

ENHANCED DAILY CHIRPS PRECIPITATION USING SEQUENTIAL K-NEAREST NEIGHBORS CORRECTION AND KALMAN FILTER BLENDING FOR THAILAND

Winai Chaowiwat, Jirayuth Srisat, Kritanai Torsri, Kanoksri Sarinnapakorn
Hydro-Informatics Institute (Public Organization),
Ministry of Higher Education, Science, Research and Innovation, Bangkok, Thailand

Received: 8/9/2025; Accepted: 22/9/2025

Abstract: Accurate precipitation data is essential for hydrological modeling and water resource management, particularly in tropical regions with complex topography and limited ground-based observation networks. This study develops an integrated two-stage framework combining K-nearest neighbors (KNN) machine learning bias correction with Kalman filter blending to enhance Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS) daily precipitation estimates across Thailand's diverse geographical and climatic conditions. The methodology utilized comprehensive meteorological data from 628 stations across Thailand spanning 44 years (1981-2024), with temporal partitioning into training (1981-2015) and validation (2016-2024) periods. The first stage implemented seasonal KNN bias correction using 11-dimensional feature vectors incorporating CHIRPS satellite precipitation, auxiliary meteorological variables (maximum/minimum temperature, relative humidity, evaporation), and station coordinates. The second stage applied adaptive Kalman filter blending with dual-update processing, combining raw CHIRPS data with KNN-corrected estimates. Results demonstrate exceptional performance improvements across both periods. Correlation coefficients increased dramatically from 0.42 to 0.94 during training (124% improvement) and from 0.41 to 0.91 during validation (122% improvement). Systematic bias correction transformed raw CHIRPS overestimation of 34.03% to controlled underestimation of -10.00% (BC CHIRPS) and -10.78% (BBC CHIRPS) during training, with similar validation patterns. Regional analysis revealed differential effectiveness across Thailand's climatic zones. The most challenging DJF dry season showed severe raw CHIRPS overestimation of 588.27% (training) and 167.12% (validation), reduced by 95-99% with both corrections. Spatial validation confirmed operational applicability, effectively eliminating widespread overestimation while preserving legitimate precipitation signals. The integrated framework successfully addresses systematic biases in satellite precipitation products while maintaining computational efficiency. This research demonstrates that sophisticated machine learning integrated with optimal filtering theory can significantly enhance satellite precipitation accuracy for operational applications in data-scarce tropical regions, with demonstrated effectiveness across Thailand's diverse conditions and strong potential for broader tropical applications.

Keywords: CHIRPS bias correction, K-nearest neighbors (KNN), Kalman filter blending, satellite precipitation, machine learning, Thailand.

1. Introduction

Accurate precipitation data is essential for hydrological modeling, water resources management, climate monitoring, and

agricultural planning. However, ground-based rain gauge networks remain inadequate, particularly in remote forested regions and areas with limited infrastructure [1], [2]. This limitation is especially pronounced in tropical regions where complex topography and dense vegetation challenge comprehensive

Corresponding author: Winai Chaowiwat
E-mail: winai@hii.or.th

meteorological networks. Satellite-based precipitation products address these spatial data gaps. CHIRPS (Climate Hazards Group InfraRed Precipitation with Station data) provides daily precipitation estimates at 0.05° resolution from 1981 to present [3], combining satellite infrared observations, microwave precipitation estimates, and ground station data. However, systematic biases in satellite-based retrievals require correction procedures for regional applications [4], [5]. Traditional bias correction methods include multiplicative or additive adjustments based on statistical relationships with ground observations [6]. However, these linear methods often fail to capture complex, nonlinear relationships between satellite observations and surface precipitation. Machine learning algorithms address these limitations. Random Forest algorithms improve precipitation estimates by learning complex relationships between satellite channels and ground observations [7]. Support Vector Machines effectively correct precipitation biases across climatic regions [8]. Neural networks capture nonlinear precipitation processes [9], [10].

K-nearest neighbors (KNN) algorithms preserve local patterns while providing robust nonparametric estimation [11]. KNN methods successfully apply to precipitation downscaling [12], drought forecasting [13], and streamflow prediction [14]. The algorithm identifies similar meteorological conditions for prediction, making it suitable for precipitation correction where local patterns influence bias characteristics. Recent studies demonstrate KNN effectiveness in satellite precipitation correction. Baez-Villanueva et al. [15] showed machine learning methods, including KNN, outperformed traditional techniques for CHIRPS data. Similarly, Fang et al. [16] demonstrated improved estimates using KNN-based methods incorporating multiple meteorological variables.

Optimal data fusion techniques combine multiple precipitation sources. Kalman filtering provides theoretically sound sequential data assimilation, optimally combining observations

with varying uncertainties [17]. Extended and Ensemble Kalman filters have improved precipitation estimation [18], [19]. Recent advances focus on multi-source integration. Sinclair & Pegram [20] developed optimal interpolation for combining satellite and gauge data. Habib et al. [21] demonstrated Kalman filter effectiveness for merging precipitation datasets. However, limited research explores combining machine learning bias correction with optimal filtering techniques. This study develops an integrated two-stage framework combining K-nearest neighbors bias correction with Kalman filter optimization. Research objectives include: (1) Developing seasonal KNN bias correction incorporating multi-dimensional meteorological features and spatial optimization for CHIRPS data, (2) Implementing adaptive Kalman filter blending combining raw satellite data with bias-corrected estimates, and (3) Comprehensively evaluating framework performance across temporal periods and spatial scales using Thailand's meteorological network. The methodology employs a two-stage approach using 628 meteorological stations across Thailand (1981-2024). Stage 1 implements seasonal KNN bias correction with ensemble learning, incorporating CHIRPS data with auxiliary variables (temperature, humidity, evaporation) through feature engineering and spatial optimization testing 1-15 neighboring stations. Stage 2 applies adaptive Kalman filtering combining raw CHIRPS with KNN-corrected estimates through parameter optimization, extending from station-based optimization to comprehensive gridded coverage across 18,513 grid points. This integrated approach aims to enhance precipitation accuracy while maintaining computational efficiency, contributing to improved satellite precipitation correction for hydrological and climatological applications in data-scarce regions.

