




Article

A Strategy to Optimize the Implementation of a Machine-Learning Scheme for Extreme Meiyu Rainfall Prediction over Southern Taiwan

Jung-Lien Chu ¹, Chou-Chun Chiang ¹, Li-Huan Hsu ¹, Li-Rung Hwang ¹, Yi-Chiang Yu ^{1,*}, Kuan-Ling Lin ¹, Chieh-Ju Wang ¹, Shih-Hao Su ² and Ting-Shuo Yo ³

- ¹ National Science and Technology Center for Disaster Reduction, New Taipei 23143, Taiwan; jlchu@ncdr.nat.gov.tw (J.-L.C.); cgjou@ncdr.nat.gov.tw (C.-C.C.); lhhsu@ncdr.nat.gov.tw (L.-H.H.); sharonny@ncdr.nat.gov.tw (L.-R.H.); aling119@ncdr.nat.gov.tw (K.-L.L.); julie@ncdr.nat.gov.tw (C.-J.W.)
- ² Department of Atmospheric Science, Chinese Culture University, Taipei 11114, Taiwan; ssh3@g.pccu.edu.tw
- ³ Department of Atmospheric Science, National Taiwan University, Taipei 10617, Taiwan; tingshuo.yo@gmail.com
- * Correspondence: yuyc@ncdr.nat.gov.tw

Abstract: This study aims to propose a strategy to optimize the performance of the Support Vector Machine (SVM) scheme for extreme Meiyu rainfall prediction over southern Taiwan. Variables derived from Climate Forecast System Reanalysis (CFSR) dataset are the candidates for predictor selection. A series of experiments with different combinations of predictors and domains are designed to obtain the optimal strategy for constructing the SVM scheme. The results reveal that the accuracy (ACC), positive predictive values (PPV), probability of detection (POD), and F1-score can exceed 0.6 on average. Choosing the predictors associated with the Meiyu system and determine the domain associated with the correlations between selected predictors and predictand can improve the forecast performance. Our strategy shows the potential to predict extreme Meiyu rainfall in southern Taiwan with lead times from 16 h to 64 h. The F1-score analysis further demonstrates that the forecast performance of our scheme is stable, with slight inter-annual fluctuations from 1990 to 2019. Higher performance would be expected when the north of the South China Sea is characterized by stronger southwesterly flow and abundant low-level moisture for a given year.

Keywords: support vector machine; machine learning; extreme Meiyu rainfall



Citation: Chu, J.-L.; Chiang, C.-C.; Hsu, L.-H.; Hwang, L.-R.; Yu, Y.-C.; Lin, K.-L.; Wang, C.-J.; Su, S.-H.; Yo, T.-S. A Strategy to Optimize the Implementation of a Machine-Learning Scheme for Extreme Meiyu Rainfall Prediction over Southern Taiwan. *Water* **2021**, *13*, 2884. <https://doi.org/10.3390/w13202884>

Academic Editors: Momcilo Markus and Renato Morbidelli

Received: 2 September 2021

Accepted: 12 October 2021

Published: 14 October 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Taiwan is a mountainous island located in the East Asia Summer Monsoon (EASM) region. The north-south elongated Central Mountain Range (CMR) in Taiwan is about 2000 m in height on average. The rainfall of Taiwan ties closely to its topography and the complex spatial-temporal variability of EASM. The rainfall in Taiwan is mainly contributed by the Meiyu fronts and typhoons in boreal summer [1]. During the Meiyu season, May and June, the EASM region is often characterized by a quasi-stationary front (Meiyu front) and its associated rain belt, which is elongated northeast to southwest from the Sea of Japan to the Bashi Channel and the northern South China Sea [2]. The Meiyu regime may then be found over the Yangtze River in the period of mid-June to mid-July [3]. When the Meiyu front moves to Taiwan, torrential rainfall will happen and may cause many disastrous events. For example, the heavy rainfall of the Meiyu front on 12 June 2005, resulted in large area inundation and landslide in western Taiwan. The total amount of rainfall in four days reached 1645 mm in Majia Township over the region of southwestern Taiwan. The damage loss is about USD 160 million from 12 June to 15 June in Taiwan [4].

While the Meiyu front is located to the north of Taiwan, there is a low-pressure area to the north of Taiwan and a high-pressure area to the south of Taiwan, corresponding to a large northward pressure gradient force. This pressure pattern is favorable to the formation

of southwesterly flow [5], also known as low-level jet [6,7]. The heavy rainfall in southern Taiwan is found to be correlated with the strong (e.g., 12 m s^{-1}) southwesterly flow at low levels (e.g., 850 hPa) that occurred in southern Taiwan. When the axis of the southwesterly flow moves toward southern Taiwan, the moist air is transported from the northern South China Sea to southern Taiwan and provides a favorable environment for continuous rain and heavy rainfall [5,8,9].

Recently, artificial intelligence (AI) techniques are broadly used in many disciplines. To provide valuable information to decision-makers, AI and related data science methods generally work with big data in different disciplines [10]. In the past decades, the resolution of model outputs and the amount of observed data has become finer and denser. High amounts of high-quality data make it possible to apply AI techniques in the atmospheric science discipline, especially on those types of high-impact weather. Although the dynamical model can capture the characteristics of extreme regional events, the high costs of computation would be the weakness. Owing to the lower requirement of computation, the AI techniques give another choice in the prediction of high-impact weather while the computational cost is still high during the training period. The AI schemes that are generally used on high-impact weather include traditional model output statistics (MOS), artificial neural network (ANN), decision trees methods, support vector machine (SVM), convolutional neural network (CNN), etc. [11,12]. An objective system was conducted to predict synoptic-scale fronts by means of the CNN scheme [13]. The system performs better than the numerical frontal analysis in front detection. In addition, CNN was used to estimate the intensity of tropical cyclone (TC) based on satellite images [12]. The results reveal high-quality estimations on TC intensity as the operational forecasts. The AI is also applied to the study of air pollution using the support vector machine (SVM) method [14]. Furthermore, the SVM method was also used to identify synoptic weather types [15]. Their results showed that the SVM method outperformed methods based on the traditional objective diagnosis. The equitable threat Score (ETS) in their study can reach 0.33 in frontal system identification, which is higher than the ETS that using the method based on thermal front parameters [16].

For the simulation of Meiyu rainfall, some problems have been solved in recent decades. The overall amount and general spatial distribution of Meiyu rainfall over Taiwan can be well performed by the WRF Model [17]. Moreover, the phase-locked topographic extreme Meiyu rainfall can also be captured to a reasonable extent by a high-resolution cloud-resolving model, the cloud-resolving storm simulator (CReSS). Even so, poor forecast performance is still found in migratory events [18]. Although the dynamical model can capture the characteristics of extreme events, a finer resolution is required (the horizontal resolution of a cloud-resolving model is usually less than 5 km) to enhance the forecast performance. The higher the resolution of the model, the heavier the loading of computational cost. The AI techniques provide another way to look at this problem. The focus of the study is to propose an optimal strategy in predicting the regional extreme rainfall events using AI techniques. Therefore, the SVM-based schemes for daily rainfall will be developed over the region of southern Taiwan during the Meiyu season. The performance and associated skill scores for these schemes are evaluated through a series of experiments. The remaining parts of the article are as follows. Section 2 describes the data, the SVM method, and the design of the experiments. Section 3 demonstrates the analysis results and the performance of these schemes. The conclusion and discussion are given in Section 4.

