EnviroScienteae Vol. 20 No. 4, November 2024 ISSN 2302-3708 (online)
Halaman 433-448

AUTOMATED MULTI-MODEL PREDICTION AND EVALUATION FOR CONNECTING RAINFALL PREDICTION INFORMATION AND SINGLE-YEAR OPERATIONAL PLAN OF LAHOR-SUTAMI DAM

Otomatisasi Prediksi dan Evaluasi Multimodel untuk Menghubungkan Informasi Prediksi Curah Hujan dan Rencana Tahunan Operasional Waduk (RTOW) Bendungan Lahor-Sutami

Rikha Rizki Mahmudiah[1,2], Cahyo Crysdian[1*], Mokhamad Amin Hariyadi[1], Andang Kurniawan[2]

[1] *Program Studi Magister Informatika, Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim, Malang, Indonesia*
[2] *Stasiun Klimatologi Jawa Timur, Malang, Indonesia*
[*] *email:* cahyo@ti.uin-malang.ac.id

Abstract

There is a gap between existing climate information and the needs of annual dam operational planning. This study aims to demonstrate that the percentile approach currently used for planning is not optimal, especially now that automation has become more accessible. The purpose of this study is to design an automated forecasting and evaluation system based on 36 10-days rainfall projections using a multi-model approach. This approach comprises a percentile, ARIMA, ECMWF+ARIMA, IOD DMI regression, ERSST regression, and ensemble methods models. Additionally, this study aims to demonstrate how a verified, multi-model-based rainfall forecast can provide more reliable assurance for the annual operational planning of Lahor-SutamiDam, simulated operationally in November 2022 for the 2022/2023 planning cycle. Data utilized include historical 10-days rainfall data from 1991 to 2023, ECMWF raw and corrected model outputs, Nino-Dipole index, and global sea surface temperature. The verification method employs four criteria based on MAE and fit index. An operational simulation approach is used for training-testing period segmentation, while a 10-year window is applied to account for possible climate-change-induced shifts in relationships. Single linear regression is used to avoid overfitting. The automation system was developed using R-Statistics. Results indicate that the current approach is only optimal for 58% of locations. Superior methods identified include ECMWFcorrected, ERSST regression, and Ensemble models. A case study for 2022/2023 demonstrates that the forecast results outperform the existing plan for at least 78% of the projected periods.

Keywords: DAM operational plan; rainfall prediction; operational simulation

INTRODUCTION

As the central institution providing meteorological, climatological, and geophysical information in Indonesia (BMKG, 2019), the Agency of Meteorology, Climatology, and Geophysics (BMKG) produces seasonal forecasts used in dam operational planning. Knowledge of the rainy and dry seasons allows dam managers to anticipate fluctuations in water supply (Hurkmans *et al.,* 2023). BMKG's information includes forecasts of season's start, characteristics, rainfall, and duration, which are determined using rainfall data and specific rainfall limits to distinguish between the two seasons, with a threshold of 50 mm per 10-days for at least three consecutive 10-days.

Referring to the Area Annual Water Allocation Plan (RAAT WS) for the Brantas River Basin in the 2023/2024 period (Bestari, 2023), there are several gaps between the seasonal forecast information provided by BMKG and the practical dam operational needs. BMKG's seasons period vary each year, depending on the development of dynamic atmospheric conditions, while the single-year operational plan of dam has a fixed period and uses specific thresholds based on 35th and 65th percentiles, produced every November for the following year. This discrepancy highlights an opportunity to study forecast quality in the context of required operational scenarios, a topic not previously addressed.

Currently, seasonal forecast information produced by BMKG, especially the dry-normal-wet classification, is used as the input for single-year operational plan (RTOW) of dam (Bestari, 2023). Previous analyses found a high correlation, reaching 0.9, between basic rainfall data and the dam's inflow in the river basin filling the Lahor-Sutami dam. This correlation indicates the potential to use basic rainfall forecasts as a more precise and relevant input in dam operational planning. Recognizing the importance of forecast accuracy, this study explores the use of basic rainfall forecasts as a more suitable alternative to seasonal characteristics currently used in dam operations.

Various models used for rainfall forecasting include ARIMA, ECMWF, and regression models. ARIMA has an average verification rate of 60% when used for monthly rainfall forecasts in East Java (Kurniawan *et al.,* 2022). The ECMWF model has a lower error rate and higher correlation coefficient compared to the HyBMG univariate model, especially under extreme rainfall conditions in Kebumen (Ruslana *et al.,* 2021). Rainfall forecasts in Indonesia also consider dynamic atmospheric conditions on regional, local, and global scales (Abdullah, 2021). For instance, ENSO activity, observed through the ONI index, shows a strong positive correlation with rainfall in Central Java. Besides ENSO, sea surface temperatures influence rainfall with varying correlations across regions, while variables like SOI, SST, ENSO, and IOD also affect rainfall in places like Pekanbaru. Each of these variables has a unique influence across different regions, indicating the potential for using a multi-variable approach in rainfall prediction (Ardhitama and Sholihah, 2013).

The BMKG Research and Development Center has created Hybrid BMG (HyBMG), a statistical prediction model based on climate forecasts as a preventive measure to reduce the impact of hydrometeorological disasters such as drought and high rainfall. HyBMG is the result of climate modeling which is expected to be effective for areas with monsoon rainfall patterns, including East Java. Rainfall forecasting in HyBMG uses the ARIMA, ANFIS (Adaptive Neuro Fuzzy Inference System), wavelet-ARIMA, and wavelet-ANFIS methods. Research conducted by (Kurniawan *et al.,* 2022) evaluated the extent to which the HyBMG prediction model can provide accurate monthly rainfall forecasts for the East Java region, with a focus on verification based on the percentage of accuracy in categorizing rainfall according to the SNI 8196:2015 standard. The model output is analyzed in the context of Neutral, El Niño and La Niña conditions using the Nino 3.4 index (Oceanic Nino Index/ONI) to determine ENSO conditions. The ARIMA method has the highest verification rate among other models, namely 60% on average. In general, the HyBMG model output has verification below 60% during the La Niña phase, while in the El Niño phase only ARIMA reaches a verification level of between 60% and 70%. Meanwhile, the results of research conducted by Noviasari *et al.* (2023) shows that the AR1 model is able to produce rainfall forecasts with similar patterns to observational data in Madura, East Java.

