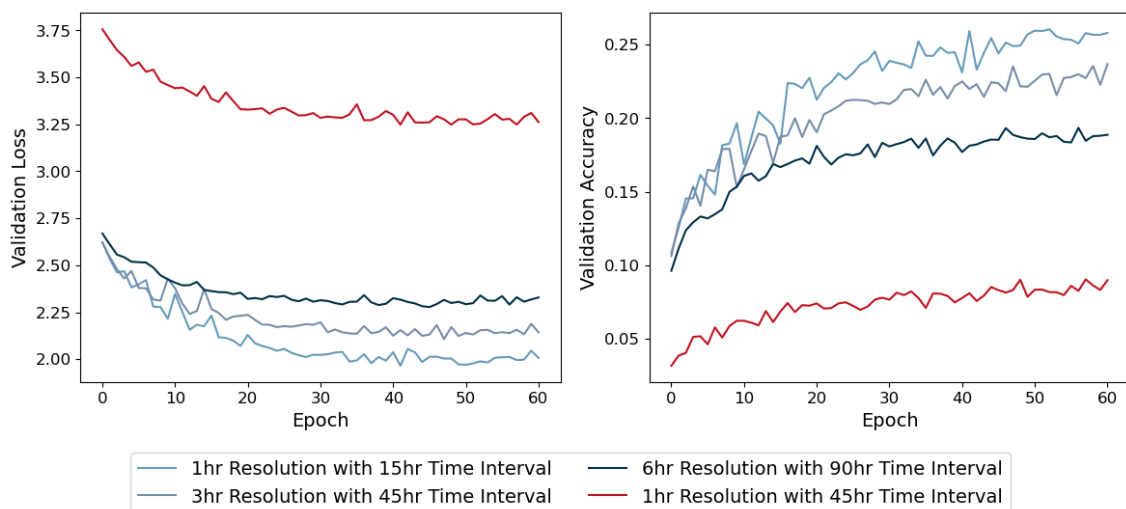


Figure B.4 Validation Loss and Accuracy Comparisons for Different Temporal Resolutions and Time Intervals

The final optimization focused on temporal resolution and sampling intervals. Figure B.4 compares different configurations:



- **1-hour resolution with 15-hour intervals**
- **3-hour resolution with 45-hour intervals**
- **6-hour resolution with 90-hour intervals**
- **1-hour resolution with 45-hour intervals**

The results indicate that a 3-hour resolution with 45-hour intervals achieved the lowest validation loss and highest accuracy among the tested configurations. This setting provides a balance between maintaining sufficient time intervals to capture temporal patterns and achieving higher accuracy. It also effectively balances temporal detail and computational efficiency, making it the final choice for the model configuration.