2. Data and Methodology

2.1. Predictand and Predictors

The station's daily rainfall data from Central Weather Bureau Taiwan are used as the predictand over the region of southern Taiwan to conduct the SVM-based prediction schemes (Figure 1). The 32-year (1988–2020) daily rainfall data for May and June in southwestern Taiwan is ranked. Generally, an extreme weather event can be defined as a case that is as rare as or rarer than the 10th or 90th percentile of a probability density

function estimated from observations [19]. Following the definition, the 90th percentile of daily rainfall is 35 mm (including zero events) and is taken as the threshold to define extreme rainfall events. The occurrence frequency of daily rainfall over 35 mm for all stations for 32 years is calculated. Moreover, if the frequency of extreme rainfall occurrence for a given station is lower than 10%, then most of the event log may be labeled to 0. The constructed model may perform skillfully by only predicting no extreme rainfall event for most of the time. To construct a well-trained model, stations for which the frequency of extreme rainfall occurrence is greater than 10% are selected. The stations for that the extreme rainfall occurrence frequency is greater than 10% are mostly located in the mountain region in southern Taiwan (Figure 1). The extreme rainfall events are then labeled in daily records if any of these stations' daily rainfall is greater than 35 mm. On the other hand, predictors are adopted from the Climate Forecast System reanalysis (CFSR) [20]. The CFSR data provides high-resolution temporal (6-hourly) and spatial ($0.5^\circ \times 0.5^\circ$) information of weather systems to the corresponded period for the station's daily rainfall data.

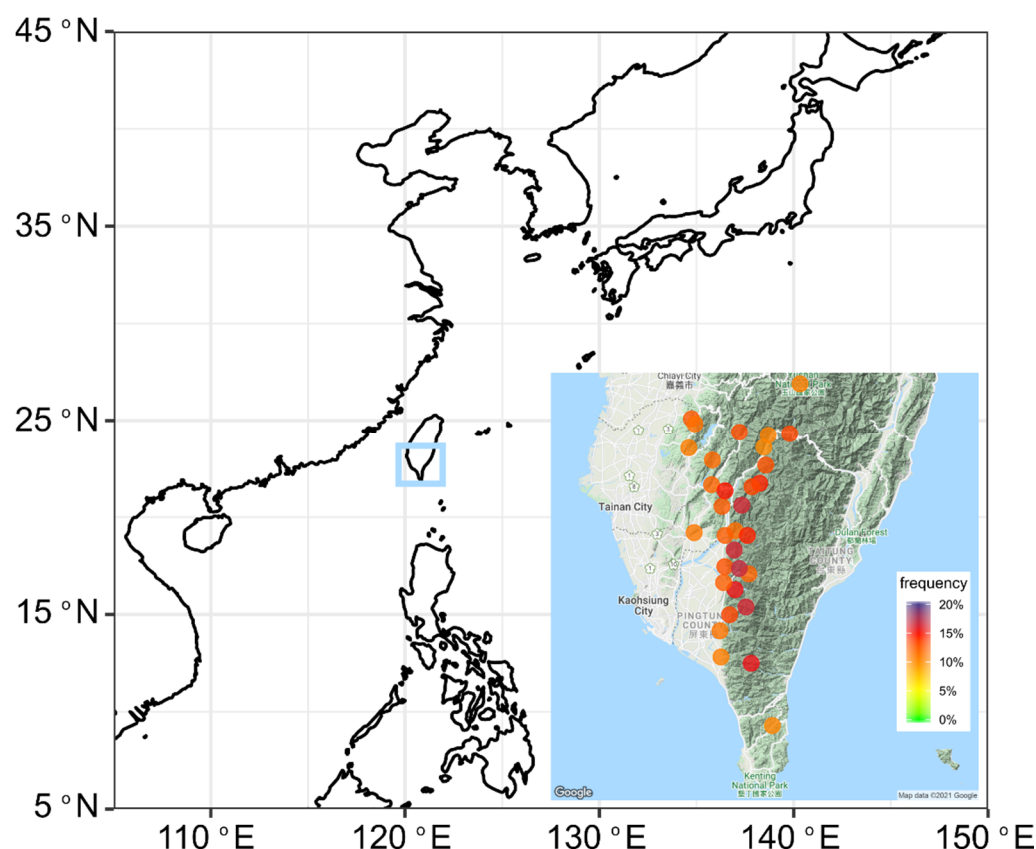


Figure 1. Locations of the CWB stations in southern Taiwan that used as the predictand to conduct the SVM-based prediction schemes. The stations are chosen according to the occurrence frequency that is greater than 10% for the daily rainfall over 35 mm for 32 years.

2.2. The SVM-Based Prediction Scheme

The SVM is one of the best algorithms for the classification of real-world data [21]. SVM has the potential to classify data clearly in high-dimensional space. For the extreme event prediction, the SVM classifier is applied to build the connection between the daily rainfall and the CFSR variables with a shift of time. Figure 2 demonstrates the process of the prediction scheme. First, the CFSR data is used as the predictors to construct the scheme. For the purpose of dimension reduction, the randomized principal component analysis (PCA) is applied to the CFSR variables [22,23]. PCA has been widely used in the environmental science field for feature extraction from a big dataset. Mathematically, the

k th principal feature (i.e., mode) of a given time series of variable X with $m \times n$ points can be expressed in the following equation.

$$P_k(t) = \sum_{i=1}^m \sum_{j=1}^n E_{i,j,k} X_{i,j}(t), \quad (1)$$

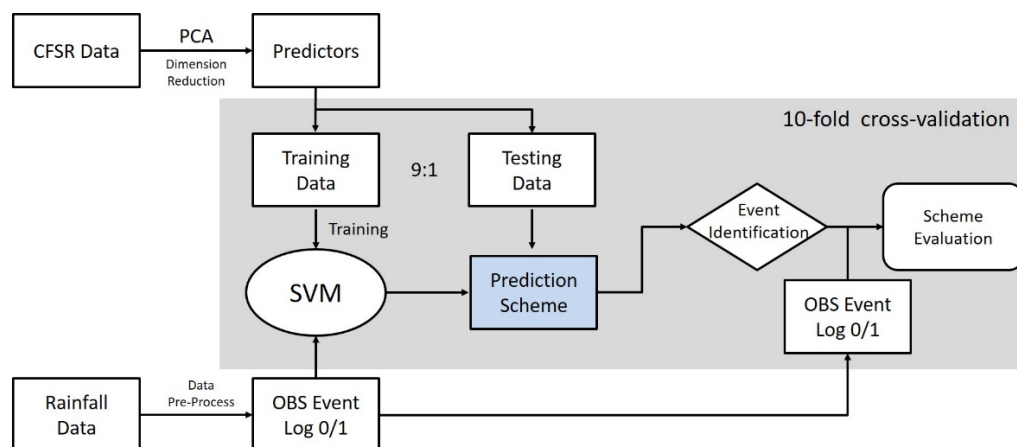


Figure 2. The schematic diagram for constructing the SVM-based prediction scheme.

