International Journal on Advanced Science Engineering Information Technology

Vol.11 (2021) No. 5
ISSN: 2088-5334

The Assessment of Rainfall Prediction Using Climate Models Results and Projections under Future Scenarios: the Machángara Tropical Andean Basin Case

Angel Vázquez-Patiño [a,b,*], Mario Peña [c], Alex Avilés [d]

[a] *Facultad de Ingeniería, Universidad de Cuenca, Cuenca, 010203, Azuay, Ecuador*
[b] *Departamento de Ingeniería Civil, Universidad de Cuenca, Cuenca, 010203, Azuay, Ecuador*
[c] *Departamento de Química Aplicada y Sistemas de Producción, Universidad de Cuenca, Cuenca, 010203, Azuay, Ecuador*
[d] *Carrera de Ingeniería Ambiental, Facultad de Ciencias Químicas, Universidad de Cuenca, Cuenca, 010203, Azuay, Ecuador*
Corresponding author: [*]angel.vazquezp@ucuenca.edu.ec

***Abstract*—Rainfall is vital in the biosphere and predicting it is essential under the possible adverse effects of climate change. Rainfall behavior is linked to the availability of fresh water and the development of almost all the activities necessary for human subsistence. Therefore, knowing their patterns under future scenarios could help decision-makers to plan water use policies. This study used the random forest algorithm to predict rainfall in Chanlud and El Labrado stations, located in the tropical Machángara high mountain basin in Ecuador. Data from the Ecuador project's third national communication (TNC) were used to train the prediction models. First, those models' performance was analyzed to know which climate model results of the TNC provide more information to learn observed rainfall patterns. Then, the rainfall signal was projected under the RCP 4.5 and 8.5 scenarios. Among the most important results obtained, it stands out that the assembly results of the TNC provided the best information to learn rainfall patterns in the present. The performance is the best from January to July, but from August to December it is lower. Rainfall projections under RCP 8.5 are, in general, lower than under RCP 4.5. No significant trends were found in the future. However, a very slight increase (decrease) of rainfall was observed for the driest (wettest) months in both stations, although slightly more accentuated in El Labrado.**

***Keywords*—Machángara basin; rainfall prediction; random forest; climate models; projection; future scenarios; RCP.**

Manuscript received 14 Mar. 2021; revised 30 May. 2021; accepted 2 Sep. 2021. Date of publication 31 Oct. 2021.

I. INTRODUCTION

Rainfall prediction is essential to decide mitigation plans in a possible scenario of climate change. Since rainfall is one of the most significant freshwater sources, knowing its future availability is crucial to developing water use policies. Moreover, possible future changes in rainfall patterns could impact agriculture and the production of hydropower plants with negative economic results. These impacts could be a more tangible reality in countries with fragile irrigation systems whose energy production is based mainly on hydroelectric generation. If it is possible to anticipate such changes, decision-makers could develop plans that counteract adverse effects or propose policies of water use based on scientific evidence.

Climate models are the most used tools to simulate future climate. They are mathematical models representing the general circulation of the planet and assemble different climate system components (e.g., atmosphere, cryosphere) to predict its behavior under different scenarios of greenhouse gases, e.g., RCP 4.5 and 8.5. Their spatial resolution varies from 1° to 3° (approx. 111 to 333 km), which has been sufficient to propose hypotheses about the change in climate dynamics on a global scale.

Climate model results are presented at different time scales, e.g., daily or monthly. Different time scales may help to understand different processes that act under different phenomena. For example, weather systems act on a daily scale or atmospheric blocking weekly [1]. Rainfall on a high temporal resolution could help predict extremes (e.g., floods), which often result in human deaths. On the other hand, the prediction of rainfall in a lower temporal resolution, for example, monthly, would be helpful to carry out plans to control the catchment of water in dams. However, climate model results do not represent rainfall variability, especially the high spatiotemporal variability of high mountain zones.

Prediction of rainfall is a complex task in high mountain zones, like the tropical Andean ones. In these regions, the high spatiotemporal variability is poorly represented by climate models, especially on a daily scale. Different efforts have been made for downscaling rainfall data and use these to perform predictions. Two main approaches have been used in downscaling, namely, dynamical and statistical downscaling [2]. These two approaches were used in the Third National Communications of Ecuador on Climate Change (TNC) [3] to produce a high spatial resolution of climate variables in the present, including rainfall. Additionally, the TNC generated high spatial-resolution climatic projections for 2011-2040, 2041-2070, and 2071-2100 under two scenarios of greenhouse gases, the RCP 4.5 and 8.5. These data can be leveraged to predict rainfall based on machine-learning models that learn patterns between climatic variables in the present and simulate the future using the climatic projections of the TNC.