Another model used by BMKG is ECMWF with the latest version ECS4 (ECMWF System-4), is a refinement of the previous model, namely System-3, with the difference being the use of the ocean model from the Nucleus for European Modeling of the Ocean and the NEMO/NEMOVAR data assimilation system. In addition, this model utilizes ERA-Interim re-analysis as initial conditions for the atmosphere and uses the latest integrated forecasting system cycle. All of this combined improves the capabilities of ECS4 over previous versions in several aspects, including horizontal, vertical resolution, number of ensemble members, and prediction range. Previous research by Weisheimer and Palmer (2014) showed that ECS4 has probabilistic prediction capabilities for seasonal rainfall that vary depending on the region and season period considered. They concluded that these predictions have a level of reliability that is still very useful in tropical regions during the Winter and Summer periods. A study on how reliable ECMWF is in forecasting basic rainfall in extreme conditions, especially during flood events in Kebumen, was carried out by Ruslana *et al.* (2021). The sensitivity of the HyBMG and ECMWF univariate models was tested through visual methods of spatial fit, simple correlation, and RMSE to compare their performance. One of the results obtained shows that the ECMWF raw output provides the best results compared to other models. ECMWF is a dynamic model that has a major role in producing summaries and explanations for basic and monthly operational forecasts. This is because this model takes into account dynamic factors from atmospheric circulation and sea surface temperature, including climate variability such as La Nina and El Nino. The analysis was carried out based on ZOM areas involving a sample of 1 rain post in each ZOM. The forecast results are evaluated with RMSE and correlation coefficient which shows that ECMWF is the best in terms of small RMSE among other models.

Rainfall in East Java is also influenced by ENSO activity, dipole mode and sea surface temperature (Deman *et al.,* 2022). Regression analysis allows identifying the relationship between two variables (Maulita and Nurdin, 2023). Previous study conducted by Ardhitama and Sholihah (2013) attempted to improve the quality of weather forecast results in Pekanbaru by involving a simulation of forecasting the amount of rainfall in Pekanbaru City for 2011 and 2012 using input predictors such as SOI, SST, Nino 3.4, and IOD. The method used is multiple linear regression. Evaluation of rainfall forecast results is carried out using Root Mean Square Error (RMSE) and Standard Deviation (SD). The main objective of this study is to identify the most significant predictor factors influencing rainfall conditions in Pekanbaru. The research results show that weather conditions, especially rainfall in Pekanbaru, are influenced by global, regional and local factors. It was found that SOI predictors have a relationship with rainfall that has a high level of correlation. This shows that the multiple linear regression method can be an option to determine the influence of several variables on rainfall in Pekanbaru. Study conducted by Stockdale *et al.* (2010) revealed that seasonal predictions depend on changes in weather probability, especially related to slow changes such as sea surface temperature anomalies from El Niño-Southern Oscillation (ENSO). However, seasonal weather is also influenced by many other factors and internal variations in the atmosphere, so comprehensive models are needed to identify what can be predicted.

In ongoing operational activities, BMKG forecasts use several existing model options such as ARIMA, ECMWF, and atmospheric index regression. In reality, there is no one model that consistently provides the best results in all regions of Indonesia, so adjustments for which model is the best one in each place are left to forecasters in related regions. Rather than forming a new model, the ensemble method allows the use of existing models to optimize existing forecasts.

Ensemble techniques allow the combination of multiple models with different performances to produce more accurate forecasts (Weyn *et al.,* 2021; Specq *et al.,* 2020; Pakdaman *et al.,* 2022). Studies have shown that using ensembles improves forecast quality (Gu *et al.,* 2022; Kullahci and Altunkaynak, 2023; Kundu *et al.,* 2023; Anggraeni *et al.,* 2018), with the performance of each modelinfluencing its contribution to the ensemble. Assigning variable weightsyields better results rather than uniform ones, as each model has unique qualities (Wei *et al.,* 2022). The inherent uncertainty in General Circulation Models (GCMs) affects the accuracy of projections, and multi-model ensemble approaches offer promising, replicable results in reducing projection uncertainty (Raju and Kumar, 2020). The use of ensemble techniques aims to improve accuracy and minimize discrepancies in forecast values. This research will contribute to establishing standards and criteria for more accurate and reliable seasonal forecasts, applicable to practical needs like dam operation planning and water resource management.

Previous research revealed that rainfall is an important process that influences the water cycle (Chowdary and Anbarasi, 2020; Zhai *et al.,* 2022), so it is necessary to predict rainfall patterns to anticipate floods. This research was carried out using an ensemble stacking technique involving the basic models of Naive Bayes, Decision Tree, KNN, and SVM. Meanwhile, the ensemble was built using the Deep Neural Network method. The results of this study show a better level of accuracy and specificity in the ensemble results, namely 80% and 97%.

Important point found by Raju and Kumar (2020) that selection of an appropriate Global Climate Model (GCM) needs to be carried out in impact studies. Performance measurements and appropriate decision-making techniques play an important role in the ensemble that will be built. Uncertainty in GCMs affects projection accuracy. The multi-model ensemble approach provides promising and replicable results in reducing uncertainty in projections.

According to Wei *et al.* (2022), there are some effects of adding a low-performance model on forecast accuracy with an ensemble model. Different weighting techniques showed a completely different result on the ensemble forecast. The ensemble is formed using linear regression with the weights between models being the same and varying which are assumed to be inversely proportional to the error variance (MAE). The results of this study show that ensemble forecasts with varying weights are better than equal weights. It is happens not only for rainfall prediction, but also temperature prediction with various variations in time lag also shows a similar thing where the use of varying weights is better. The varying weight method can overcome the problem of decreasing accuracy due to adding models with low performance (Jose *et al.,* 2022).

A monthly rainfall forecast model was built by Gu *et al.* (2022) using an ensemble technique to combine several basic models which each have their own advantages. This research was conducted in the Taihu Watershed, China. The basic models used are k-nearest neighbors (KNN), extreme gradient boosting (XGB), support vector regression (SVR), and artificial neural networks (ANN). The ensemble is built using the weighting of each basic model. The ensemble monthly rainfall forecast produces the lowest MAE 41.65, indicating that the presence of the ensemble improves forecast performance compared to the output of the basic model. More specifically, precipitation forecasts using stacking ensembles are better in spring and winter than in summer and autumn.

Fathi *et al.* (2019) combined 9 daily rainfall forecast models in Iran to build an ensemble model based on average and weighted average. The weight of each ensemble member is determined from previous performance in the training period. The performance of the baseline and ensemble models is compared using RMSE and ACC. This research shows that the ensemble process is effective in improving forecast quality where the best results are shown by the ensemble weighted average (ENSWM) with RMSE 2.8-3.3.