Here, the P_k and E_k are the principal components (PC) and the eigenvector in the k th mode, respectively. In general, the modes would be ranked in descending order according to the portion of explained variances for the time series. The eigenvector E represents the spatial projection of the correspondence PC. The strength of this spatial pattern at a given time t is controlled by the magnitude of $P(t)$. By retaining the leading modes, the dimension of the variable X could be reduced. In this study, a selected number of PCs for CFSR variables will be substituted into the scheme as predictors.

Secondly, the labeled rainfall data (predictand) and those selected predictors will then be included to construct the prediction scheme using the SVM algorithm. When training the scheme, the polynomial kernel is used in this study as the kernel function of SVM to produce a hyperplane that can clearly separate groups of heavy rainfall and non-heavy rainfall events [24]. Meanwhile, the method of 10-fold cross-validation [25] with the same random seed is applied to all the constructed schemes for evaluation. The ratio between the amount for training data and the testing data is 9:1 at each time when the cross-validation is made. After the cross-validation is repeated 10 times, the total size of the testing data is equal to the total events. The testing data set will be used to evaluate the performance of the SVM-based prediction scheme in this study. Finally, a binary label for the extreme rainfall event will be given when the scheme is applied to a set of CFSR data that are excluded from the training period.

2.3. Experiments

2.3.1. Choice of Predictors

The determination of the choice of predictors and analysis domain play important roles in a successful prediction scheme. First, most of the variables in CFSR are considered as the predictors in our experiments, named group1. On the other hand, the selected variables that may cause extreme rainfall in southern Taiwan will be considered in the experiment of group2. The principles of selecting variables are based on a large number of studies about the Meiyu system and torrential rainfall in Taiwan during the Meiyu season. These Meiyu studies could be categorized into five topics: (1) the structure of front, (2) the frontal genesis, (3) the evolution of front, (4) the low-level Jet (LLJ), and (5) the interaction between the Meiyu system and topography [2]. The environment that favors the torrential Meiyu rainfall happening in southwestern Taiwan is characterized by the development

of meso-scale convective systems over the Southern Taiwan Strait within the warm and moist southwesterly monsoon flow. Meanwhile, the interaction between LLJs and Taiwan topography may enhance the heavy rainfall in the windward slopes of southwestern Taiwan. During the process of heavy rainfall development caused by the Meiyu system, the latent heat release plays an important role in maintaining the convective system and the propagation of the Meiyu regime [26]. Following the concepts of the aforementioned studies, we facilitate the choice of variables and focus on the low-level standard isobaric surface, such as 700 hPa, 850 hPa, 925 hPa, and 1000 hPa. For higher levels, we choose the 200 hPa and 400 hPa wind fields associated with upper-level jet and divergence fields. The associated predictors are chosen as group2 and are listed in Table 1.

Table 1. List of variables at selected levels that were used as predictors in group1 and group2 experiments.

Variables	Group1	Group2
Mean sea level pressure (MSLP)	Surface	Surface
Height (H)	100–950 hPa, interval 50 hPa	700 hPa, 850 hPa
Temperature (T)	100–1000 hPa, interval 50 hPa	700 hPa, 850 hPa, 925 hPa, 1000 hPa
Relative humidity (RH)	100–1000 hPa, interval 50 hPa	700 hPa, 850 hPa, 925 hPa, 1000 hPa
Zonal wind (U)	100–1000 hPa, interval 50 hPa	200 hPa, 400 hPa, 700 hPa, 850 hPa, 1000 hPa
Meridional wind (V)	100–1000 hPa, interval 50 hPa	200 hPa, 400 hPa, 700 hPa, 850 hPa, 1000 hPa
Vertical velocity (W)	100–1000 hPa, interval 50 hPa	

2.3.2. Domain Selection

In addition to the choice of predictors, we also investigate the method of domain selection. Generally, a selected domain should be large enough to describe the large-scale circulation which is relevant to the predictand [27]. Furthermore, the choice of analysis domain can also be defined by the boundary of zero correlation coefficient between predictors and the predictand [28]. In the present study, the two methods above are applied to the experiments. Domain-1 is the region from 90° E to 150° E and 10° S to 50° N, where the Meiyu-related rain belt associated with the revolution of the EASM is found [29]. Domain-2 will be defined according to the combination of the correlation coefficients among the main predictors and the rainfall in southern Taiwan between May and June. This method corresponds to the chosen predictors and will be further explained in Section 3. The performance of the schemes based on domain-1 and domain-2 will be examined in Section 3.

Several experiments are designed to obtain an optimal strategy for the scheme construction. Table 2 list the experiments and their descriptions. The scheme based on predictors of group1 and domain1 is defined as EXP-G1D1. The scheme constructed by domain1 using the variables of group2 as predictors (hereafter EXP-G2D1) will then be compared with EXP-G1D1. The scheme with higher skill between EXP-G1D1 and EXP-G2D1 will then be selected for further investigation on domain determination. EXP-G2D2 is chosen as the base experiment to test different forecast lead times further, named EXP-G2D2-L2, etc.

Table 2. List of experiments.

Experiments	Variable Type	Domain Type	Lead Time
EXP-G1D1	Group1	Domain-1	16 h
EXP-G2D1	Group2	Domain-1	16 h
EXP-G2D2	Group2	Domain-2	16 h
EXP-G2D2-L2	Group2	Domain-2	28 h
EXP-G2D2-L3	Group2	Domain-2	40 h
EXP-G2D2-L4	Group2	Domain-2	52 h
EXP-G2D2-L5	Group2	Domain-2	64 h
EXP-G2D2-L6	Group2	Domain-2	76 h

2.4. Evaluation Methods

The binary labels of extreme rainfall events can be determined by both the SVM prediction scheme and the corresponding observation pairs. The contingency table (Table 3) can be made to evaluate the performance of these experiments. The four elements in the contingency table, true positive, false positive, false negative, and true negative are denoted as A, B, C, and D, respectively.

Table 3. Contingency table for identifying the extreme rainfall events. Elements A, B, C, and D indicate true positive, false positive, false negative, and true negative, respectively.

		Observation	
		True	False
Predict	True	A (true positive)	B (false positive)
	False	C (false negative)	D (true negative)

Scores such as accuracy (ACC), positive predictive values (PPV), probability of detection (POD), and F1-score [30] can be calculated based on these four elements. These scores are generally used to compare the ability of constructed schemes in identifying extreme rainfall events. The formulas of these scores are explained as follows:

- ACC depicts the level of agreement between the result of identification and observation. The lower accuracy would be 0, and the higher accuracy would be 1.