This study predicts rainfall of a high mountain tropical Andean basin located in Ecuador on a sub-monthly scale. Specifically, the objectives are as follows:

- to investigate which results from the climate models of the TNC content more information to learn rainfall patterns to boost the prediction of the prediction models,
- to find out the performance of the prediction models in each month, and
- to use the best performing models to predict rainfall under the RCP 4.5 and 8.5 scenarios, investigating trends and possible changes in seasonality.

II. Materials and Methods

A. Study area

The study area is the Machángara basin in southern Ecuador (Fig. 1), close to Cuenca, considered the third most important city in the country. The basin is essential for freshwater supply and industrial activities. The altitude of the basin varies from 2440 to 4420 m a.s.l. [4], [5]. This accentuated gradient is a characteristic of the Ecuadorian mountain range. Rainfall observations are available in two stations in the Machángara Alto and Chulco river sub-basins [6], namely, Chanlud and El Labrado stations. The seasonality of rainfall in the two stations is shown in Fig. 2a. April is the wettest month in the Chanlud and El Labrado stations, whereas August is the driest month.

B. Climate models

The TNC project selected climate models from the Coupled Model Intercomparison Project, phase 5 (CMIP5) [7], to perform downscaling. In this study, the results of the dynamical downscaling from the TNC project are used.

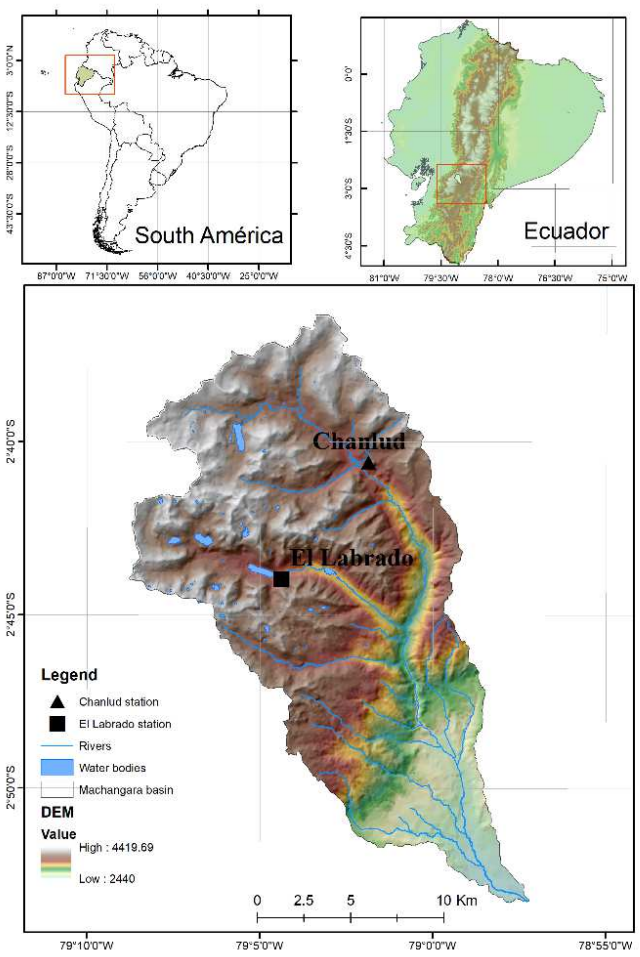

Fig. 1 Location of the study area and the two stations where rainfall data is available, Chanlud and El Labrado.

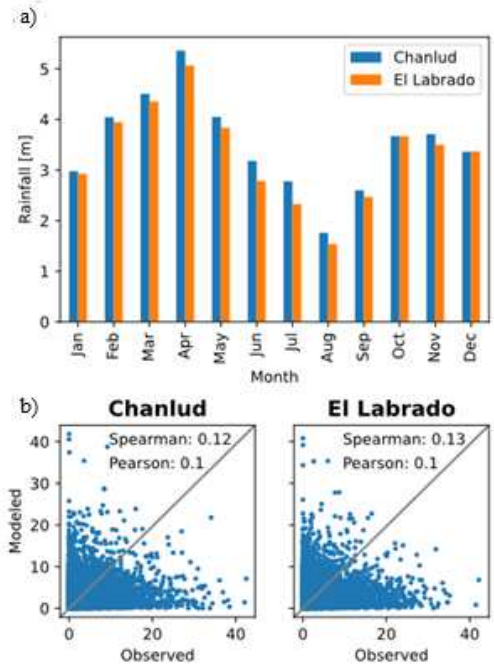

Fig. 2 a) Seasonality of rainfall observations in Chanlud and El Labrado stations and b) the relation between daily rainfall observations and simulations of the assembly model.