Not just forming an ensemble, this forecast is adjusted to resemble the scenario used in dam operational planning so that the results of this research are more applicable. This adjustment is intended to overcome existing gaps. The scenario in question is a forecast covering the next year starting from December to November according to the schedule for holding annual planning meetings. So far, dam planning usually uses forecast information about seasonal characteristics. In the latest data processing that compares dam inflow discharge with basic rainfall, it is known that the two have a quite promising correlation. Instead of continuing to use seasonal properties, this research offers baseline rainfall forecasts as a more relevant option for input in dam operational planning.

Creating seasonal forecasts for East Java using large volumes of historical data requires efficient approaches, such as automation with open-source software. Automation involves algorithms and big data techniques to uncover patterns within complex datasets, enabling faster periodic forecast updates and adaptive responses to changing weather conditions. In line with recent standards, BMKG now recommends open-source software for seasonal forecasting due to its sustainability and accessibility. In this context, R-Statistics, with R-Clim features, emerges as a practical option.

Based on the above, this research focuses on optimizing seasonal forecasts within the RTOW framework by implementing ensemble techniques to improve forecast relevance for dam operations. This study designs an automated 36 10-days rainfall prediction system using a multi-model approach, including percentile, ARIMA, ECMWF+ARIMA, IOD DMI regression, ERSST regression, and ensemble models. Additionally, it aims to demonstrate how a verified multi-model forecast approach can provide more reliable certainty for the annual operational plan of the Lahor-Sutami dam, simulated for the 2022/2023 planning period.

MATERIALS AND METHODS

The material used in this study is rainfall data based on seasonal zones, abbreviated as ZOM. Out of the 74 ZOM in East Java, 12 ZOM intersect with the area that fills the Lahor-Sutamidam. The analysis used to identify the intersecting ZOM was conducted using the zonal statistics feature from the raster package in R-Statistics. First, the area that serves as the dam catchment is converted into a binary raster, where 1 represents the catchment area and 0 represents areas outside the catchment. This raster is then aligned in resolution and extent with the existing ZOM raster. The analysis resulted in identifying 12 ZOM that cover the dam catchment, with ZOM 36 [36.40%], 40 [29.68%], 34 [9.67%], 37 [8.48%], 33 [6.83%], 20 [6.10%], and the remaining 6 ZOMs [30-31-35-38-39-41] each contributing less than 1%. A visual representation of the overlap between the dam catchment area and ZOM in East Java can be seen in Figure1 below.

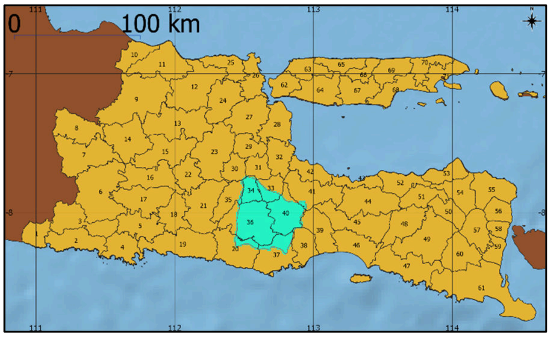

Figure 1. Dam Catchment Area (Green) with ZOM in East Java (orange)
Source: East Java Climatological Station 2024

The time dimension of rainfall used in this study is 10-days, spanning from the first 10-days of 1991 to the 36th 10-days of 2023, which is equivalent to 1188 10-days periods or 33 years. The rainfall data used is blended rainfall data, which combines ground-based rain gauge measurements and satellite-based rainfall estimates. This data is operationally used by official government agencies in Indonesia for seasonal monitoring and forecasting. The rainfall predictor data used consists of both raw and corrected ECMWF outputs, which are available operationally on the 8th of each month. Rainfall and its predictions are commonly known for having a connection with sea surface temperature (Mahera *et al.*, 2023). Other predictor data used include the Oceanic Niño Index (ONI) from CPC NCEP NOAA, representing the Pacific Ocean index, and Dipole Mode Index (DMI) from PSL NOAA, representing the Indian Ocean index. Lastly, the Sea Surface Temperature Anomaly Index from ERSST NOAA IRIDL is also used, which will be processed using principal component analysis (PCA) as a representation of the long-range teleconnection, commonly recognized in climatology. All sea surface temperature indices are on a monthly timescale.

This study uses an operational simulation approach, with the simulation years spanning from 2014 to 2022 (9 years). The year 2014 marks the first year that ECMWF was used as one of the prediction sources by government climate forecasters, while 2022 is the last year for which data was available at the time this research was conducted. This simulation approach means that the data is not split using the 80:20, 70:30, or similar ratios for training and testing. The data cut is made based on the actual availability of data when an annual dam operational plan (RTOW) is created, which occurs every November. The assumption is the meeting for the operational plan is held in mid-November, so the rainfall data is available up to the 33rd 10-days (end of October) of the current year. The ECMWF predictions for November through May of the following year are available, and the ocean-related indices are available up to the anomaly analysis of September or JAS (July-August-September) for the Oceanic Niño Index (ONI), which is typically delayed by two months.

To evaluate the accuracy of a forecast, this study proposes the use of four criteria. The first criterion is the average fit index. This fit index is similar to the confusion matrix used in other studies, but the table employed uses the 10-day rainfall categories according to the official government rainfall product legend (Kurniawan *et al.,* 2022), with a tolerance of one category to be considered a match. This method is used here and also represents the operational simulation concept. In addition to the average fit index, the absolute minimum value for each year is also usedas in operational settings, a model that experiences a decrease in fit index will attract considerable attention from forecasters. Besides the fit index, the Mean Absolute Error (MAE) also used, which represents a superior numerical approach due to its simplicity compared to Root Mean Square Error (RMSE) and also reflects a non-categorical approach (Willmott and Matsuura, 2005).

To generate forecasts for the next 36 10-days periods (from 10-days period 34 to 10-days period 33 of the following year), several approaches are used as follows. The first scenario, or percentile-based scenario, is currently used in operational status, utilizing the 35th percentile (dry), 50th percentile (normal), and 65th percentile (wet) to determine the planned discharge (Bestari, 2023). An alternative proposed in this study is the use of the ARIMA, ECMWF+ARIMA, ONI-IOD regression, ERSST Principal Component Regression, and MAE-and-fit index-based Ensemble. Since the rainfall pattern in East Java is monsoonal (Kartika *et al.,* 2021), the difference from the moving average is updated annually for ARIMA, enabling it to directly generate predictions for the next 36 10-days periods. However, this is not the case for ECMWF, which requires supplementation with ARIMA. The sea surface temperature-based regression uses a 10-year moving window approach, which accommodates the potential impact of climate change on atmospheric-ocean parameter relationships. The regression used is simple linear regression to avoid overfitting due to the short window size. The teleconnection lag approach is represented using up to a 10-month lag. This means that a sea surface temperature index up to October of previous year is considered potential for use as a predictor for the 36 10-days period rainfall forecast, calculated from the end of October of the current year.