$$ACC = (A + D) / (A + B + C + D), \quad (2)$$

- PPV demonstrates the ability of schemes in identifying cases of a true positive. The formula can be written as:

$$PPV = A / (A + B), \quad (3)$$

- POD is also known as the true positive rate. It measures the portion of hits that are correctly identified:

$$POD = A / (A + C), \quad (4)$$

- F1-score is a common measurement for anomaly detection. A higher weighting is given in the F1-score for true positive cases. The mathematic form of F1-score can be expressed as:

$$F1 = 2A / (2A + B + C), \quad (5)$$

3. Results

A preliminary examination is made based on EXP-G1D1 and EXP-G2D1 to decide the number of PCs used as predictors in the scheme. The PCs are first ranked in descending order according to the percentage of total explained variances. Figure 3 demonstrates the forecast scores of EXP-G1D1 and EXP-G2D1 for the leading 5 to the leading 100 PCs. The F1-score for EXP-G1D1 increases gradually from K5 to K45 and slightly decreases after K45. The forecast performs better when the scheme uses the leading 45 modes of PCs (K45). EXP-G2D1 also shows similar results. The highest score of PPV occurs at K70. However, the POD shows a decrease from K60 to K70 (figure not shown). This result implies that the model has too many missing events. In general, a well-trained model is able to increase the ratio of hits and reduce both the missing and the false alarms at the same time. When compared with PPV and POD, F1-score includes all three situations and is considered to be a more comprehensive score to evaluate the model performance. Therefore, we use the leading 45 modes of PCs for CFSR variables according to the F1-score and applied

them to be the predictors for constructing our scheme in all experiments. In addition, the F1-score analysis indicates that the performance of EXP-G2D1 is, in general, better than that of EXP-G1D1 from K5 to K100. The only difference between EXP-G1D1 and EXP-G2D1 is the choice of predictors. In other words, using the selected predictors associated with Meiyu rainfall in southern Taiwan in the scheme can improve the forecast performance.

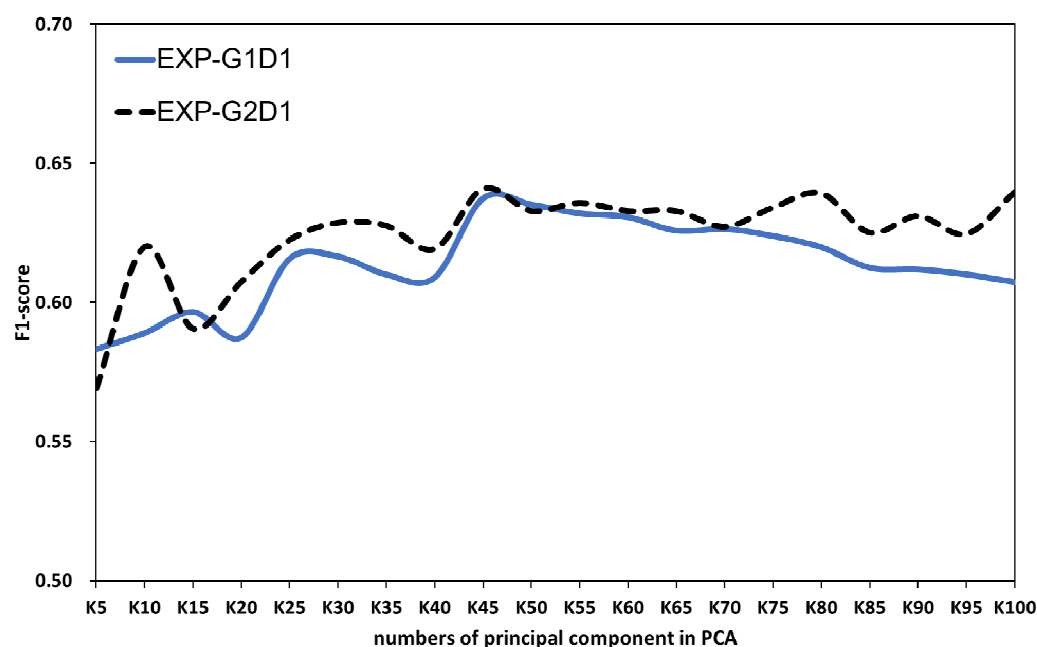


Figure 3. The F1-scores of EXP-G1D1 and EXP-G2D1 along with the numbers of PCs.

Domain selection may be another factor to influence the performance of schemes. The correlation maps between the predictors and the predictand could be one of the guidance to determine the domain for the constructed scheme [28]. Figure 4 shows the correlation maps among the main variables and Meiyu rainfall in southern Taiwan. These maps are characterized by a positive or negative center around Taiwan for most of the variables. The primary signal is more intensive in the vicinity of Taiwan, which implies that a smaller area can be defined as Domain-2. A contribution index (CI) obtained from the combination of the correlation maps proposed in Figure 4 is made to include the contribution of all the variables (Figure 5). The CI can be defined in each grid point as:

$$CI = \sum_{n=1}^{21} V_n |R_n|, \quad (6)$$

Here n is the numbers of used variables, and V_n is the percentage of total explained variances of the leading 45 PCs for each variable. $|R_n|$ is the absolute value of the correlation coefficient in Figure 4. The results indicate that a maximum center is discernible in the vicinity of Taiwan. Therefore, Domain 2 is defined as the region of 10° N– 35° N, 105° E– 135° E.

Forecast scores are compared between EXP-G2D1 and EXP-G2D2 to understand the performance of schemes based on different domains. Figure 6 depicts the performance of EXP-G2D1 and EXP-G2D2. As shown in Figure 6, the four kinds of forecast scores, F1-score, POD, PPV, and ACC for both EXP-G2D1 and EXP-G2D2, are all higher than 0.6. The ACC and PPV for EXP-G2D2 are higher than 0.7. Furthermore, all the forecast scores in EXP-G2D2 are higher than those in EXP-G2D1. Figure 7 gives the difference ratio of ACC, PPV, POD, and F1-score for EXP-G2D1 and EXP-G2D2 relative to the result of EXP-G1D1, respectively. The difference is statistically significant under the Student's t -test at a 95% significance level. The two-tailed p -value is 0.0378 for EXP-G2D1 and 0.0014 for

EXP-G2D2. Based on the result, EXP-G2D1 and EXP-G2D2 have better performance than that in EXP-G1D1, especially the PPV in EXP-G2D2. PPV improves nearly 5% relative to EXP-G1D1 and is about 4.5% higher than that of EXP-G2D1. The F1-score of EXP-G2D2 is nearly 3.5% higher than that of EXP-G2D1. The result reveals that higher forecast performance can be found when the scheme is constructed by Domain2.