The dynamical downscaling was performed using the Weather Research and Forecasting model (WRF) version 3.6.1 [8]. Table I shows the models on which the data used in the study are based; additionally, an ensemble of the models used in the TNC project and explained in Porras *et al.* [9].

TABLE I
CLIMATE MODELS ON WHICH THE DATA USED IN THE STUDY ARE BASED.

Model	Center	Country
Assembly	TNC	Ecuador
IPSL-CM5A-MR	Institut Pierre-Simon Laplace	France
CSIRO-Mk3-6-0	CSIRO Atmospheric Research	Australia
GISS-E2-R	NASA/Goddard Institute for Space Studies	USA
MIROC-ESM	Japan Agency for Marine-Earth Science and Technology, Atmosphere and Ocean Research Institute (The University of Tokyo), and National Institute for Environmental Studies	Japan

C. Data

Available daily rainfall observations from the Chanlud and El Labrado stations span from 1981 to 2018 (38y). However, the historical results from the models of the TNC are available from 1981 to 2005 (25y). Therefore, the latter period was used to train and test the machine learning models to learn present patterns. For this, daily data of eight variables shown in Table II were used to predict rainfall observations. The data of variables shown in Table II has a resolution of 10 km. The same variables are used in the projections of rainfall.

TABLE II
CLIMATIC VARIABLES OF THE TNC MODELS USED AS PREDICTORS OF RAINFALL OBSERVATIONS IN THE PRESENT AND FOR THE PROJECTIONS.

Variable	Description	Units
hr	Relative humidity at the surface	%
precip	Total precipitation	mm/day
rad	Surface downwelling shortwave radiation	W/m^2
tmax	Maximum temperature	Degree Celsius (°C)
temp	Average temperature	°C
tmin	Minimum temperature	°C
u10m	Eastward near-surface wind	m/s
v10m	Northward near-surface wind	m/s

For rainfall prediction in Chanlud and El Labrado, the data of the TNC (models in Table I) for RCP 4.5 and 8.5 scenarios were used. The prediction is performed for the period 2021-2070 (50y). Due to the complex dynamic processes controlling rainfall in the study area, climate models cannot satisfactorily reproduce daily rainfall [10], [11]. To evidence this fact in the TNC data, Fig. 2b shows the relation between observed rainfall and the assembly model's total precipitation (precip in Table II). Neither in Chanlud nor El Labrado, the results of the assembly model can assimilate rainfall on a high temporal resolution.

In early attempts to model rainfall on a daily scale, the results were quite poor because of the low ability of climate models to simulate observations. Even the assembly model's precipitation variable (which is expected to be similar to the observations) presents a correlation lower than 0.13 in both stations (Fig. 2b). Therefore, an aggregation of the data was carried out. In this way, a balance between the performance of the prediction models and a low temporal resolution was satisfied.

Rainfall observations and data of variables in Table II. were aggregated, averaging values of 15 days to represent the intermediate day. Specifically, for a daily time series $X = \{X_1, X_2, \ldots, X_n\}$ of n data, the aggregation results in another daily time series $Y = \{Y_1, Y_2, \ldots, Y_n\}$ where

$$Y_t = \frac{1}{t+7}\sum_{1}^{t+7} X_i \quad \text{if} \quad 1 \leqslant t < 8$$
$$Y_t = \frac{1}{15}\sum_{t-7}^{t+7} X_i \quad \text{if} \quad 8 \leqslant t \leqslant n-7 \qquad (1)$$
$$Y_t = \frac{1}{n-t+8}\sum_{t-7}^{n} X_i \quad \text{if} \quad n-7 < t \leqslant n$$

D. Random Forest

The random forest (RF) algorithm [12] is a perturb-and-combine technique based on an assembly of decision trees [13] to boost a prediction. Decision trees commonly have high variance and tend to present overfitting. In the RF algorithm, each decision tree is formed with a subset of input variables at each node to decrease the variance. Moreover, each tree is built using a randomly selected sample of the training subset.

RF has become popular in climate and meteorology due to its efficiency when tackling big datasets and of high dimensionality [14]–[16]. Moreover, the algorithm allows for estimating the importance of the different input variables [17]. Different hyperparameters can be optimized when training a model based on RF [18], [19]. Three of the most important ones are the number of trees tested in the forest, the maximum depth of the trees, and the number of features to consider when splitting the data.