All automation, including the generation of prediction values, verification, and plotting, is performed using R-Statistics software with the help of packages such as forecast, raster, and ncdf4, primarily for processing ARIMA with auto.arima (Hyndman and Khandakar, 2008) and sea surface temperature data in net-CDF extension. The total number of 36 10-days period predictions generated is 450, consisting of 3 percentile-based scenarios, 1 ARIMA scenario (auto.arima), 2 ECMWF (raw/corrected)+ARIMA scenarios, 24 ONI DMI regression scenarios, 120 principal component regression scenarios (10 components), 150 MAE-based ensemble scenarios (Mahmudiah *et al.,* 2019), and 150 fit index-based ensemble scenarios. The ensemble system is performed using a weighting approach (Fathi *et al.,* 2019) where each weight divided by the total weight, as in previous studies, and the fit index value divided by the total weight. Once again, the weights will be adjusted annually following the operational simulation principle, so that the evaluation results (MAE and fit index) for 2014 will be used for the 2015 ensemble, the 2014-2015 evaluation for the 2016 ensemble, and so on, until the 2014-2021 evaluation is used for the 2022 ensemble.

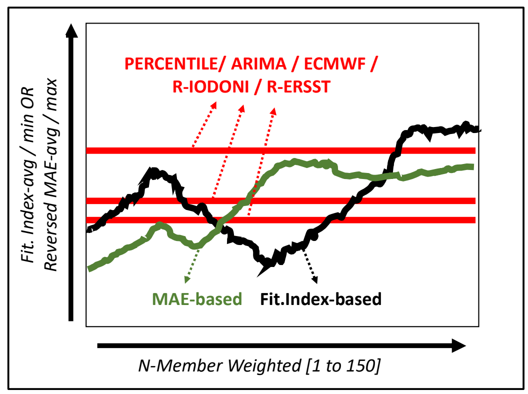

Figure 2. Graph Format for Visualizing Model Output and Verification Results

The multi-model approach, which provides multiple model outputs and verification results, is handled by proposing a standard presentation format as shown in the figure above. The green and black lines represent the addition of ensemble members (from 1 to 150, black and green), while the red line shows the verification of non-ensemble outputs. A challenge that arises is the large number of red lines, which can become overwhelming, so the solution is to group the plots into 5 categories (percentile, ARIMA, ECMWF(+ARIMA), IOD ONII/R-IODONI regression, and ERSST/R-ERSST principal component regression). The proposed scheme is as shown in the figure below. The Y-axis represents the average or minimum fitness index. In other plots, the Y-axis also shows the average and maximum MAE. Since a model is considered better if it has a high fitness index and low MAE, a reversed Y-axis is used for plotting so that forecasters can interpret the results more uniformly, with higher lines indicating better verification of the model output. This plot is expected to provide an intuitive overview to forecasters regarding the impact of increasing ensemble members on the verification results, while also offering a comparison of verifications across the different methods.

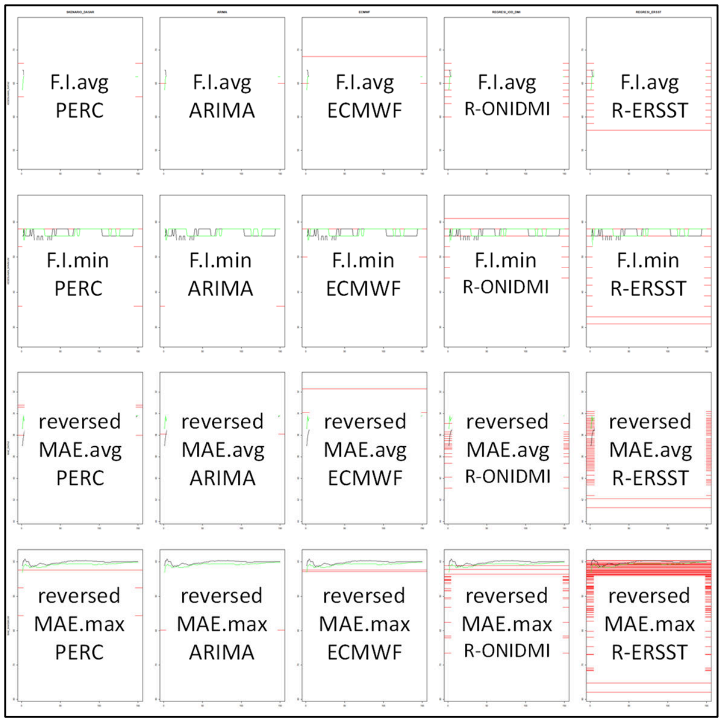

Figure 3. Overview of Each Model's Output in Specific Seasonal Zone (ZOM)

To simplify the summarization of model outputs, a table containing scenarios that perform best according to one of the four criteria (highest average fitness index, highest minimum fitness index, lowest average MAE, OR lowest maximum MAE) is automatically generated to assist forecasters. This result table is also simulated operationally for the 2022/2023 single-year dam operational plan. The 2022/2023 data used includes both planned inflows and actual observed inflows. Using this data, the study will demonstrate the potential use of the 36 10-days-ahead rainfall forecasts in the single-yeardam operational plan. The rainfall forecast results, based on the recommendation table generated from the automated system, are considered to be handed over to the forecaster. In this case, the forecaster is represented by a random selection process acting as a blind forecaster. This random selection process is repeated through bootstrapping 10,000 times, and the median is then calculated to represent the final result. The forecast results are then compared with the observed rainfall by constructing an empirical equation based on wet and dry periods. The rainfall data is converted into discharge values, enabling a direct comparison with the existing discharge plan. The prediction output is then compared by calculating the absolute difference for each 10-day period with the existing plan. It is important to emphasize that using the operational simulation approach means the rainfall prediction results obtained will be identical to the results that would have been produced in November 2022.

RESULTS AND DISCUSSION

This study automates various steps in generating 36 10-days-ahead predictions using a multi-model approach and verification. For quality control, each prediction output and observation are visualized in graphs like the one below, allowing forecasters to assess the alignment of each model and directly observe the results. This plotting process is automated. In this study, as many as 5,400 images like the one below are automatically generated. The results, in the form of forecast values based on operational simulations, are also saved in CSV file tables.