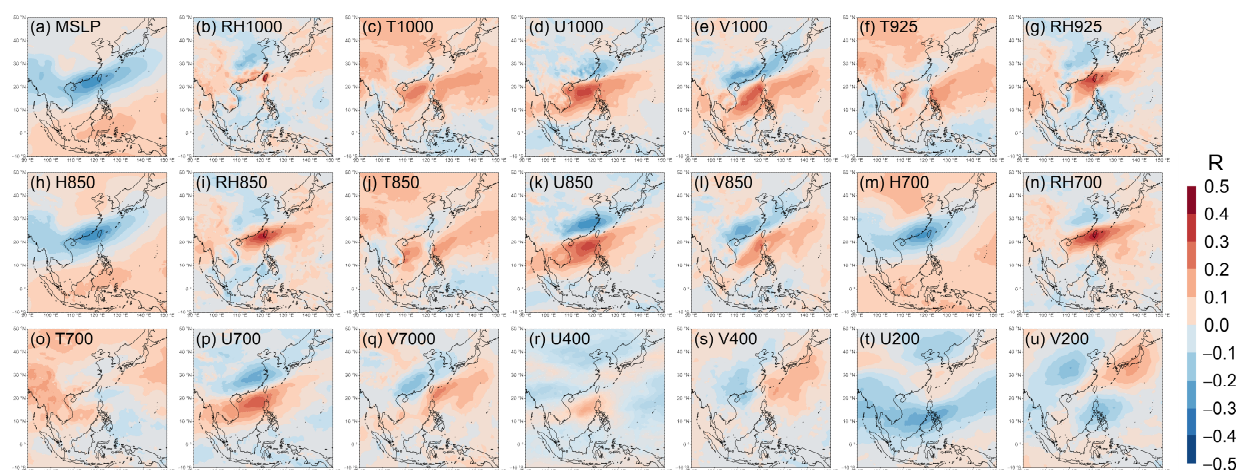


Figure 4. Correlation coefficients of the CFSR variables and Meiyu rainfall in southern Taiwan. (a) MSLP, (b) RH1000, (c) T1000, (d) U1000, (e) V1000, (f) T925, (g) RH925, (h) H850, (i) RH850, (j) T850, (k) U850, (l) V850, (m) H700, (n) RH700, (o) T700, (p) U700, (q) V700, (r) U400, (s) V400, (t) U200, and (u) V200.

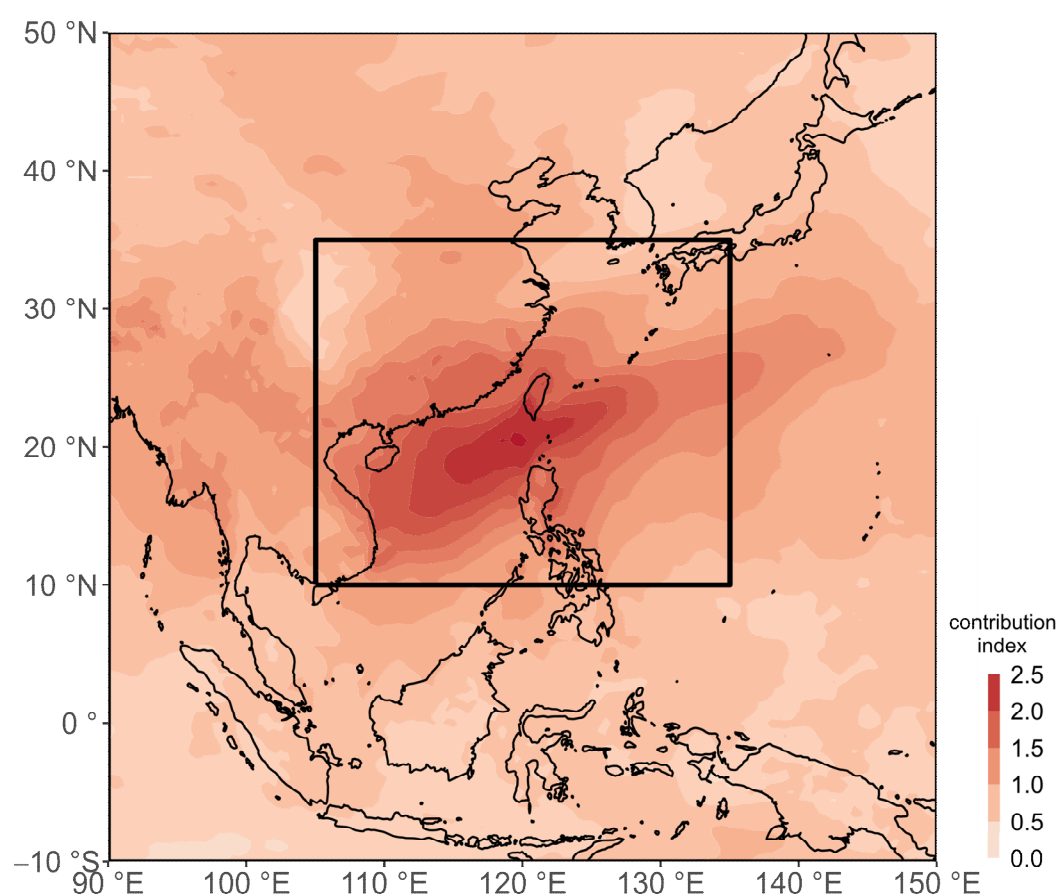


Figure 5. The composite of contribution index (CI). CI is obtained from the absolute values of correlation coefficients and the percentage of total explained variances of the leading 45 PCs for the selected CFSR variables.

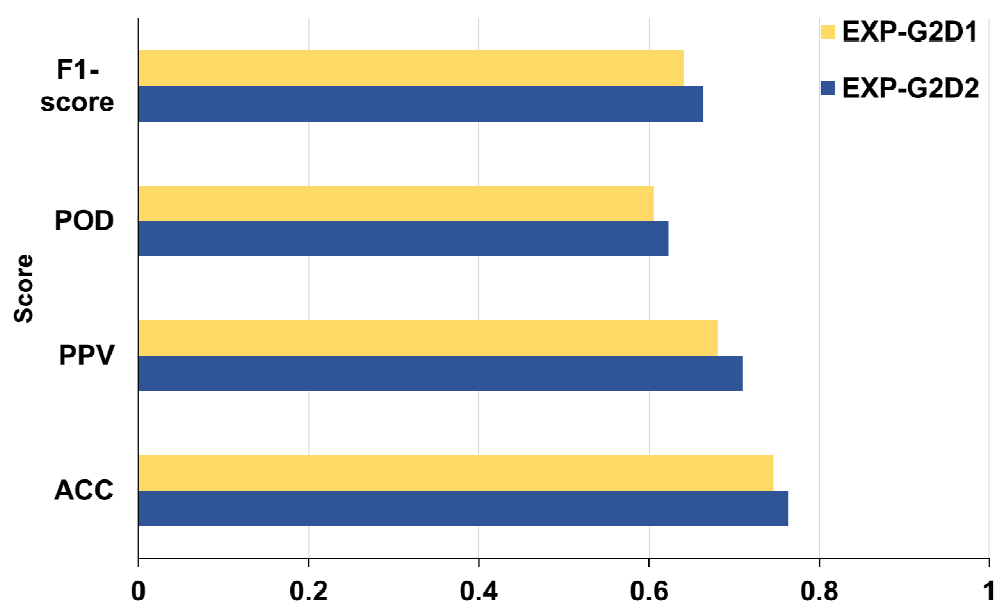


Figure 6. F1-score, POD, PPV, and ACC for EXP-G2D1 and EXP-G2D2.