E. Settings and Workflow

In order to generate the instances for training the prediction models and posterior prediction, the time series of the eight variables in Table II are extracted from the nearest pixel to the location of the stations. Moreover, four different time series are extracted to provide information on nearby locations (for each variable). These locations correspond to 11 pixels north, 11 pixels south, 11 pixels east, and 11 pixels west, representing approximately a radius of 100 km or one degree around the Chanlud and El Labrado stations. In this way, five-time series are extracted from the results of climate models for each station.

In order to leverage past and posterior information in the prediction of rainfall, lagged information of the predictor variables (the five-time series) was also added to form the

instances used in RF. Lags of 15 and -15 days were used, resulting in instances containing (aggregated) information from 22 days before to 22 days after the day of the rainfall value due to the data aggregation explained earlier. Finally, each instance is formed of 120 predictor values (five-time series for each of the eight variables of Table II × the lagged values) and the rainfall value as the target value.

An RF-based model was trained for each month. To train a model for a specific month, the information for that month was used together with that of the previous and subsequent month. This decision was made in order to help the algorithm learn the season-specific relationships between the predictors and rainfall. This prevents the information from another station from becoming noise since this information could correspond to relationships that are tied to seasonality. Since five climate models (Table I) were used, the total number of prediction models was 60.

When training the RF-based prediction models for each month, the information taken into account is divided to use 80% of the data for training the models and 20% for testing the performance of the trained models on fresh data. The following values were used in a random grid search 5-folds cross-validation fashion [20] to optimize the hyperparameters of the models:

- Number of trees in the forest: 90, 128, 181, 256, 362, 512, 724, 1024, and 1448.
- Maximum depth of the tree: 20, 35, 50, 65, 80, 95, 110, and until all leaves are pure or until all leaves contain <2 samples.
- The number of features to consider when looking for the best split: the square root of the number of input variables, the $\log_2$ of the number of input variables, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80, 90%, and 100% of the number of input variables.

After testing the performance of the 60 trained models, the climate model whose data shows the best RF-based prediction is used to predict the rainfall in Chanlud and El Labrado under the two RCP scenarios.

F. Performance metrics

Given the daily rainfall time series $Y = \{Y_1, Y_2, \ldots, Y_n\}$ and the predicted time series $\hat{Y} = \{\hat{Y}_1, \hat{Y}_2, \ldots, \hat{Y}_n\}$, four performance metrics, independent of the data scale, were calculated (first four one listed below). Moreover, two of the most commonly used metrics (dependent on data scale) were also calculated to compare with other future studies.

1) Nash-Sutcliffe Efficiency: The Nash-Sutcliffe Efficiency coefficient (NSE) [21] is commonly used to assess the performance of hydrological models but also in assessing predictive rainfall models [22]. It is defined as follows:

$$NSE = 1 - \frac{\sum_1^n (Y_i - \hat{Y}_i)^2}{\sum_1^n Y_i - mean(Y)} \quad (2)$$

where mean (Y) is the mean of the time series Y.

The value range of NSE goes from $-\infty$ to 1. NSE = 1 means a perfect match, NSE = 0 indicates that the mean of the observations is as good as the prediction, and NSE < 0 indicates a model performing worse than using the mean of the observations.

2) Kling-Gupta Efficiency: Kling-Gupta Efficiency (KGE) [21] is commonly used in the context of hydrological modeling. It is a decomposition of NSE and is based on analyzing the relative importance of variability, linear correlation, and the bias ratio between Y and $\hat{Y}$. It is defined as

$$KGE = 1 - \sqrt{(r-1)^2 + (\alpha - 1)^2 + (\beta - 1)^2} \quad (3)$$

where r is the Pearson's correlation coefficient

$$\alpha = \frac{std(\hat{Y})}{std(Y)} \quad (4)$$

and

$$\beta = \frac{\sum_1^n \hat{Y}_i}{\sum_1^n Y_i} \quad (5)$$

The KGE ranges from $-\infty$ to 1. The closer to 1, the more accurate the model is.

3) Percent Bias: The percent bias (PBIAS) measures how much the mean trend of the simulated values is higher or lower than the observed ones. It is defined as

$$PBIAS = 100 \times \frac{\sum_1^n (Y_i - \hat{Y}_i)}{\sum_1^n Y_i} \quad (6)$$

The range of |PBIAS| goes from 0 to ∞. The optimal value is zero, while positive (negative) values indicate a tendency to overestimate (underestimate) the simulated values.