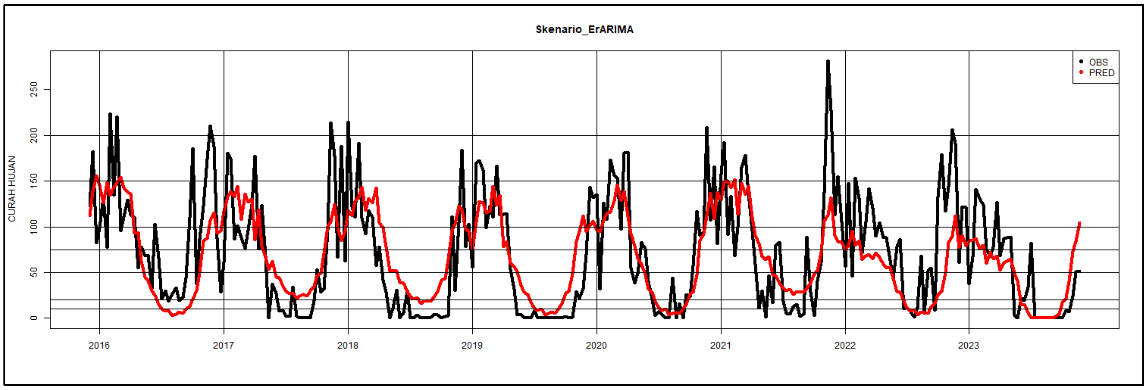

Figure 4. Time Series of Model's Prediction (Red Line) vs Observation (Black Line) for Each ZOM

The forecast results are then verified for each ZOM, producing the suitability index values and MAE for each year from 2014 to 2022 (9 years). From these values, the average suitability index, minimum suitability index, average MAE, and maximum MAE are calculated and automatically generated. These four criteria are then plotted, as shown in the example image below. The example below represents ZOM=36, where it can be seen that the ECMWF method (middle box, column 3) excels in the fit index average and MAE average. Meanwhile, when looking at the highest fit index minimum, the regression method based on DMI outperforms, and when examining the MAE average, the ensemble method is identified as the one most effective in minimizing the maximum MAE.

Based on the data in Figure 5, the system will automatically generate a recommendation table that can be provided to the forecaster. It is common practice in climate operations that the final determination of a climate model is left to the forecaster, as climate information sometimes has socio-economic implications. Therefore, questions such as which of the four criteria above is more important become less relevant. It is a consensus that the role of automation should be limited to providing recommendations to the forecaster. It is also less relevant to inquire about the fit index or MAE of each forecast model. The purpose of this automation process is to identify the most appropriate model for each ZOM, rather than using a single model that is considered to represent the majority at the expense of specific ZOM interests.

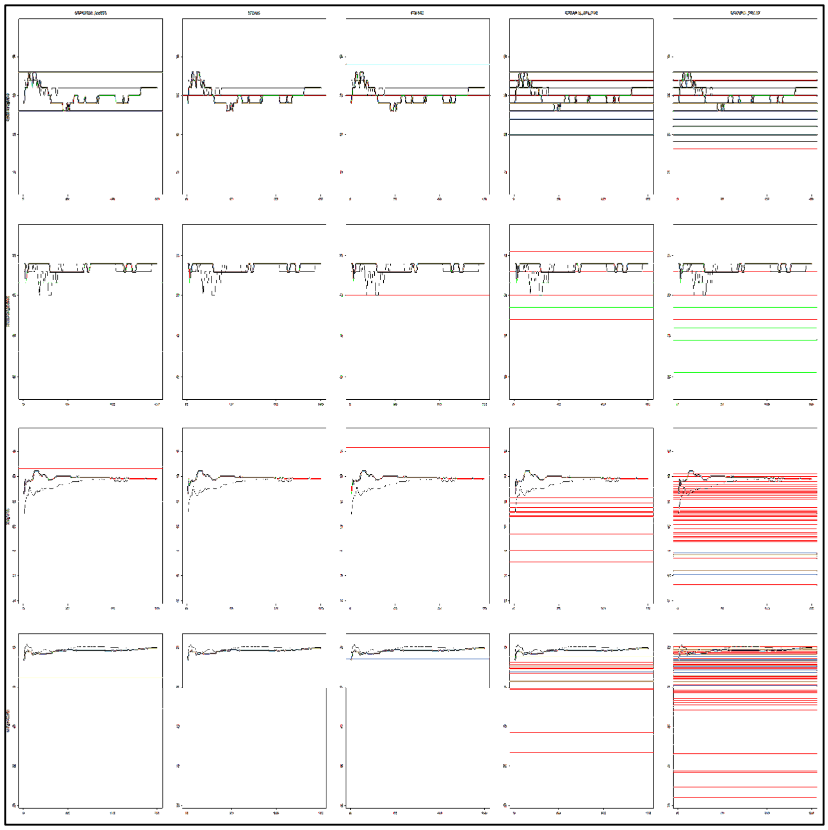

Figure 5. Overview of Each Model's Output (Left to Right: Percentile Based, ARIMA, ECMWF-ARIMA, ONI-DMI Regression, ERSST Regression) for ZOM=36

Table 1. Example of Best Model Recommendation and its Verification Results for ZOM=36

RECOMMENDATION ZOM=36	F.Index AVG	F.Index MIN	MAE AVG	MAE MAX
Ens_F.Index-based, M=4	66	56	35.5	**39.1**
R_DMI_Aug	68	**61**	34.9	42.2
ECMWF_cor+ARIMA	**69**	58	**31.7**	42.4

The use of the four criteria can also be justified by looking at the figure below, which shows that the approach using the minimum Fit Index provides an overview of models that have a lower range of prediction failure each year. The figure below is a plot of the fit index for each year from ZOM 40. The blue line represents the fit index values for each year. It can be seen that this approach ensures the fit index value does not drop below 60%. On the other hand, the approach using the maximum MAE (represented by the orange line) shows that this approach consistently results in forecasts with an MAE that does not exceed 40 mm. This is important because, in essence, even though a model may have a good average fit index or average MAE, any drop in performance at certain times may be a noteworthy issue. In the case of ZOM=40, the drop in performance that occurred in 2022 can be partly explained by the triple-dip La Niña phenomenon. It turns out that the ERSST regression method, which relies on ocean-atmosphere teleconnections, and the Ensemble method were able to produce more reliable forecasts than ECMWF.