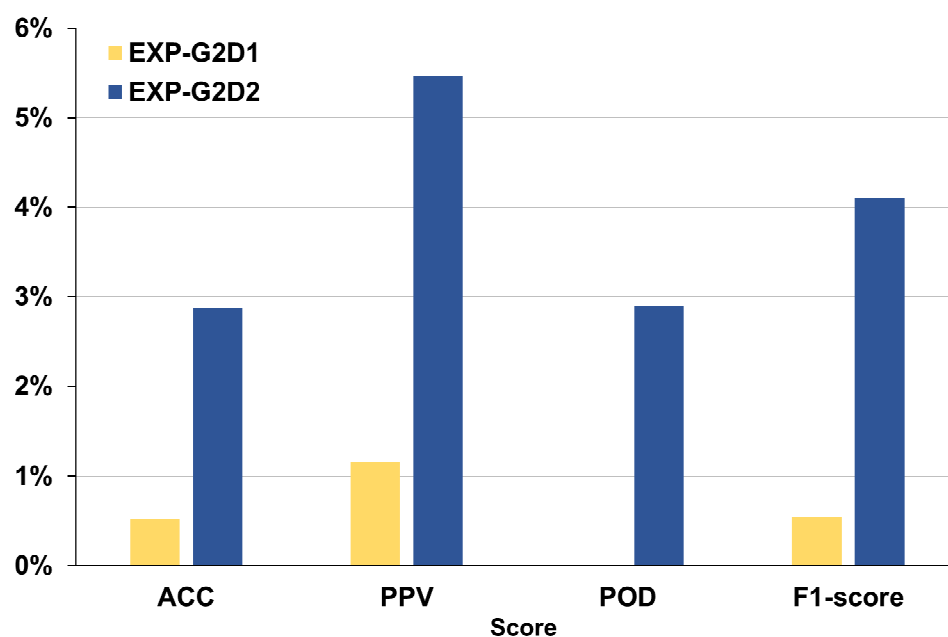


Figure 7. The difference ratio of ACC, PPV, POD, and F1-score for EXP-G2D1 (yellow) and EXP-G2D2 (blue) relative to EXP-G1D1.

The inter-annual variability of the F1-score in EXP-G2D2 is shown in Figure 8, which examines the SVM-based prediction scheme's ability to capture the extreme rainfall signals in the Meiyu season from 1990 to 2019. The result of a testing year can be obtained by excluding the data of the testing year and then training the SVM scheme by the data from the other 29 years [31]. This process is carried out for one year at a time. The SVM module will be re-trained when changing the target testing year. After repeating the process for 30 different testing years, the 30-year's evaluation score would be obtained. Figure 8 indicates that the maximum of the F1-score is about 0.8 in 1994 and 0.25 for the minimum in 1992. The average of the F1-score is 0.6. A composite map for low-level moisture and circulation difference is made between the top 5 and the last 5 years of the F1-score (Figure 9). Shadings in Figure 9 represent the composite difference of low-level water vapor flux (integrated water vapor flux from 1000 hPa to 850 hPa), and vectors represent

the difference of 850 hPa winds (less than 1 m s^{-1} are not shown). The composite map indicates that the low-level water vapor flux in the region of 15° N – 20° N , 115° E – 120° E is higher for top-5 years. In contrast, the low-level water vapor flux is found lower in the region of the Philippian Sea to the southern South China Sea, accompanied by the weaker southwesterly winds. Based on the result, a higher forecast performance would be expected when stronger southwesterly and abundant moisture are observed in the region of 15° N – 20° N , 115° E – 120° E for a given year.

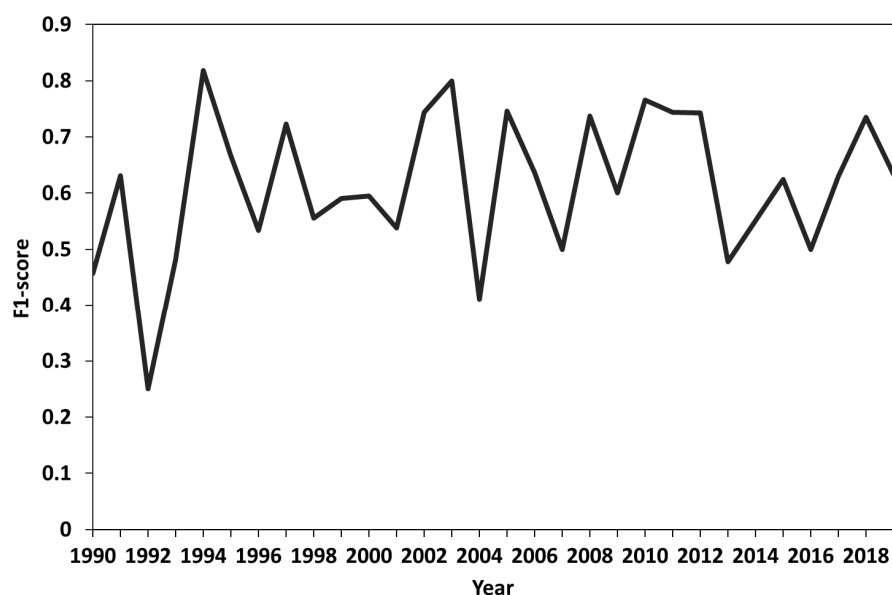


Figure 8. F1-score for EXP-G2D2 from 1990 to 2019.

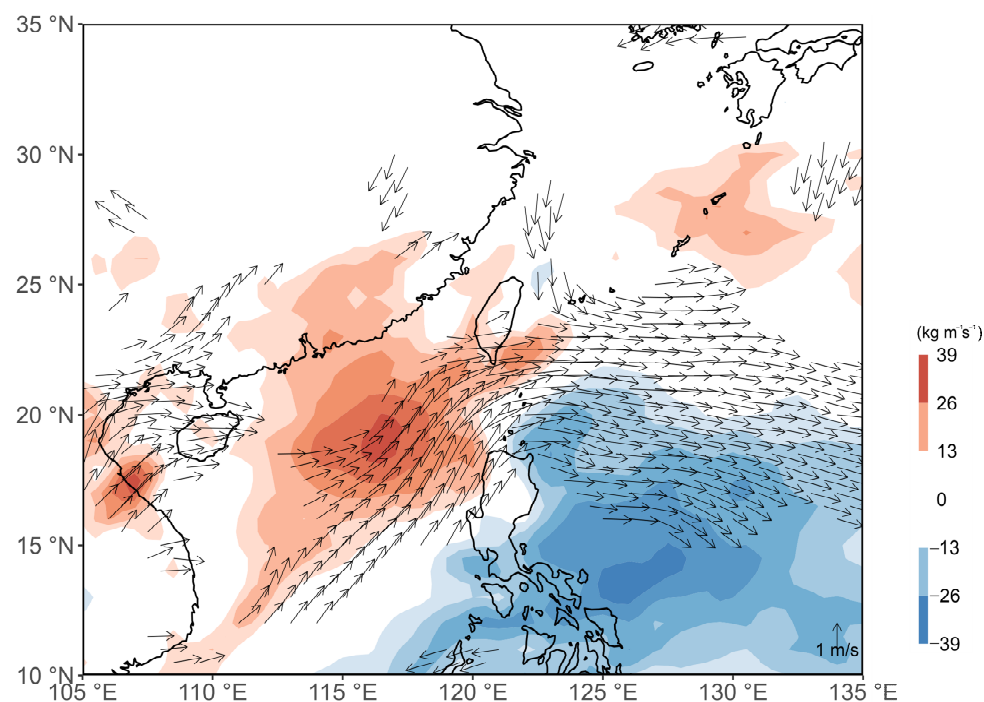


Figure 9. Composite for low-level (integrated from 1000 hPa to 850 hPa) water vapor flux (shadings) and 850 hPa winds (vectors) difference between the top 5 and the last 5 years of the F1-score from 1990 to 2019.