2. Study Area and Data Used

Thailand was selected due to its tropical climate, diverse topography, and comprehensive meteorological network [22], [23]. Located

between 5°37'N to 20°27'N and 97°22'E to 105°37'E, Thailand covers 513,120 km² including mountainous terrain, central plains, the Khorat Plateau, and Southern peninsula [24,25]. This diversity creates complex precipitation patterns influenced by Southwest (May-October) and Northeast (November-February) monsoons, making it ideal for testing satellite precipitation correction methodologies [26]. The study utilized data from 628 meteorological stations across Thailand spanning 44 years (1981-2024) [27] (Figure 1). Station elevations range from sea level to over 1,500 meters, capturing topographic precipitation effects [28]. CHIRPS satellite data provided quasi-global precipitation estimates at 0.05° resolution covering 18,513 grid points across Thailand [3]. CHIRPS combines satellite infrared observations, microwave estimates, and ground station data, selected for its long temporal record and proven tropical performance [29,

30]. Ground observations included daily precipitation, maximum/minimum temperature, relative humidity, and evaporation from 628 stations following WMO standards [31], [32]. All data underwent quality control including range checks and spatial validation [33]. Missing data was generally low (<5%) [34]. The dataset was partitioned into training (1981-2015) and validation (2016-2024) periods [35]. Data were stratified by seasons: December-January-February (DJF), March-April-May (MAM), June-July-August (JJA), and September-October-November (SON) [36]. Data preprocessing included temporal alignment, missing value identification, and outlier detection [37]. The final dataset maintained high completeness with 2,512 station-season combinations, providing ideal foundation for developing the integrated bias correction framework across Thailand's diverse conditions [38].

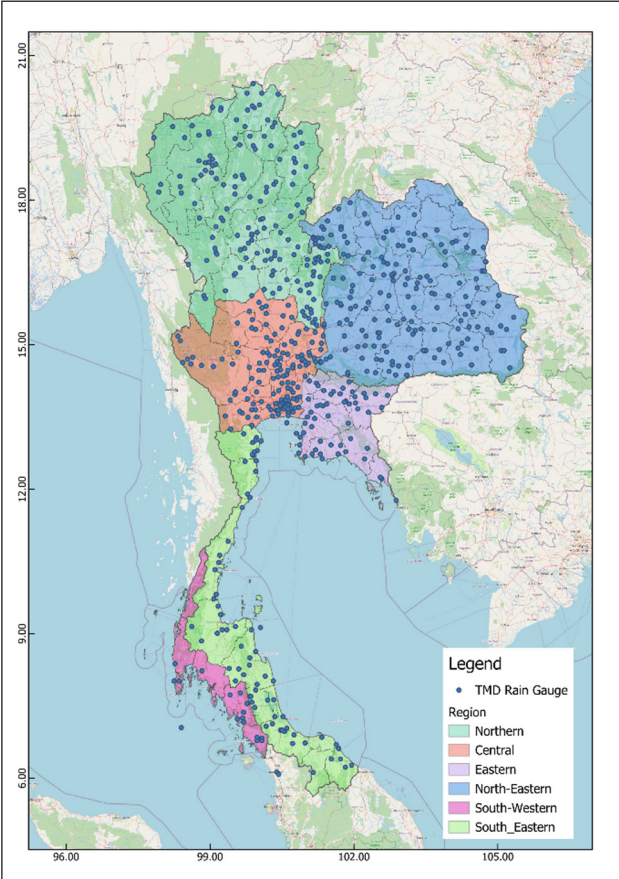


Figure 1. Distribution of TMD rainfall gauge stations across six regions of Thailand

3. Methodology

This study employed an integrated two-stage framework combining K-nearest neighbors machine learning with Kalman filter blending to enhance CHIRPS precipitation estimates across Thailand [39], [17]. The methodology addresses systematic biases while maintaining computational efficiency [40]. The dataset utilized 628 stations across Thailand with elevations from sea level to over 1,500 meters [27]. CHIRPS provided precipitation estimates at 0.05° resolution covering 18,513 grid points [3]. Ground observations included daily precipitation, temperature, humidity, and evaporation following WMO standards [31]. Data were partitioned into training (1981-2015) and validation (2016-2024) periods and stratified into four seasons: DJF, MAM, JJA, and SON [35], [36]. For each station, 11-dimensional feature vectors incorporated CHIRPS precipitation, meteorological variables, and coordinates [41]. Spatial optimization determined neighboring configurations using Euclidean distance.

KNN implementation tested K values of 1, 3, and 5 with Euclidean, Minkowski, and Manhattan distance metrics [42]. Bagging ensemble used bootstrap sampling with arithmetic mean aggregation [43].

Feature scaling applied standardization:

$$X_{scaled} = \frac{X - \mu_{train}}{\sigma_{train}} \quad (1)$$

Seasonal model development created 2,512 models (628 stations × 4 seasons) with cross-validation [44].

The Kalman filter stage mapped each station to nearest CHIRPS grid point:

$$Grid_{point} = \operatorname{argmin}(\operatorname{distance}(\operatorname{station}_{coords}, \operatorname{grid}_{coords})) \quad (2)$$

State-space model used persistence:

$$x(t) = x(t-1) + w(t) \quad (3)$$

with observation equations:

$$Z_{raw}(t) = x(t) + v_{raw}(t) \quad Z_{corr}(t) = x(t) + v_{corr}(t) \quad (4)$$

The dual-update process included prediction:

$$x_{pred}(t) = x_{est}(t-1)P_{pred}(t) = P_{est}(t-1) + Q \quad (5)$$

and sequential updates with non-negativity constraints:

$$x_{final}(t) = \max(0, x_{est}(t)) \quad (6)$$

Performance evaluation used Pearson correlation and Percent Bias [45]. Multi-temporal evaluation covered historical and recent periods with comparative analysis across raw CHIRPS, KNN-corrected, and Kalman blended products [46].

These metrics were chosen because they directly address the two fundamental aspects of satellite precipitation correction: Temporal correspondence (correlation) and systematic bias magnitude (PBIAS), which are critical for hydrological modeling and water resource management applications. Correlation coefficient assesses the ability to capture precipitation variability and timing, while PBIAS quantifies systematic over- or under-estimation relating to water balance calculations. The comprehensive evaluation framework combined these statistical metrics with temporal analysis across training and validation periods, regional stratification across Thailand's six climatic zones, seasonal assessment through DJF, MAM, JJA, and SON periods, and spatial validation (Figures 2-8) demonstrating operational applicability. Accuracy improvements were reported using standardized calculations throughout Tables 1-2 and all discussion sections: For correlation coefficients, relative improvement = $((R_{corrected} - R_{raw}) / R_{raw}) \times 100\%$, and for bias reduction, percent reduction = $((|PBIAS_{raw}| - |PBIAS_{corrected}|) / |PBIAS_{raw}|) \times 100\%$, ensuring transparent and reproducible improvement quantification.