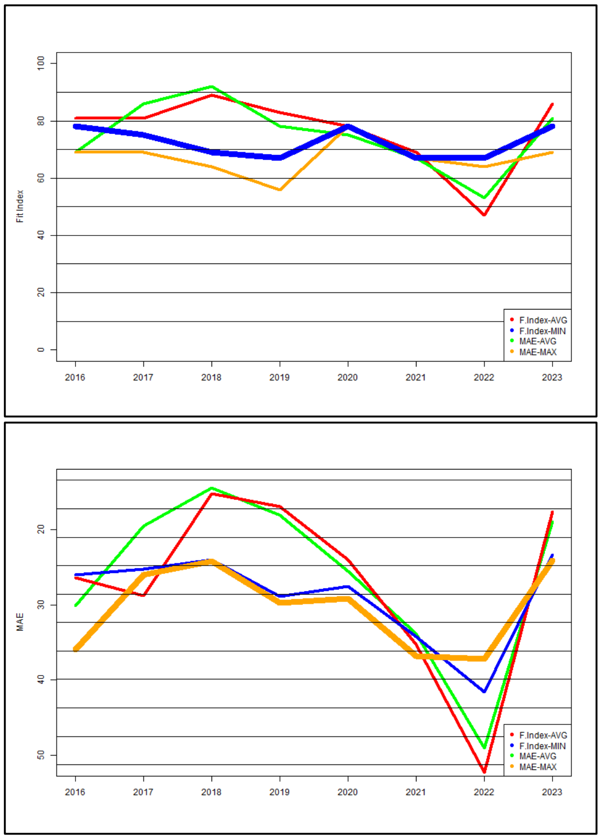

Figure 6. Verification Results for Fit-Based (Above) and MAE-Based (Bottom) Ensemble Forecasts for ZOM=40

The table below shows that the percentile-based prediction approach is not optimal. Although this approach was the best based on the four criteria in 7 ZOM (58%), it was not optimal for predicting rainfall in 3 ZOM with the highest proportion of the dam inflow area (34, 36, and 40). The three methods that can be considered most optimal for forecasting rainfall over the next 36 10-days are Reg-ERSST (11 ZOM), ECMWF (10 ZOM), and Ensemble (9 ZOM). Regarding the three ZOM with the highest proportions, the ECMWF method was the most reliable, followed by Ensemble and Reg-ERSST. The table below also indicates that the information from the analysis, or the data available at the time of preparation (rather than the forecast itself), from ONI and DMI is not crucial for predicting rainfall over the next 36 10-days. Rather, the SST teleconnection pattern needs further investigation related to long-term predictions. The next question is which ECMWF model is more optimal to use. Based on the evaluation of the four criteria, the ECMWF corrected approach was only suboptimal for ZOM=35, while ECMWF raw was optimal only for ZOM=20. This further emphasizes the importance of model correction, as supported by previous studies.

Table 2. Summary of Forecast Verification at Each ZOM

ZOM	PERCENTILE	ARIMA	ECMWF	Reg. ONI-DMI	Reg. ERSST	ENS
20	1	1	1	0	1	1
30	1	0	1	1	1	1
31	1	0	1	0	1	1
33	1	0	1	0	1	1
34	0	0	1	0	1	0
35	1	0	0	1	1	1
36	0	0	1	1	0	1
37	0	0	1	0	1	0
38	0	0	1	0	1	0
39	1	1	1	1	1	1
40	0	1	1	0	1	1
41	1	0	0	0	1	1
TOTAL	**7**	**3**	**10**	**4**	**11**	**9**
%	**58%**	**25%**	**83%**	**33%**	**92%**	**75%**

To demonstrate how 36 10-days optimized rainfall predictions can enhance the operational plan for the dam, a case study of the Sutami-Lahor Dam was conducted. The Sutami-LahorDam is assumed to receive no inflow from other dams, so the 36 10-days rainfall directly impacts the inflow discharge of the dam. The data used in the case study includes the Single-year Dam Operational Plan (RTOW) and its evaluation from the 2022/2023 Annual Water Allocation Plan (RAAT). It is hoped that through this demonstration, stakeholders will be more motivated to collaborate on data sharing. The simulation for the preparation of the RTOW 2022/2023 was then carried out by calculating the median of randomly selected models considered optimal based on the four criteria (Fit Index AVG, Fit Index MIN, MAE AVG, and MAE MAX), with calculations repeated up to the 2021 limit to align with the operational simulation principle.

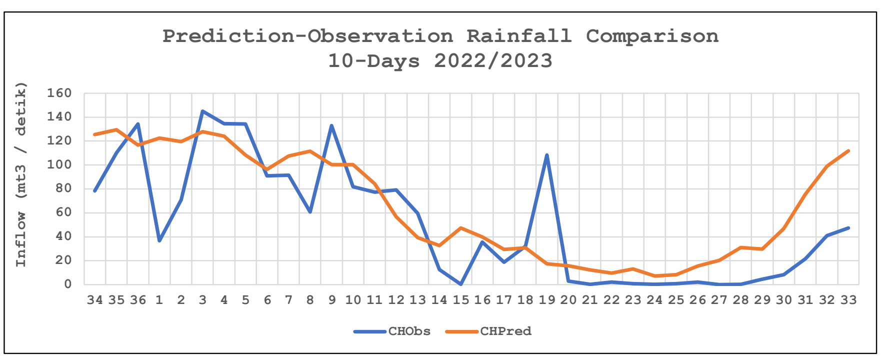

Figure 7. Comparison of Prediction (Orange Line) and Observation (Blue Line) Rainfall in Sutami-Lahor Dam

The rainfall results for each ZOM are then multiplied by their respective weights and accumulated to form the Rainfall Prediction for 2022/2023 over 36 10-days. The multiplication of weights and rainfall values for each ZOM is also applied to the observation data, allowing for comparisons as shown in the figure above. It can be observed that the predicted and observed values follow an identical pattern. There are several spikes in the observed rainfall, expected caused by the influence of

Madden-Julian Oscillation on East Java rainfall. A noticeable difference appears to widen and remain constant from 21st 10-days to 33rd 10-days. However, the next question that needs to be answered is whether this prediction is better compared to the existing plan that has already been implemented.

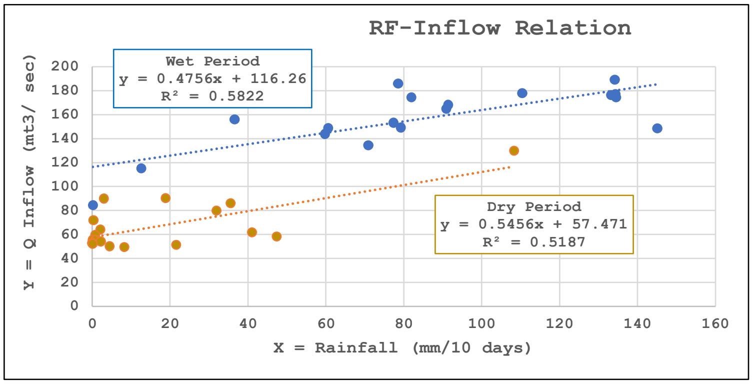

Figure 8. Rainfall-Inflow Relation in Sutami-Lahor Dam

In the RTOW, the required value is not rainfall but inflow discharge (Q inflow in cubic meters per second). In this study, an empirical regression formula is used to convert rainfall in millimeters per 10-days to inflow discharge in cubic meters per second. Since the RTOW separates the rainy period (December-May) and the dry period (June-November), the empirical formula is also divided accordingly. Due to limited information, the data used comes from the rainfall and discharge records in the RAAT 2022/2023. The empirical regression used in this study is shown in the figure above. It can be seen that the relationship between rainfall per 10-days and inflow during the wet period differs from the dry period. The two linear regression constants indicate that in wet conditions, even without rain, the inflow discharge is about 116 cubic meters per second, which is much higher than the dry period, where the inflow is around 57 cubic meters per second. This study also encourages further research on the relationship between rainfall and inflow using longer datasets and more comprehensive interpolation methods.