Appendix C — Evaluation of Different Weather Analog Methods

C.1 Single-Event SSR Evaluation Across Variables for Multiple Methods

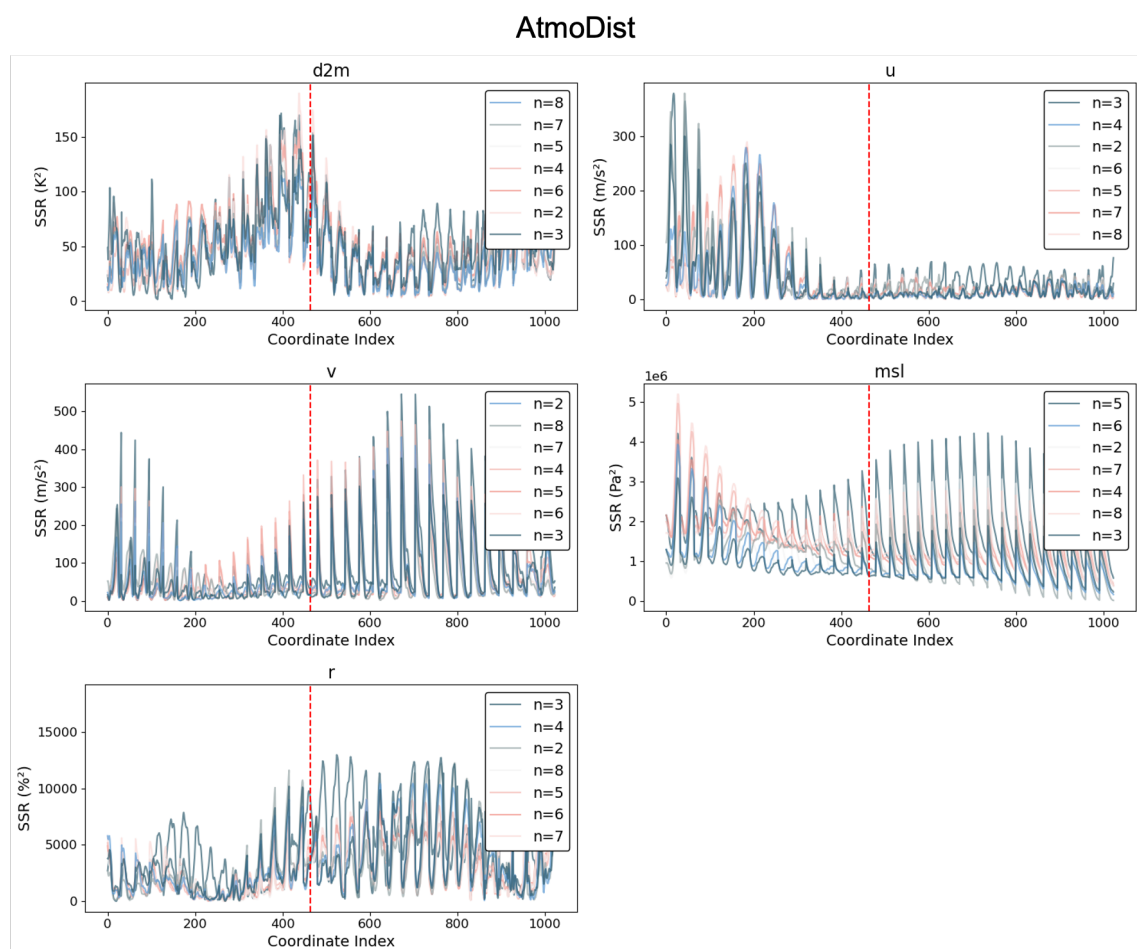


Figure C.5 Sum of Squared Residuals Analysis for Five Atmospheric Variables across Different Analogues Based on AtmoDist Results