For the KNN configuration, we tested K values of 1, 3, and 5 neighbors with three distance metrics (Euclidean, Minkowski, Manhattan), with K=5 and Euclidean distance providing optimal performance across most stations and seasons after cross-validation. The bagging ensemble employed bootstrap sampling with 10 iterations

and arithmetic mean aggregation for final predictions, while all features were standardized using training set statistics (μ_{train} , σ_{train}) and applied consistently to validation data. Kalman filter parameters, including process noise (Q) and measurement noise covariances (R), were optimized through grid search on the training period with values ranging from 0.1 to 10.0 for different regions. To prevent information leakage, we implemented strict temporal partitioning with several critical safeguards: (1) Clean separation between training (1981-2015) and validation (2016-2024) periods with no overlap, (2) Station-wise standardization with scaling parameters computed exclusively from training data and frozen before validation, (3) Independent training of each of the 2,512 models (628 stations $\times$ 4 seasons) using only historical data, (4) KNN neighbor selection and Kalman filter updates using only past observations with no future information, and (5) Time-series cross-validation within the training period that respected temporal ordering.

4. Results

4.1. Evaluation of Bias Corrected Precipitation

Both Bias-Corrected CHIRPS (BC CHIRPS) and Blended Bias-Corrected CHIRPS (BBC CHIRPS) demonstrated exceptional improvements over raw CHIRPS during training period, as detailed in Table 1. Year-round correlation coefficients increased 124% from 0.42 to 0.94 for both methods, with regional performance variations illustrated in Figures 2-3 showing consistent improvements across all 628 meteorological stations. South-Western region showed the largest improvement (221%), followed by South-Eastern (191%) and Eastern (98%) regions. Northern, Central, and North-Eastern regions achieved 86-116% improvements. For bias reduction, BC CHIRPS achieved 70.6% overall improvement, transforming 34.03% overestimation to -10.00% underestimation, while BBC CHIRPS provided 68.3% improvement (-10.78% underestimation). The distributional transformation is visualized in Figures 4-5, showing systematic shift from wet to controlled dry bias. South-Western region showed largest

enhancement at 85.3% (BC) and 78.5% (BBC), followed by South-Eastern at $\sim$ 80%. Central and Eastern regions demonstrated 59-66% improvements, while Northern and North-Eastern regions achieved 23-32% reductions.

The transformation from systematic overestimation to controlled underestimation (-10.00% for BC CHIRPS and -10.78% for BBC CHIRPS) represents a strategic outcome known as "controlled dry bias." This intentional slight underestimation is preferable to overestimation in water resource management applications, as it provides a conservative estimate for water availability assessments and reduces the risk of overestimating water resources. This controlled bias pattern remained consistent during validation (Table 2), where 23.11% overestimation was reduced to -7.41% (BC) and -7.79% (BBC), demonstrating the framework's ability to maintain controlled underestimation across different time periods.

4.2. Seasonal rainfall analysis

Training period (1981-2015) box plot analysis revealed systematic Raw CHIRPS overestimation averaging 31.1%, with severe DJF bias (92.2%). South-Western and South-Eastern regions experienced highest overestimations (150% and 125% during DJF), while Northern regions showed moderate overestimations (40-50%). Both correction methods transformed systematic wet bias to controlled underestimations (-3.1% and -4.1%). MAM season showed 45.8% Raw CHIRPS overestimation, particularly in South-Western (71.4%) and Eastern (60%) regions. Wet seasons JJA and SON demonstrated moderate overestimations (20.6% and 25.2%). Both corrections achieved excellent bias control with underestimations from -2.1% to -8.9%. Validation period (2016-2024) showed improved Raw CHIRPS performance (18.9% average overestimation) and sustained correction effectiveness. BC CHIRPS and BBC CHIRPS maintained excellent control (-2.0% and -2.7% underestimation). DJF remained challenging (48.7% Raw CHIRPS overestimation), but corrections achieved excellent control (BC: -3.9%, BBC: -2.0%). Both methods demonstrated

robust transferability, with Raw CHIRPS naturally improving from 31.1% to 18.9% overestimation

while corrections maintained effectiveness across Thailand's diverse conditions.

Table 1. Goodness-of-fit test in the training period (1981-2015)

Season	Region	R			PBIAS (%)		
		Raw CHIRPS	BC CHIRPS	BBC CHIRPS	Raw CHIRPS	BC CHIRPS	BBC CHIRPS
DJF	Northern	0.26	0.79	0.78	254.86	-11.07	-6.66
	Central	0.26	0.81	0.81	236.91	-7.82	-3.15
	Eastern	0.31	0.87	0.87	163.62	-14.30	-12.71
	North-Eastern	0.29	0.82	0.82	129.96	-11.52	-8.68
	South-Western	0.25	0.81	0.80	1370.04	-8.93	-2.33
	South-Eastern	0.30	0.81	0.80	1374.21	-7.33	1.70
	Average	0.28	0.82	0.81	588.27	-10.16	-5.31
MAM	Northern	0.42	0.93	0.93	34.62	-12.71	-11.89
	Central	0.43	0.93	0.93	59.81	-10.58	-9.73
	Eastern	0.42	0.94	0.94	39.03	-11.33	-11.99
	North-Eastern	0.46	0.94	0.94	31.11	-10.36	-9.54
	South-Western	0.27	0.92	0.92	104.41	-9.56	-13.71
	South-Eastern	0.31	0.91	0.91	56.39	-15.14	-15.18
	Average	0.38	0.93	0.93	54.23	-11.61	-12.01
JJA	Northern	0.36	0.93	0.93	15.06	-8.99	-8.51
	Central	0.36	0.92	0.92	34.77	-9.45	-8.92
	Eastern	0.37	0.92	0.92	29.25	-9.93	-10.78
	North-Eastern	0.42	0.93	0.93	17.23	-7.75	-7.13
	South-Western	0.24	0.89	0.89	57.93	-8.88	-13.62
	South-Eastern	0.22	0.90	0.90	47.03	-6.98	-7.15
	Average	0.33	0.92	0.92	33.54	-8.66	-9.35
SON	Northern	0.43	0.94	0.94	18.61	-9.41	-8.69
	Central	0.48	0.94	0.94	36.58	-8.57	-8.00
	Eastern	0.49	0.94	0.94	35.28	-8.94	-9.54
	North-Eastern	0.52	0.94	0.94	13.54	-8.83	-8.14
	South-Western	0.20	0.92	0.92	71.65	-9.91	-14.30
	South-Eastern	0.27	0.90	0.91	60.64	-10.53	-10.73
	Average	0.40	0.93	0.93	39.38	-9.37	-9.90
All year round	Northern	0.44	0.95	0.95	13.36	-10.27	-9.82
	Central	0.47	0.94	0.94	28.22	-9.87	-9.54
	Eastern	0.48	0.94	0.95	27.14	-10.13	-11.06
	North-Eastern	0.51	0.95	0.95	12.64	-9.08	-8.56
	South-Western	0.29	0.93	0.93	67.17	-9.86	-14.44
	South-Eastern	0.32	0.93	0.93	55.66	-10.80	-11.23
	Average	0.42	0.94	0.94	34.03	-10.00	-10.78