Figure 10 shows the performance of the forecast schemes for different lead times. These schemes are constructed based on EXP-G2D2, except for predicting in different lead times (Table 2). The PPV and ACC are over 0.6 for all the lead times. The POD and F1-score reveal a decaying along with the increase in the lead time. For the F1-score, most of the schemes are more than 0.5 concerning different lead times, except for the lead time of 76 h. In general, the settings of EXP-G2D2 are capable of performing good forecast scores 64 h earlier to the extreme Meiyu rainfall event in southern Taiwan.

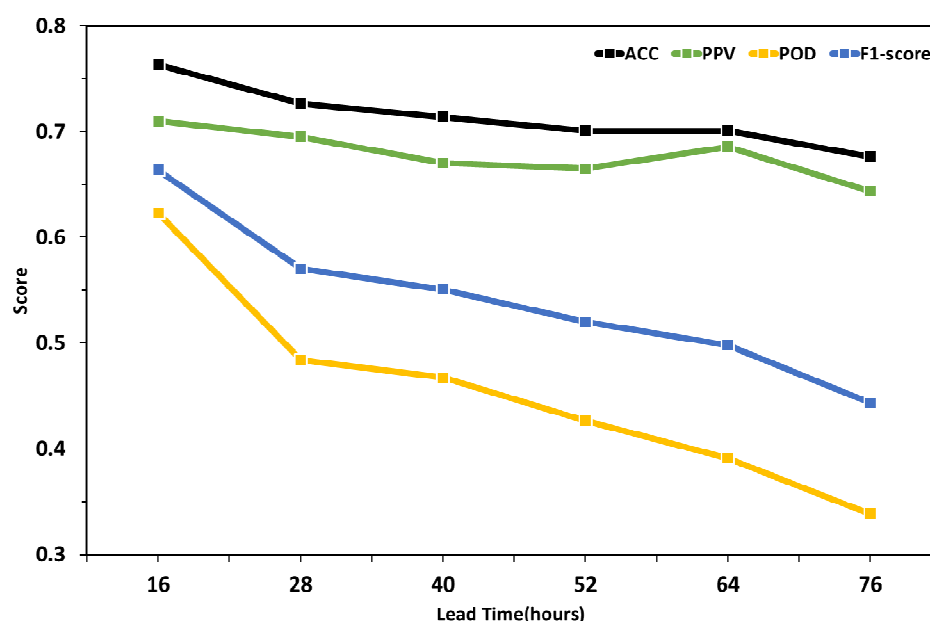


Figure 10. The ACC (black), PPV (green), POD (yellow), and F1-score (blue) along with different lead times from 16 to 76 h. The results of lead times 16, 28, 40, 52, 64, and 76 h are based on EXP-G2D2, EXP-G2D2-L2, EXP-G2D2-L3, EXP-G2D2-L4, EXP-G2D2-L5, and EXP-G2D2-L6, respectively.

4. Discussion and Conclusions

This study investigates the strategy to construct an optimal SVM-based scheme for predicting the extreme Meiyu rainfall events in southern Taiwan. The choice of predictors and the selection of domain play crucial roles in scheme construction. Different strategies concerning the combination of predictors and domains are implemented to design several experiments for different schemes. The ACC, PPV, POD, and F1-score can exceed 0.6 on average. The strategy of EXP-G2D2 performs best among all experiments. The results indicate that the strategy of EXP-G2D2 can improve the forecast performance of the SVM prediction scheme. In other words, choosing the predictors associated with the Meiyu system and determine the domain associated with the correlation coefficients of selected predictors and predictand can benefit the construction of the prediction scheme. The performance of WRF rainfall forecasting during the Meiyu season in Taiwan has been reported in recent years [17,18]. For example, the threat score (TS) for the 24 h forecasts is about 0.23 at the 50 mm threshold and 0.3 at the 25 mm threshold from 2012–2014 Meiyu seasons [18]. Although the case, the rainfall threshold, and the leading time are different, it is still interesting to show our TS evaluation for reference. As a result, the TS of the SVM-based prediction scheme in this study reaches 0.4 for the 28 h forecast at 35 mm threshold over southern Taiwan.

The SVM-based scheme's ability to capture the inter-annual variability of extreme Meiyu rainfall events is also investigated. According to the results, higher forecast performance will be expected for the scheme when stronger southwesterly as well as abundant low-level moisture is found over the region to the north of the South China Sea (15° N–20° N, 115° E–120° E) for a given year. In addition, the predictability of the scheme

based on EXP-G2D2 is further examined. The scheme shows a stable capability to predict the extreme Meiyu rainfall events in southern Taiwan for lead times from 16 h to 64 h.

This study proposes a strategy to optimize the implementation of an SVM-based scheme for predicting the extreme Meiyu rainfall events in southern Taiwan. Some critical issues such as the choice of predictors and the determination of domain are discussed in the article. A stable scheme could be obtained when following this strategy, while some uncertainty on the inter-annual variability still exists. The result reveals the high potential of AI in extreme rainfall prediction. However, the prediction of rainfall amount is not addressed yet. Further investigation is required for the application of AI techniques to meteor-hydrological disciplines.

Author Contributions: Conceptualization, J.-L.C. and C.-C.C.; methodology, S.-H.S.; software, T.-S.Y.; validation, L.-R.H., K.-L.L. and C.-J.W.; formal analysis, J.-L.C. and C.-C.C.; investigation, L.-H.H.; resources, L.-R.H. and Y.-C.Y.; data curation, J.-L.C., C.-C.C. and L.-H.H.; writing—original draft preparation, J.-L.C.; writing—review and editing, L.-H.H.; visualization, C.-C.C.; supervision, Y.-C.Y.; project administration, J.-L.C. and Y.-C.Y.; funding acquisition, Y.-C.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Central Weather Bureau (CWB) Taiwan station data. Available online: <https://www.cwb.gov.tw/V8/C/D/DailyPrecipitation.html> (accessed on 5 January 2021); Climate forecast system reanalysis (CFSR) data. Available online: <https://climatedataguide.ucar.edu/climate-data/climate-forecast-system-reanalysis-cfsr> (accessed on 10 January 2021).