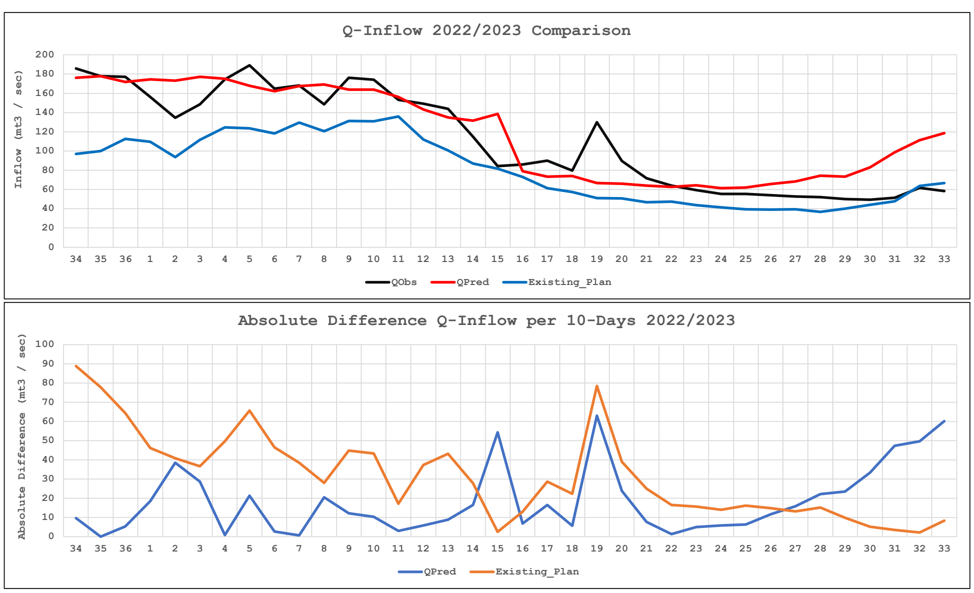

Figure 9. Inflow Comparison (Above) and Absolute Difference of Inflow (Bottom) in Sutami-Lahor Dam 2022/2023

The predicted rainfall, which has been converted into discharge, is then compared with the existing plan. It can be observed that the RTOW for 2022/2023 would have a smaller absolute difference if the automatically verified multi-model 36 10-days rainfall prediction is used. Out of the 36 10-days for which the discharge is planned, the multi-model results are only worse in 8 of the 36 planning 10-days. One of these occurred in 15th 10-day, and the other seven occurred in 27th – 33rdat the end of the prediction period. This demonstrates the potential for more accurate planning when the optimized 36 10-days rainfall prediction is used.

CONCLUSIONS

The first conclusion drawn from this study is that the status quo method, which links seasonal rainfall characteristics to determine streamflow conditions [3], proves to be no better than relying on a multi-model approach and direct rainfall predictions over 36 10-days. The first piece of evidence is that forecasts based on the 35th, 50th, and 65th percentiles are not superior to the multi-model approach. These percentile-based forecasts only emerge as the best approach in a few seasonal zones. The output from the EMCWF model, combined with ARIMA and ensemble methods, shows more promising results in terms of fit criteria (average and minimum) and Mean Absolute Error (MAE, average and maximum). The second piece of evidence comes from direct simulations using random forecasters provided with model recommendations from evaluations, which show lower absolute deviations compared to the annual plans used in 2022/2023. The use of an operational simulation approach in this study strengthens the aforementioned evidence, as it accurately reflects real conditions and employs training-testing data segmentation consistent with the conditions in November each year (including 2022). Based on this evidence, it is recommended to shift from using seasonal rainfall characteristics to forecasting rainfall over 36 pentads to enhance the value of operational dam planning. Further research is necessary to establish a more empirical formula for the rainfall-streamflow relationship and to integrate dam planning with projected rainfall data (as opposed to predictions) under future climate scenarios. It is also suggested that similar studies be conducted for other dams.

ACKNOWLEDGEMENTS

We would like to thank the Coordination Team of Water Resources Management (TKPSDA) of Brantas for making the discussion on linking climate information with the single-year operational plan data for dam possible. This research was conducted as a result of the insightful discussions that took place in that forum.

REFERENCES

Abdullah, S. E. A. N. (2021). Analisis Hubungan Indeks Nino 3.4 dengan Curah Hujan di Jawa Tengah. *Buletin Meteorologi, Klimatologi, dan Geofisika,* 2 (5), 24-30. P-ISSN 2716-0130.

Anggraeni, D., Kurnia, I. F., & Hadi A. F. (2018). Ensemble averaging and stacking of ARIMA and GSTAR model for rainfall forecasting. *IOP Conf. Series: Journal of Physics: Conf. Series.* doi:10.1088/1742-6596/1008/1/012019.

Ardhitama, A. & Sholihah, R. (2013). Model Simulasi Prakiraan Curah Hujan Bulanan pada Wilayah Riau dengan Menggunakan Input Data SOI, SST, NINO 3.4, dan IOD. *Jurnal Sains & Teknologi Modifikasi Cuaca*, 14 (2), 95-104.

Badan Meteorologi Klimatologi dan Geofisika. (2019). Peraturan Badan Meteorologi Klimatologi dan Geofisika Nomor 7 Tahun 2019 tentang Penyediaan dan Penyebaran Prakiraan Musim. Jakarta.

Bestari, A. (2023). *Laporan Akhir: Penyusunan Alokasi Air WS Brantas*. Surabaya: Balai Besar Wilayah Sungai Brantas.

Chowdary, P.S.M. & M. Anbarasi. (2020). Enhanced Rainfall Predictions using Stacking Technique. *Journal of Emerging Technologies and Innovative Research*, 7(7).