Table 2. Goodness-of-fit test in the validation period (2016-2024)

Season	Region	R			PBIAS (%)		
		Raw CHIRPS	BC CHIRPS	BBC CHIRPS	Raw CHIRPS	BC CHIRPS	BBC CHIRPS
DJF	Northern	0.14	0.75	0.74	110.75	15.95	19.76
	Central	0.25	0.73	0.72	211.78	0.98	5.34
	Eastern	0.19	0.79	0.78	146.27	12.95	15.14
	North-Eastern	0.14	0.71	0.70	135.14	23.06	26.81
	South-Western	0.35	0.91	0.91	145.72	-0.62	-1.72
	South-Eastern	0.46	0.85	0.85	253.10	-17.90	-14.84
	Average	0.25	0.79	0.78	167.12	5.74	8.42
MAM	Northern	0.35	0.91	0.91	56.70	-5.54	-4.15
	Central	0.44	0.89	0.90	52.36	-8.48	-7.25
	Eastern	0.39	0.91	0.91	46.79	-2.16	-2.83
	North-Eastern	0.35	0.92	0.92	34.42	-5.47	-4.29
	South-Western	0.39	0.92	0.92	36.77	-10.11	-13.91
	South-Eastern	0.38	0.86	0.87	59.57	5.03	5.79
	Average	0.38	0.90	0.90	47.77	-4.45	-4.44
JJA	Northern	0.30	0.89	0.89	16.11	-6.12	-5.32
	Central	0.33	0.87	0.87	31.61	-8.74	-7.59
	Eastern	0.28	0.88	0.88	28.81	-4.31	-5.85
	North-Eastern	0.36	0.90	0.90	22.19	-1.47	-0.48
	South-Western	0.32	0.92	0.92	27.82	-8.20	-12.40
	South-Eastern	0.23	0.85	0.85	31.68	0.72	1.10
	Average	0.30	0.89	0.89	26.37	-4.69	-5.09
SON	Northern	0.37	0.91	0.91	31.00	-1.50	-0.37
	Central	0.41	0.91	0.91	32.14	-5.32	-4.39
	Eastern	0.39	0.91	0.91	39.19	-1.70	-2.11
	North-Eastern	0.47	0.92	0.92	20.98	2.16	3.09
	South-Western	0.32	0.92	0.92	28.99	-11.91	-16.12
	South-Eastern	0.35	0.88	0.88	38.88	-9.22	-8.50
	Average	0.38	0.91	0.91	31.86	-4.58	-4.73
All year round	Northern	0.38	0.92	0.92	17.85	-5.98	-5.23
	Central	0.42	0.91	0.91	24.60	-8.75	-7.96
	Eastern	0.39	0.91	0.91	27.98	-4.68	-5.55
	North-Eastern	0.43	0.92	0.92	17.56	-2.89	-2.02
	South-Western	0.39	0.93	0.93	25.22	-11.05	-15.26
	South-Eastern	0.42	0.87	0.88	25.44	-11.10	-10.75
	Average	0.41	0.91	0.91	23.11	-7.41	-7.79