Euclidean

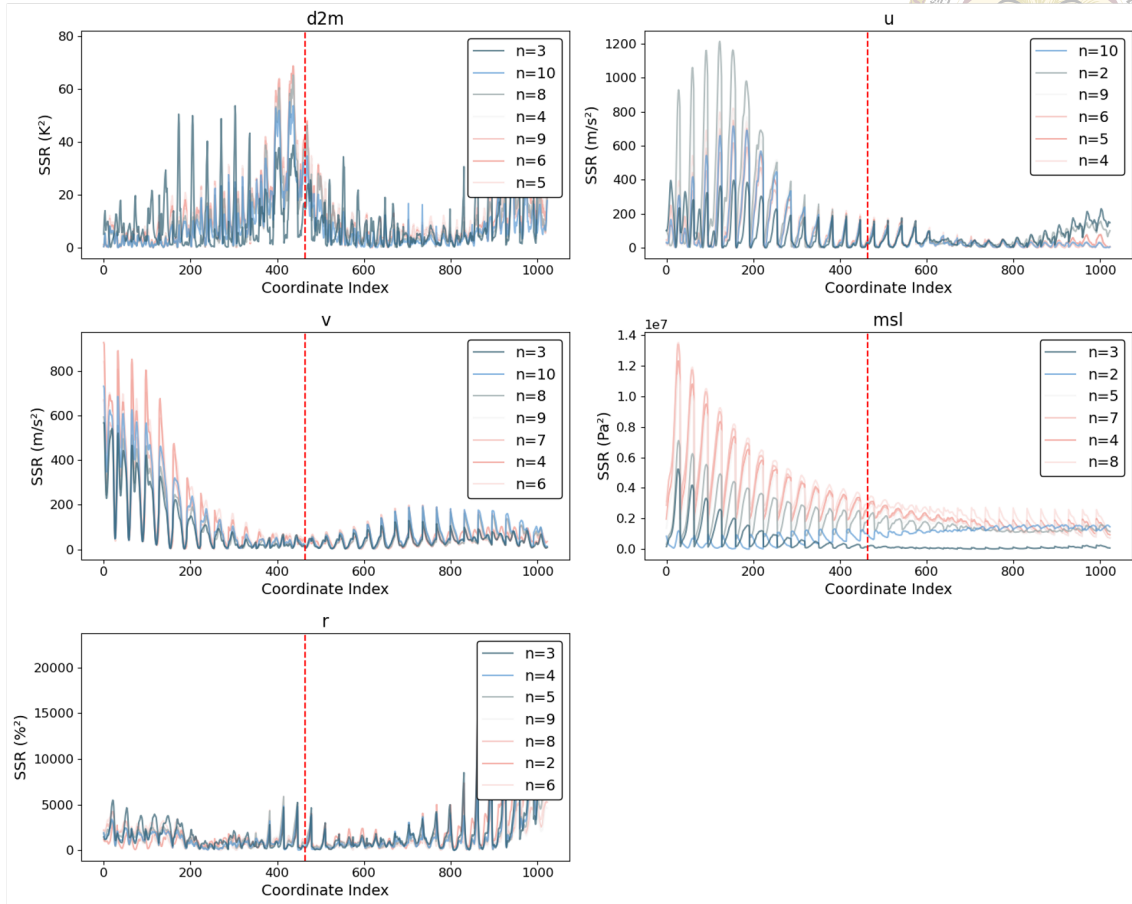


Figure C.6 Sum of Squared Residuals Analysis for Five Atmospheric Variables across Different Analogues Based on Euclidean Results