Acknowledgments: The authors appreciate Chia-Chun Wu to support the edit of the article. We also thank the National Science and Technology Center for Disaster Reduction for the support of funding and other resources.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Chen, C.-S.; Chen, Y.-L. The Rainfall Characteristics of Taiwan. *Mon. Weather Rev.* **2003**, *131*, 1323–1340. [\[CrossRef\]](#)
- Chen, G.T.J. Research on the phenomena of Meiyu during the past quarter century: An overview. In *World Scientific Series for Meteorology of East Asia, East Asian Monsoon*; Chang, C.P., Ed.; World Scientific Publishing Co.: Hackensack, NJ, USA, 2004; Volume 2, pp. 357–403.
- Chen, G.T.J. Observation aspects of the Mei-Yu phenomena in subtropical China. *J. Meteorol. Soc. Jpn.* **1983**, *61*, 306–312. [\[CrossRef\]](#)
- Estimated Production Loss of Major Agricultural Disasters. 1994–2020. Available online: https://agrstat.coa.gov.tw/sdweb/public/book/Book_File.ashx?chapter_id=371_80_2 (accessed on 16 August 2021).
- Chien, F.-C.; Chiu, Y.-C.; Tsou, C.-H. A Climatological Study of Southwesterly Flows and Heavy Precipitation in Taiwan during Mei-yu Seasons from 1979 to 2018. *J. Meteorol. Soc. Jpn.* **2021**. [\[CrossRef\]](#)
- Chen, G.T.J.; Yu, C.C. Study of low-level jet and extremely heavy rainfall over northern Taiwan in the Mei-Yu season. *Mon. Weather Rev.* **1988**, *116*, 884–891. [\[CrossRef\]](#)
- Chen, G.T.J.; Wang, C.-C.; Lin, D.T.-W. Characteristics of low-level jets over northern Taiwan in Mei-yu season and their relationship to heavy rain events. *Mon. Weather Rev.* **2005**, *133*, 20–43. [\[CrossRef\]](#)
- Chen, G.T.J.; Hsu, Y.S. Composite structure of a low level jet over southern China observed during the TAMEX period. *J. Meteorol. Soc. Jpn.* **1997**, *75*, 1003–1018. [\[CrossRef\]](#)
- Chien, F.-C.; Chiu, Y.-C. A composite study of southwesterly flows and rainfall in Taiwan. *J. Meteorol. Soc. Jpn.* **2019**, *97*, 1023–1040. [\[CrossRef\]](#)
- Karstens, C.D.; Stumpf, G.; Ling, C.; Hua, L.; Kingfield, D.; Smith, T.M.; Correia, J., Jr.; Calhoun, K.; Ortega, K.; Melick, C.; et al. Evaluation of a probabilistic forecasting methodology for severe convective weather in the 2014 hazardous weather testbed. *Weather Forecast.* **2015**, *30*, 1551–1570. [\[CrossRef\]](#)
- McGovern, A.; Elmore, K.L.; Gagne, D.J.; Haupt, S.E.; Karstens, C.D.; Lagerquist, R.; Smith, T.; Williams, J.K. Using artificial intelligence to improve real-time decision making for high-impact weather. *Bull. Am. Meteorol. Soc.* **2017**, *98*, 2073–2090. [\[CrossRef\]](#)
- Wimmers, A.; Velden, C.; Cossuth, J.H. Using deep learning to estimate tropical cyclone intensity from satellite passive microwave imagery. *Mon. Weather Rev.* **2019**, *147*, 2261–2282. [\[CrossRef\]](#)

13. Lagerquist, R.; McGovern, A.; Gagne, D.J. Deep Learning for Spatially Explicit Prediction of Synoptic-Scale Fronts. *Weather Forecast.* **2019**, *34*, 1137–1160. [[CrossRef](#)]
14. Shahriar, S.A.; Kayes, I.; Hasan, K.; Salam, M.A.; Chowdhury, S. Applicability of machine learning in modeling of atmospheric particle pollution in Bangladesh. *Air Qual. Atmos. Health* **2020**, *13*, 1247–1256. [[CrossRef](#)]
15. Su, S.H.; Chu, J.L.; Yo, T.S.; Lin, L.Y. Identification of synoptic weather types over Taiwan area with multiple classifiers. *Atmos. Sci. Lett.* **2018**, *19*, e861. [[CrossRef](#)]
16. Hope, P.; Keay, K.; Pook, M.; Catto, J.; Simmonds, I.; Mills, G.; McIntosh, P.; Risbey, J.; Berry, G. A comparison of automated methods of front recognition for climate studies: A case study in southwest Western Australia. *Mon. Weather Rev.* **2014**, *142*, 343–363. [[CrossRef](#)]
17. Wang, C.-C.; Chien, F.-C.; Paul, S. An evaluation of WRF rainfall forecasts in Taiwan during three Mei-Yu seasons from 2008 to 2010. *Weather Forecast.* **2017**, *32*, 1329–1351. [[CrossRef](#)]
18. Wang, C.-C.; Chuang, P.-Y.; Chang, C.-S.; Tsuboki, K.; Huang, S.-Y.; Leu, G.-C. Evaluation of Mei-Yu heavy-rainfall quantitative precipitation forecasts in Taiwan by a cloud-resolving model for three seasons of 2012–2014. *Nat. Hazards Earth Syst. Sci. Discuss.* **2021**. in review. [[CrossRef](#)]
19. IPCC. Annex III: Glossary [Planton, S. (Ed.)]. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. In *Climate Change 2013: The Physical Science Basis*; Stocker, T.F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S.K., Boschung, J., Nauels, A., Xia, Y., Bex, V., Midgley, P.M., Eds.; Cambridge University Press: Cambridge, UK; New York, NY, USA, 2013; pp. 1447–1466. [[CrossRef](#)]
20. Saha, S.; Moorthi, S.; Pan, H.-L.; Wu, X.; Wang, J.; Nadiga, S.; Tripp, P.; Kistler, R.; Woollen, J.; Behringer, D.; et al. The ncep climate forecast system reanalysis. *Bull. Am. Meteorol. Soc.* **2010**, *91*, 1015–1058. [[CrossRef](#)]
21. Fernández-Delgado, M.; Cernadas, E.; Barro, S.; Amorim, D. Do we need hundreds of classifiers to solve real world classification problems. *J. Mach. Learn. Res.* **2014**, *15*, 3133–3181.
22. Halko, N.; Martinsson, P.-G.; Shkolnisky, Y.; Tygert, M. An algorithm for the principal component analysis of large data sets. *SIAM J. Sci. Comput.* **2011**, *33*, 2580–2594. [[CrossRef](#)]
23. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine learning in python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
24. Karatzoglou, A.; Smola, A.; Hornik, K.; Zeileis, A. Kernlab—An S4 package for kernel methods in R. *J. Stat. Softw.* **2004**, *11*, 1–20. [[CrossRef](#)]
25. Stone, M. Cross-validated choice and assessment of statistical predictions. *J. R. Stat. Soc. Ser. B* **1974**, *36*, 111–147. [[CrossRef](#)]
26. Wang, C.-C.; Chen, G.T.J. Case Study of the Leaside Mesolow and Mesocyclone in TAMEX. *Mon. Weather Rev.* **2002**, *130*, 2572–2592. [[CrossRef](#)]
27. Feddersen, H.; Andersen, U. A method for statistical downscaling of seasonal ensemble predictions. *Tellus A Dyn. Meteorol. Oceanogr.* **2005**, *57*, 398–408. [[CrossRef](#)]
28. Benestad, R.E. Empirical-statistical downscaling in climate modeling. *Eos Trans. AGU* **2004**, *85*, 417–422. [[CrossRef](#)]
29. Huang, R.; Huang, G.; Wei, Z. Climate variations of the summer monsoon over China. In *World Scientific Series for Meteorology of East Asia, East Asian Monsoon*; Chang, C.P., Ed.; World Scientific Publishing Co.: Hackensack, NJ, USA, 2004; Volume 2, p. 217.
30. Jolliffe, I.T.; Stephenson, D.B. *Forecast. Verification: A Practitioner's Guide in Atmospheric Science*, 2nd ed.; Wiley-Blackwell: Oxford, UK, 2012.
31. Michaelsen, J. Cross-validation in statistical climate forecast models. *J. Clim. Appl. Meteorol.* **1987**, *26*, 1589–1600. [[CrossRef](#)]