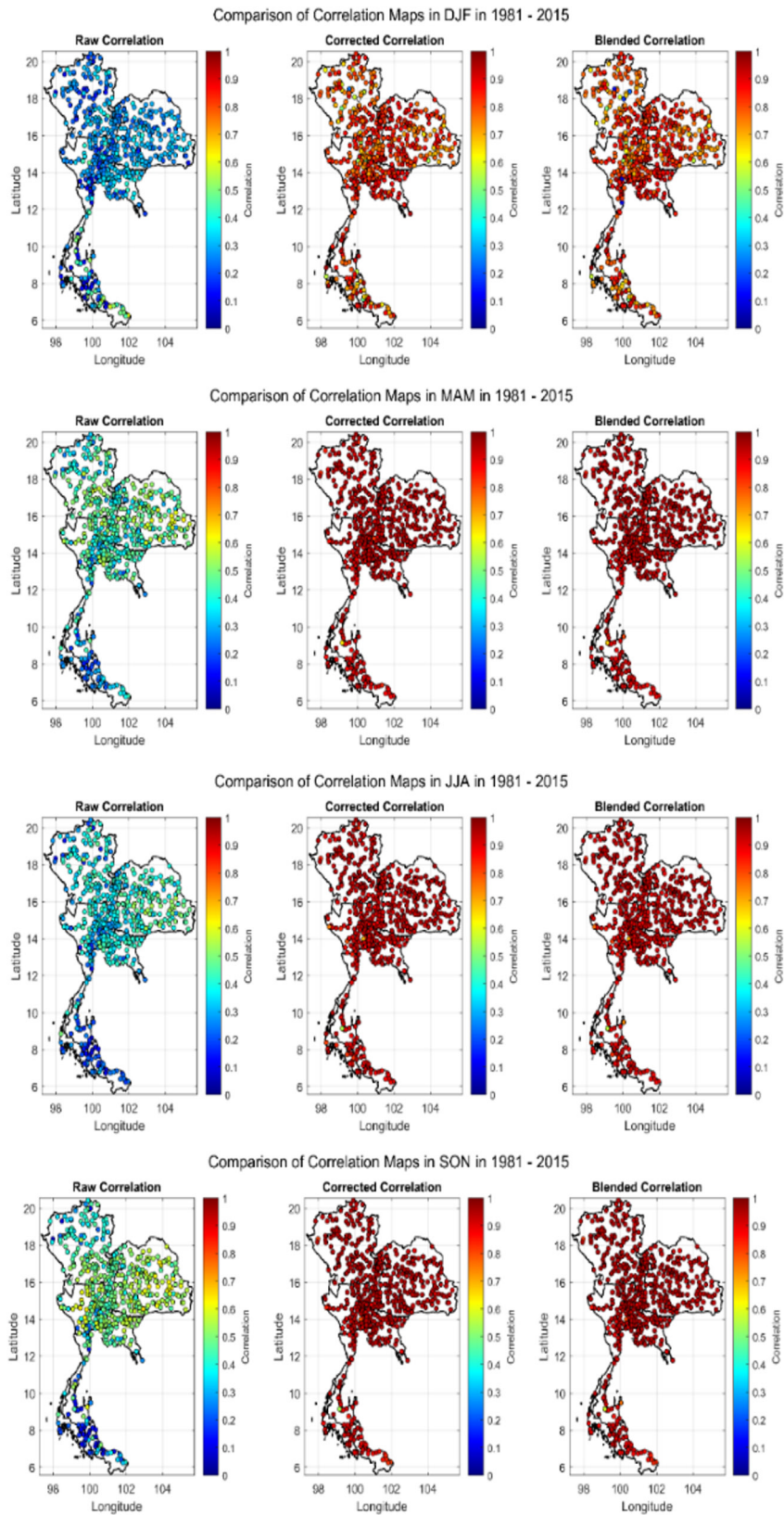


Figure 2. Comparison of seasonal correlation maps in the training period (1981-2015)

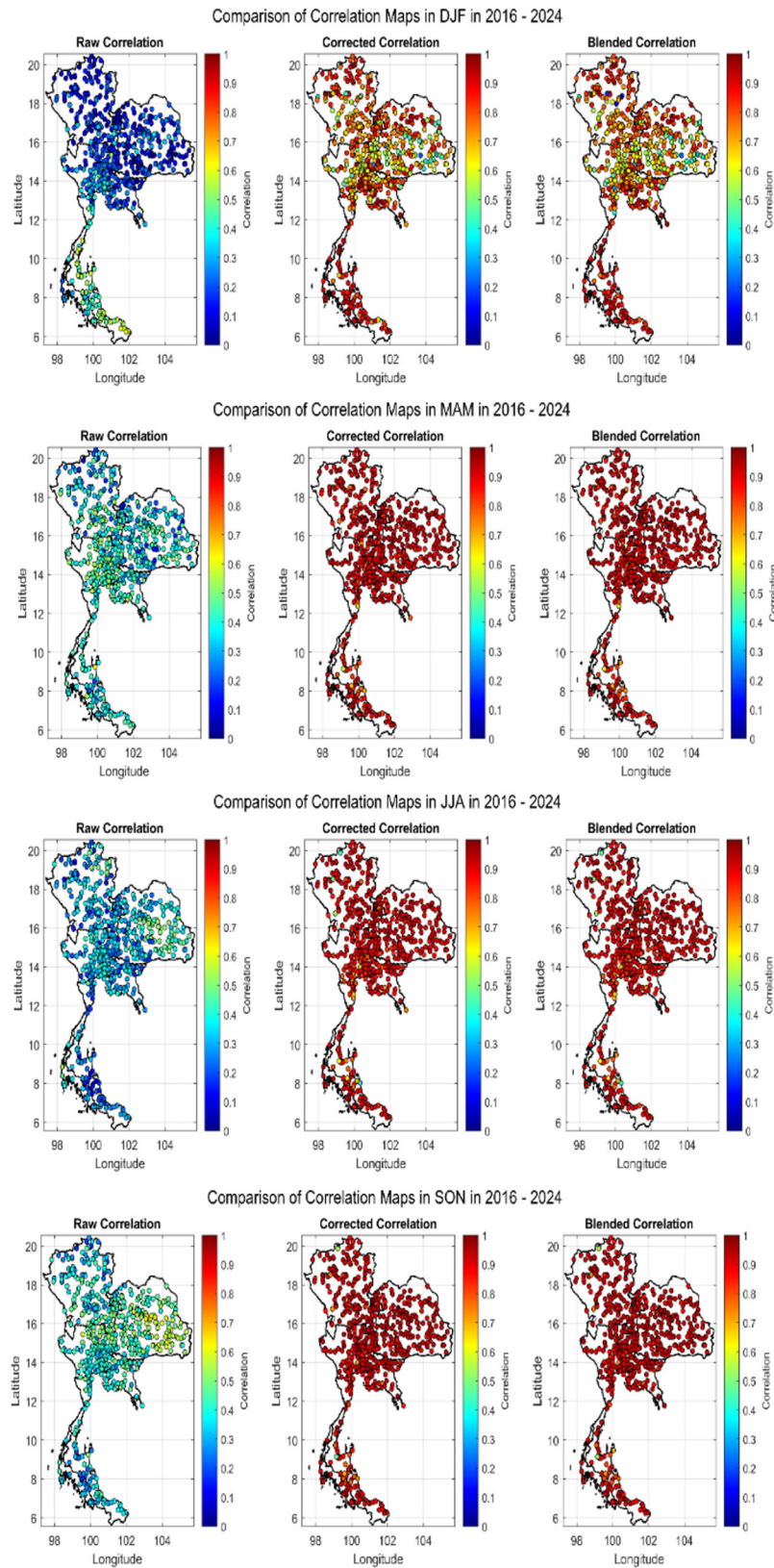


Figure 3. Comparison of seasonal correlation map in the validation period (2016-2024)

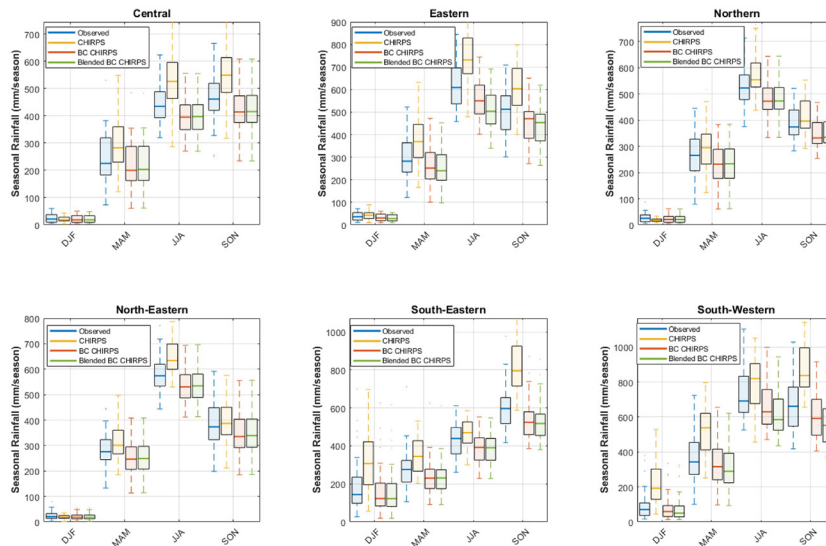


Figure 4. Box plots of seasonal rainfall in each region of Thailand in the training period (1981-2015)