Cosine

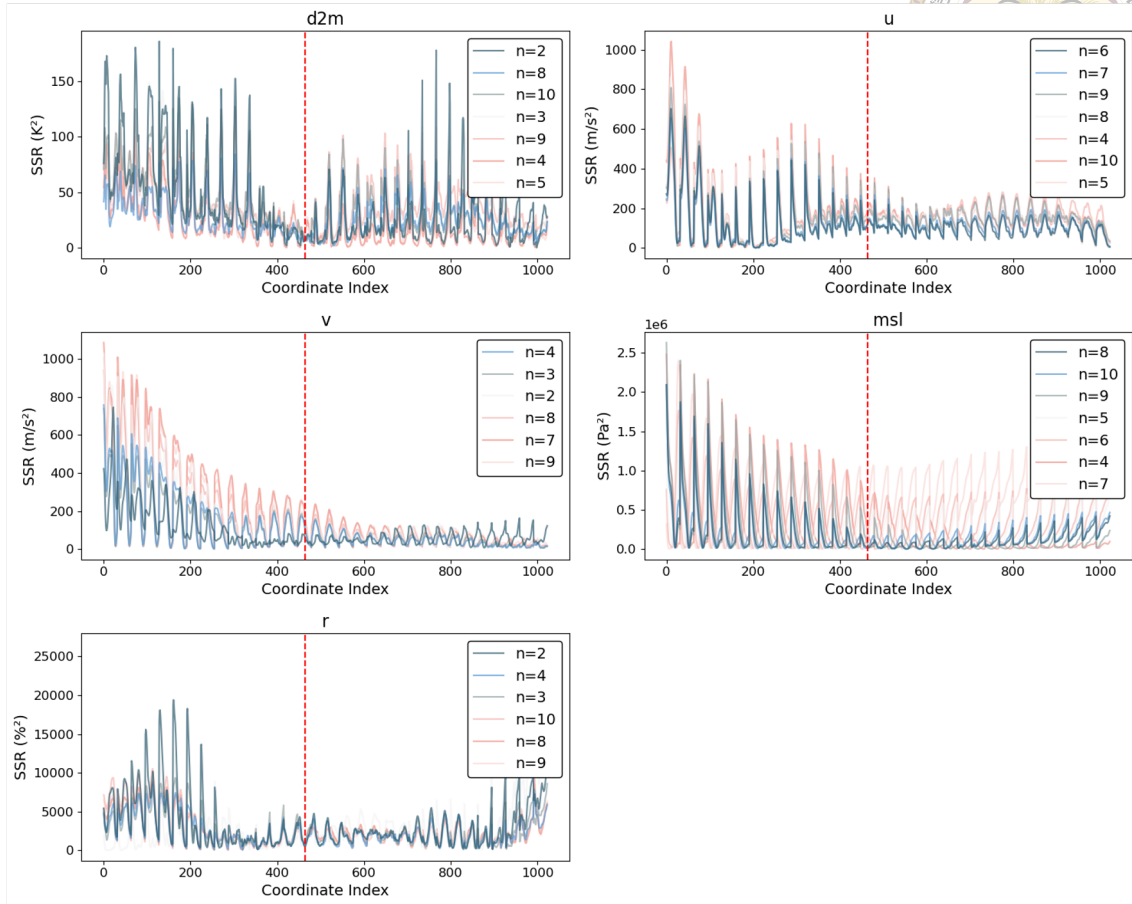


Figure C.7 Sum of Squared Residuals Analysis for Five Atmospheric Variables across Different Analogues Based on Cosine Results

SSIM

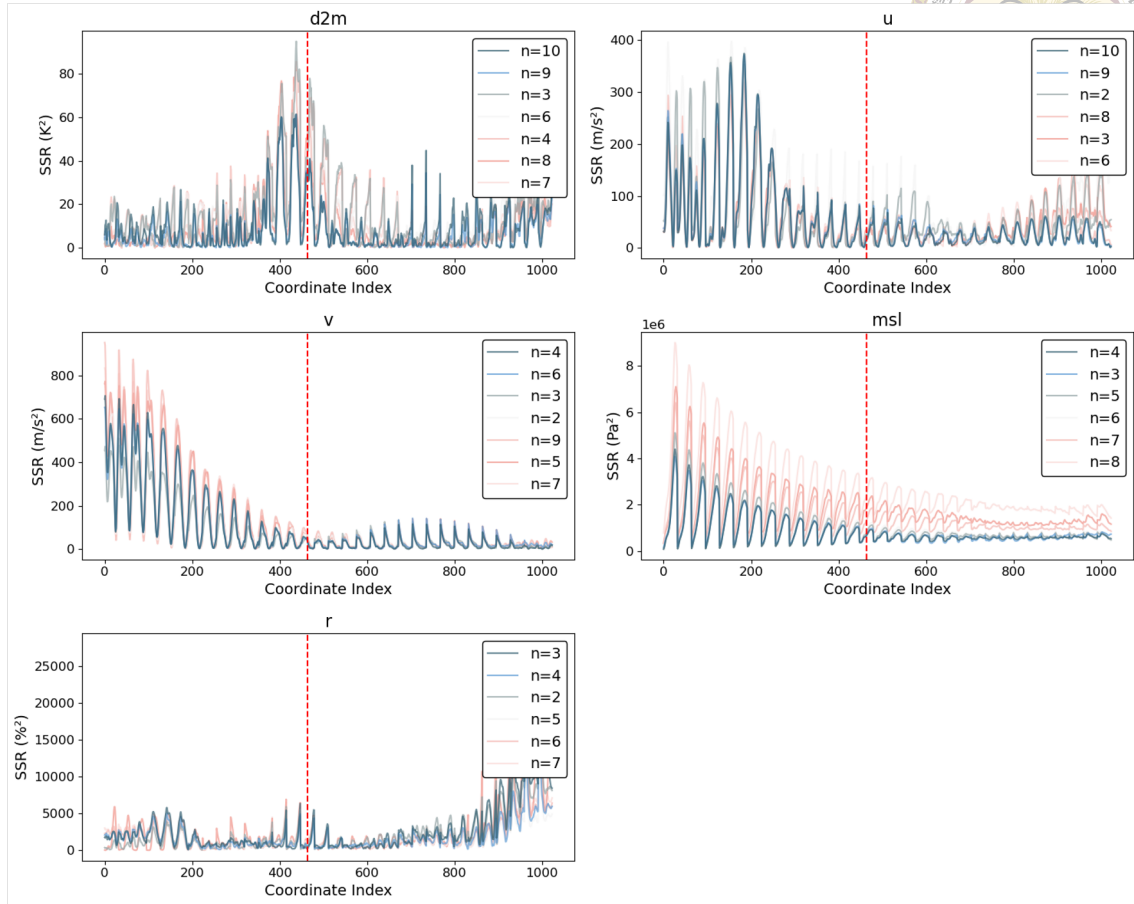


Figure C.8 Sum of Squared Residuals Analysis for Five Atmospheric Variables across Different Analogues Based on SSIM Results