Deman, M. H., Koppa, A., Waegeman, W., MacLeod, D. A., Singer, M. B., & Miralles D. G. (2022). Seasonal prediction of Horn of Africa long rains using machine learning: The pitfalls of preselecting correlated predictors. *Water*, 4:1053020. doi: 10.3389/frwa.2022.1053020.

Fathi, M., Azadi, M., Kamali, G., & Meshkatee, H. (2019). Improving Precipitation Forecasts over Iran using A Weighted Average Ensemble Technique. *J. Earth Syst. Sci.*, 128 (133). doi: 10.1007/s12040-019-1145-2.

Gu, J., Liu, S., Zhou, Z., Chalow, S. R. & Zhuang, Q. (2022). A Stacking Ensemble Learning Model for Monthly Rainfall Prediction in the Taihu Basin, China. *Water*, 14. doi: 10.3390/w14030492.

Hyndman, R. J. & Khandakar, Y. (2008). Automatic time series forecasting: The forecast package for R. *Journal of Statistical Software*, 27(1), 1–22.

Hurkmans, R. T. W., Hurk, B. V. D., Schmeits, M. S., Wetterhall, F. & Pechlivanidis, I. G. (2023). Seasonal Streamflow Forecasting for Fresh Water Reservoir Management in the Netherlands: An Assessment of Multiple Prediction Systems. *Journal of Hydrometeorology*, 24 (7): 1275–1290. doi: 10.1175/JHM-D-22-0107.1

Jose, D. M, Vincent, A. M. & Dwarakish, G. S. (2022). Improving Multiple Model Ensemble Predictions of Daily Precipitation and Temperature through Machine Learning Techniques. *Scientific Reports*, 12: 4678, 1-25. doi: 10.1038/s41598-022-08786-w.

Kartika, Q. A. Faqih, A., Santikayasa, I. P., & Setiyawan, A. M. (2021). Sea Surface Temperature Anomaly Characteristics Affecting Rainfall in Western Java, Indonesia. *Agromet*, 37 (10), 54-65, doi: 10.29244/j.agromet.37.1.54-65.

Kullahci, A. G. K. & Altunkaynak, A. (2023). Integrating Wavelet Decomposition and Stacking Ensemble learning for Accurate Daily Rainfall Forecasting.

Kundu, S., Biswass, S. K., Tripathi, D., Karmakar, R., Majumdar, S., & Mandal S. (2023). A review on rainfall forecasting using ensemble learning techniques. *Advances in Electrical Engineering, Electronics and Energy*, 6. doi: 10.1016/j.prime.2023.100296.

Kurniawan, A., Muzayanah, L. F., & Sudirman. (2022). Verifikasi Luaran HyBMG saat Kejadian El Niño & La Niña di Jawa Timur. *Buletin GAW Bariri (BGB)*, 3(2), 8-15. doiI: 10.31172/bgb.v3i2.69. P-ISSN: 2721-7752 | E-ISSN: 2721-9704.

Mahera, U., Ilhamsyah, Y., & Octavina, C. Pengaruh Suhu Muka Laut terhadap Awal Masuk Musim hujan di Wilayah Utara-Timur Aceh. *Jurnal*

Lingkungan Almuslim, 2 (2), 035-049. doi: 10.51179/jla.v2i2.2110

Mahmudiah, R. R., Kurniawan, A. & Makmus, E. E. S. (2019). Determining best East Java monthly rainfall projection using spatial-based validation. *IOP Conf. Ser.: Earth Environ. Sci.,* 303 012023.

Maulita, M., & Nurdin. (2023). Pendekatan Data Mining untuk Analisis Curah Hujan Menggunakan Metode Regresi Linear Berganda (Studi Kasus: Kabupaten Aceh Utara). *Jurnal Online Manajemen Informatika*, 6 (2), 99-106. ISSN 2684-7280.

Noviasari, T., Nuzula, N. I., Efendy, M., Febrianto, A. A., & Darmadi, A. (2023). Peramalan Curah Hujan terhadap Produktivitas Garam di Gresik Putih Sumenep. *Jurnal Kelautan Tropis*, 26 (1): 9-18. doi: 10.14710/jkt.v26i1.16139.

Pakdaman, M., Babaeian, I., & Bouwer, L. M. (2022). Improved Monthly and Seasonal Multi-Model Ensemble Precipitation Forecasts in Southwest Asia Using Machine Learning Algorithms. *Water*, 14. doi: /10.3390/w14172632.

Ruslana, Z. N., Tresnawati, R., Rosyidah, , Harmoko, I. W., & Siswanto. (2021). Reliabilitas Prediksi Curah Hujan Dasarian Pada Kejadian Curah Hujan Ekstrim Pemicu Banjir 26 Oktober 2020 di Kebumen: Model Statistik (HyBMG) versus Model Dinamik (ECMWF). *Jurnal Geosains dan Teknologi*, 4(2).

Raju, K. S. & Kumar, D. N. (2020). Review of Approaches for Selection and Ensembling of GCMs. *Journal of Water and Climate Change*, 11, 577-599. doi: 10.2166/wcc.2020.128.

Specq, D., Batte, L., Deque, M., & Ardilouze, C. (2020). Multimodel Forecasting of Precipitation at Subseasonal Timescales Over the Southwest Tropical Pacific. *Earth and Space Science*, 7(9). doi: 10.1029/2019EA001003.

Tuysuzoglu, G., Birant, U. K., & Birant, D. (2023). Rainfall Prediction Using an Ensemble Machine Learning Model Based on K-Stars. *Sustainability*, 15, doi: 10.3390/su15075889.

Wei, X., Sun, X., Sun, J., Yin, J., Sun J. & Liu, C. (2022). A Comparative Study of Multi-Model Ensemble Forecasting Accuracy between Equal and Variant Weight Techniques. *Atmosphere*, 13(4). doi: 1-.3390/atmos13040526.

Weisheimer, A. & Palmer, T. (2014). On the Reliability of Seasonal Climate Forecast. *Journal of the Royal Society Interface*, 11 (96). doi: 10.1098/rsif.2013.162.

Weyn, J. A., Durran, D. R., Caruanam Rm & Cresswell-Clay, N. (2021). Sub-Seasonal Forecasting with a Large Ensemble of Deep-Learning Weather Prediction Models. *Journal of Advances in Modeling Earth Systems*, 13. doi: 10.1029/2021MS002502

Willmot, C. & Matsuura. (2005). Advantages of the Mean Absolute Error (MAE) over the Root Mean Square Error (RMSE) in assessing average model performance. *Clim. Res*., 30: 79-82.

Zhai, A., Fan, G., Ding, X., & Huang G. (2022). Regression Tree Ensemble Rainfall–Runoff Forecasting Model and Its Application to Xiangxi River, China. *Water*, 14. doi: 10.3390/w14030463.