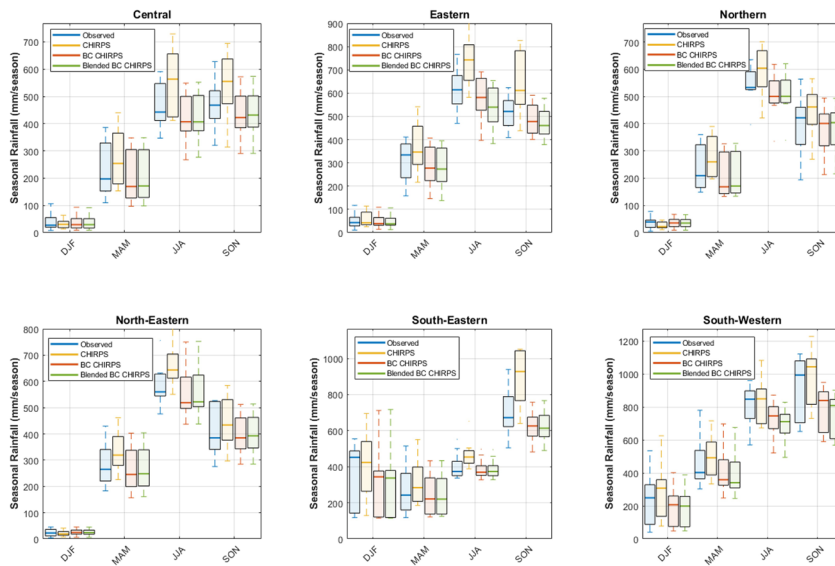
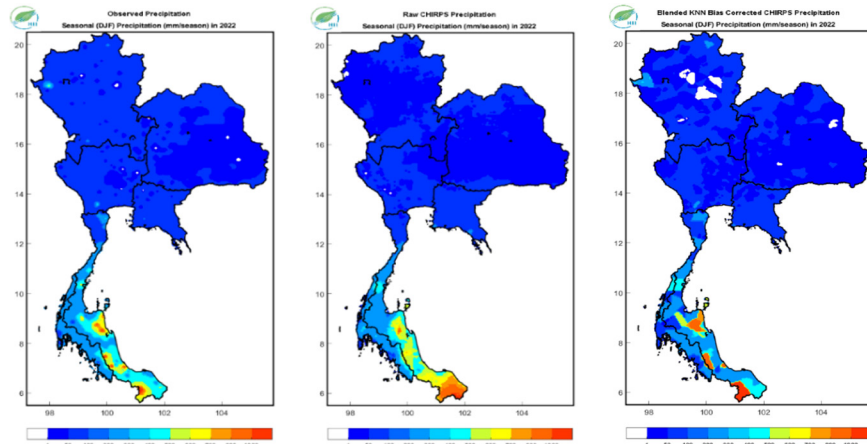


Figure 5. Box plots of seasonal rainfall in each region of Thailand in the validation period (2016-2024)

4.3. Spatial analysis

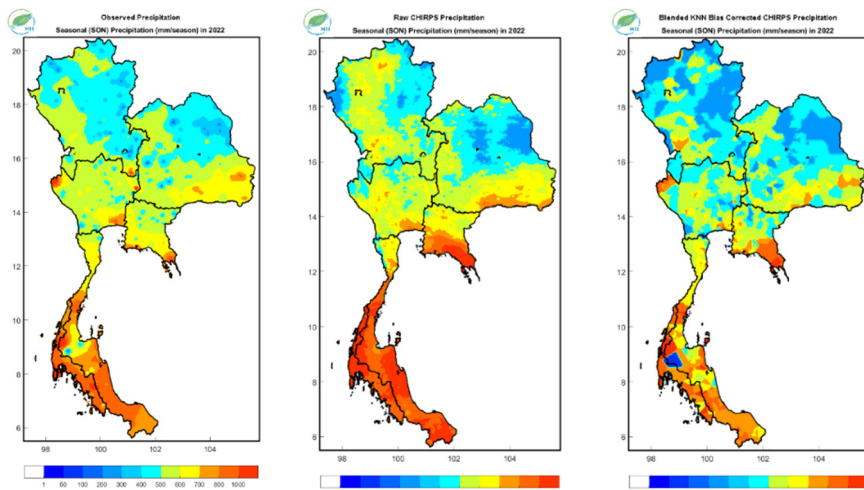
DJF 2022 spatial comparison (Figure 6) demonstrates Blended KNN BC CHIRPS effectiveness. Observed patterns showed minimal Northern/central precipitation and elevated Southern rainfall. Raw CHIRPS exhibited systematic overestimation (50-150 mm in central/Northeastern regions vs near-zero observations). Corrected dataset achieved remarkable accuracy, eliminating overestimation while preserving legitimate Southern signals (0-

25 mm Northern/central, 25-75 mm Southern). SON 2022 analysis (Figure 7) showed superior correction during post-monsoon transition. Observed patterns exhibited typical North-South gradient (50-150 mm Northern, 100-250 mm central, 200-400 mm Southern). Raw CHIRPS showed severe overestimation (300-500 mm central/Northeastern vs 100-200 mm observed). Correction transformed unrealistic uniform distribution to realistic patterns matching observations.



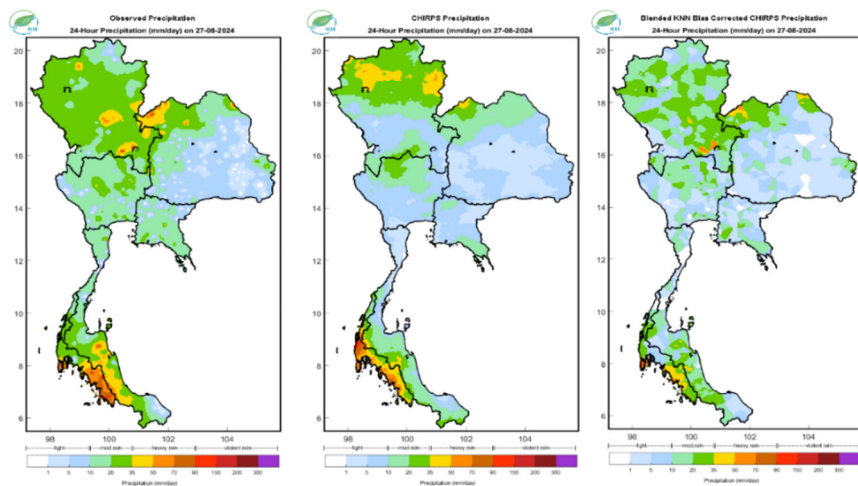
a) Observed b) Raw CHIRPS c) Blended KNN BC CHIRPS

Figure 6. Comparison of DJF seasonal rainfall map in year 2022



a) Observed b) Raw CHIRPS c) Blended KNN BC CHIRPS

Figure 7. Comparison of SON seasonal rainfall map in year 2022



a) Observed b) Raw CHIRPS c) Blended KNN BC CHIRPS

Figure 8. Comparison of daily rainfall map on 27 August 2024

Daily validation (August 27, 2024) during active monsoon conditions (Figure 8) confirmed operational reliability. Observed patterns showed intense central/eastern precipitation (100-250 mm), moderate Northern rainfall (50-150 mm), and variable Southern amounts (25-200 mm). Raw CHIRPS demonstrated significant overestimation (200-400 mm Northeastern vs 50-100 mm observed). Blended KNN BC CHIRPS exhibited exceptional daily accuracy, reducing overestimation by 70-85% in Northeastern regions while preserving genuine high-intensity signals.

These validations provide visual evidence that corrections not only reduce bias but preserve realistic geographical patterns essential for hydrological applications, demonstrating capability across diverse topographical and meteorological conditions.

5. Conclusion and discussion

This study successfully developed an integrated two-stage framework combining K-Nearest Neighbors machine learning with Kalman filter blending to enhance CHIRPS precipitation estimates across Thailand. Correlation coefficients increased from 0.42 to 0.94 (training) and 0.41 to 0.91 (validation), representing 124% and 122% improvements respectively, with consistent gains across BC CHIRPS and BBC CHIRPS methods. Regional analysis reveals differential effectiveness across climatic zones. South-Western region demonstrated dramatic improvements (221% training, 138% validation) while maintaining high performance (>0.93 correlation), likely reflecting topographical influences creating predictable bias characteristics. Northern and North-Eastern regions achieved substantial relative improvements (142% and 114% validation) despite weaker initial correlations. Bias reduction demonstrates capability to address systematic CHIRPS overestimation. Transformation from 34.03% overestimation to -10.00% (BC) and -10.78% (BBC) underestimation represents fundamental shift from wet to controlled dry bias. The framework shows exceptional adaptability across monsoonal

regimes, with DJF dry season overestimation of 588.27% (training) and 167.12% (validation) reduced by 95-99%.

Comprehensive spatial validation (Figures 6-8) confirms operational applicability beyond point-based statistics. Figure 6 illustrates DJF 2022 effectiveness in eliminating systematic overestimation (50-150 mm in central/Northeastern regions) while preserving legitimate Southern precipitation signals. Figure 7 demonstrates SON 2022 superior correction, transforming unrealistic uniform distribution (300-500 mm) to realistic patterns matching observations (100-200 mm). Figure 8 displays August 27, 2024 daily validation during active monsoon, confirming 70-85% overestimation reduction in Northeastern regions while preserving genuine high-intensity signals. The seasonal correlation maps (Figures 2-3) demonstrate consistent performance improvements across all 628 stations for both training and validation periods, while box plots analyses (Figures 4-5) show distributional transformation from systematic wet bias to controlled dry bias across all regions. Key innovations include 11-dimensional feature vectors, dual-update Kalman filtering, and systematic parameter optimization. This research demonstrates that machine learning integrated with optimal filtering significantly enhances satellite precipitation accuracy, with performance improvements enabling new hydrological applications and strong potential for broader tropical applications.

Future research directions include: (1) Investigating framework transferability to other satellite products (IMERG, GSMaP) and geographic regions with different climate regimes to assess generalizability across diverse conditions, (2) Developing real-time implementation protocols with automated model updating and quality control procedures for operational forecasting systems, (3) Integrating ensemble prediction methods to quantify uncertainty in bias-corrected estimates and provide probabilistic forecasts for risk assessment, (4) Exploring deep learning

architectures (LSTM, CNN, Transformers) for capturing complex spatio-temporal precipitation patterns while maintaining the controlled dry bias characteristic, (5) Evaluating framework performance specifically for extreme precipitation events and hydrological drought monitoring to assess applicability across the full precipitation spectrum, and (6) Extending

the methodology to sub-daily temporal scales (3-hourly, hourly) for flash flood early warning applications requiring high temporal resolution. These research avenues would enhance the framework's operational utility and broaden its applicability across diverse hydrological and climatological applications in data-scarce regions globally.

Author Contributions: Winai Chaowiwat conceptualized the study; Winai Chaowiwat and Jirayuth Srisat designed the methodology; Winai Chaowiwat and Jirayuth Srisat conducted the experiments; Winai Chaowiwat and Jirayuth Srisat analyzed the data; Kritanai Torsri and Kanoksri Sarinnapakorn revised the manuscript.

Acknowledgments: The authors would like to express their sincere gratitude to the Hydro-Informatics Institute (Public Organization) for their invaluable support in providing computer equipment and research facilities for this study.

Declaration: Conflicts of interest: The authors declare no conflicts of interest.

References

1. Huffman, G. J. et al. (2007), "The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales", *Journal of Hydrometeorology*, 8(1), 38-55.
2. Xie, P., and Xiong, A. Y. (2011), "A conceptual model for constructing high-resolution gauge-satellite merged precipitation analyses", *Journal of Geophysical Research*, 116(D21), D21106.
3. Funk, C. et al. (2015), "The climate hazards infrared precipitation with stations-a new environmental record for monitoring extremes", *Scientific Data*, 2(1), 150066.
4. Tian, Y. et al. (2009), "Component analysis of errors in satellite-based precipitation estimates", *Journal of Geophysical Research*, 114(D24), D24101.
5. AghaKouchak, A. et al. (2011), "Evaluation of satellite-retrieved extreme precipitation rates across the central United States", *Journal of Geophysical Research*, 116, D02115.
6. Tian, Y., and Peters-Lidard, C. D. (2010), "A global map of uncertainties in satellite-based precipitation measurements", *Geophysical Research Letters*, 37(24), L24407.
7. Chen, S. et al. (2013), "Evaluation of the successive V6 and V7 TRMM multisatellite precipitation analysis over the Continental United States", *Water Resources Research*, 49(12), 8174-8186.
8. Kühnlein, M. et al. (2014), "Improving the accuracy of rainfall rates from optical satellite sensors with machine learning-A random forests-based approach applied to MSG SEVIRI", *Remote Sensing of Environment*, 141, 129-143.
9. Hsu, K. L. et al. (1997), "Precipitation estimation from remotely sensed information using artificial neural networks", *Journal of Applied Meteorology*, 36(9), 1176-1190.
10. Hong, Y. et al. (2004), "Precipitation estimation from remotely sensed imagery using an artificial neural network cloud classification system", *Journal of Applied Meteorology*, 43(12), 1834-1853.
11. Lall, U., and Sharma, A. (1996), "A nearest neighbor bootstrap for resampling hydrologic time series", *Water Resources Research*, 32(3), 679-693.
12. Sharif, M., and Burn, D. H. (2006), "Simulating climate change scenarios using an improved K-nearest neighbor model", *Journal of Hydrology*, 325(1-4), 179-196.
13. Özger, M. et al. (2009), "Long lead time drought forecasting using a wavelet and fuzzy logic combination model: A case study in Texas", *Journal of Hydrometeorology*, 10(2), 396-407.

14. Karlsson, M., and Yakowitz, S. (1987), "*Rainfall-runoff forecasting methods, old and new*", *Stochastic Hydrology and Hydraulics*, 1(4), 303-318.
15. Baez-Villanueva, O. M. et al. (2018), "*RF-MEP: A novel Random Forest method for merging gridded precipitation products and ground-based measurements*", *Remote Sensing of Environment*, 217, 434-451.
16. Fang, J. et al. (2013), "*Evaluation of the TRMM 3B42 and GPM IMERG products for extreme precipitation analysis over China*", *Atmospheric Research*, 223, 24-38.
17. Kalman, R. E. (1960), "*A new approach to linear filtering and prediction problems*", *Journal of Basic Engineering*, 82(1), 35-45.
18. Crow, W. T., and Bolten, J. D. (2007), "*Estimating precipitation errors using spaceborne surface soil moisture retrievals*", *Geophysical Research Letters*, 34(8), L08403.
19. Pan, M., Li, H., and Wood, E. (2010), "*Assessing the skill of satellite-based precipitation estimates in hydrologic applications*", *Water Resources Research*, 46(9), W09535.
20. Sinclair, S., and Pegram, G. G. S. (2005), "*Combining radar and rain gauge rainfall estimates using conditional merging*", *Atmospheric Science Letters*, 6(1), 19-22.
21. Habib, E. et al. (2014), "*Evaluation of the high-resolution CMORPH satellite rainfall product using dense rain gauge observations and radar-based estimates*", *Journal of Hydrometeorology*, 15(4), 1506-1520.
22. Limsakul, A. et al. (2010), "*Observed changes in daily temperature and precipitation extremes over Thailand*", *Global and Planetary Change*, 74(2), 79-89.
23. Singhrattana, N. et al. (2005), "*Interannual and interdecadal variability of Thailand summer monsoon season*", *Journal of Climate*, 18(10), 1697-1708.
24. Takahashi, H. G. et al. (2010), "*High-resolution modelling of the potential impact of land surface conditions on regional climate over Indochina associated with the diurnal precipitation cycle*", *International Journal of Climatology*, 30(13), 2004-2020.
25. Loo, Y. Y. et al. (2015), "*Effect of climate change on seasonal monsoon in Asia and its impact on the variability of monsoon rainfall in Southeast Asia*", *Geoscience Frontiers*, 6(6), 817-823.
26. Jongjit, P. et al. (2013), "*Trends in precipitation extremes over Thailand during 1960-2009*", *Journal of Water and Climate Change*, 4(4), 367-377.
27. Thai Meteorological Department (2020), *Climate of Thailand*, Thai Meteorological Department, Ministry of Information and Communication Technology, Bangkok, Thailand.
28. Chotamonsak, C. et al. (2011), "*Projected climate change over Southeast Asia simulated using a WRF regional climate model*", *Atmospheric Science Letters*, 12(2), 213-219.
29. Katsanos, D. et al. (2016), "*Validation of a high-resolution precipitation database (CHIRPS) over Cyprus for a 30-year period*", *Atmospheric Research*, 169, 459-464.
30. Ayugi, B. et al. (2020), "*Evaluation of meteorological drought and flood scenarios over Kenya, East Africa*", *Atmosphere*, 11(3), 307.
31. World Meteorological Organization (WMO) (2018), *Guide to instruments and methods of observation. Volume I-Measurement of meteorological variables*. WMO-No. 8, World Meteorological Organization, Geneva.
32. Allen, R. G. et al. (1998), *Crop evapotranspiration-Guidelines for computing crop water requirements*. FAO Irrigation and Drainage Paper 56, Food and Agriculture Organization, Rome.
33. Durre, I. et al. (2010), "*Comprehensive automated quality assurance of daily surface observations*", *Journal of Applied Meteorology and Climatology*, 49(8), 1615-1633.
34. Peterson, T. C. et al. (1998), "*Homogeneity adjustments of in situ atmospheric climate data: a review*", *International Journal of Climatology*, 18(13), 1493-1517.
35. Wilks, D. S. (2011), *Statistical methods in the atmospheric sciences*. Academic Press.
36. Bonan, G. (2016), *Ecological climatology: concepts and applications*. Cambridge University Press.

37. Menne, M. J., and Williams Jr, C. N. (2009), "*Homogenization of temperature series via pairwise comparisons*", *Journal of Climate*, 22(7), 1700-1717.
38. New, M. et al. (2001), "*Precipitation measurements and trends in the twentieth century*", *International Journal of Climatology*, 21(15), 1889-1922.
39. Hastie, T. et al. (2009), *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media.
40. Tian, Y. et al. (2007), "*Multitemporal analysis of TRMM-based satellite precipitation products for land data assimilation applications*", *Journal of Hydrometeorology*, 8(6), 1165-1183.
41. Hofstra, N. et al. (2008), "*Comparison of six methods for the interpolation of daily, European climate data*", *Journal of Geophysical Research*, 113(D21), D21110.
42. Cover, T., and Hart, P. (1967), "*Nearest neighbor pattern classification*", *IEEE Transactions on Information Theory*, 13(1), 21-27.
43. Breiman, L. (1996), "*Bagging predictors*", *Machine Learning*, 24(2), 123-140.
44. Stone, M. (1974), "*Cross-validatory choice and assessment of statistical predictions*", *Journal of the Royal Statistical Society*, 36(2), 111-147.
45. Taylor, K. E. (2001), "*Summarizing multiple aspects of model performance in a single diagram*", *Journal of Geophysical Research*, 106(D7), 7183-7192.
46. Legates, D. R., and McCabe Jr, G. J. (1999), "*Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation*", *Water Resources Research*, 35(1), 233-241.