4) Mean Absolute Relative Error: The Mean Absolute Relative Error (MARE) measures how much error exists relative to the observed data. The error is computed in absolute terms as defined next

$$MARE = \frac{\sum_1^n (Y_i - \hat{Y}_i)}{\sum_1^n Y_i} \quad (7)$$

The values of MARE ranges from 0 to ∞ being 0, the optimum measure.

5) Root Mean Square Error: The Root Mean Square Error (RMSE) is the square root of the quadratic mean of the differences between simulated and observed values, defined as

$$RMSE = \sqrt{\frac{1}{n}\sum_1^n (Y_i - \hat{Y}_i)^2} \quad (8)$$

The RMSE range goes from 0 to ∞ being 0 the optimum value.

6) Mean Absolute Error: The Mean Absolute Error (MAE) measures the absolute errors between the simulated and observed values. It is defined as

$$MAE = \frac{1}{n}\sum_1^n (Y_i - \hat{Y}_i) \quad (9)$$

The MAE range goes from 0 to ∞ being 0 the optimum value.

7) Model Comparison Metric: For values NSE > 0, KGE >

0, |PBIAS| < 100, and MARE < 1, the model comparison metric (MCM) was defined combining NSE, KGE, PBIAS, and MARE as follows:

$$MCM = NSE \times KGE \times \left(1 - \frac{PBIAS}{100}\right)(1 - MARE) \qquad (10)$$

This metric helps in the intercomparison of models in general terms. The optimum value of MCM is 1, and the worse is 0.

III. Results and Discussion

The complete list of results is shown in the appendix. In this section, only the MCM is used to compare the performance of the RF-based prediction models. Once the climate model whose results give the best predictive information is chosen, those results are used to present rainfall predictions in the future.

Fig. 3 shows the performance (MCM metric) of the prediction models throughout the 12 months for Chanlud (Fig. 3a) and El Labrado (Fig. 3b). Fig. 3 also allows comparing the performance of the prediction models using the results of the five climate models of the TNC (Table I). In general terms, the performance reached for Chanlud surpasses that of El Labrado except for the RF-based models that use the MIROC-ESM results in August (Fig. 3b).

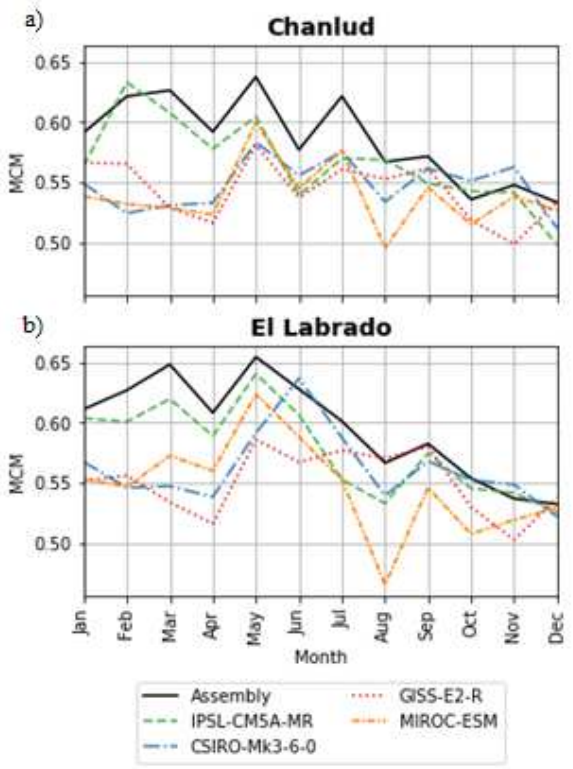

Fig. 3 Performance of the RF-based models to predict rainfall each month in a) Chanlud and b) El Labrado. The MCM metric values are shown with different lines for the different model results used.

In both stations, the RF-based prediction models perform better using the results of the assembly model from the TNC. Using those results, the performance of the prediction models is notably better than the model IPSL-CM5A-MR, whose results present the next best performance when using RF. These results show that the assembly of results must be the first option when choosing a single climate model for learning patterns between the predictor variables (Table II) and the rainfall observations in the Machángara basin.

A remarkable feature that can be noticed in both Chanlud and El Labrado is the low performance of the RF-based models in April, the wettest month of the year (see Fig. 2b). This evidences that information of the predictors is not enough to learn rainfall patterns when rainfall values are the highest.

On the other hand, the best performance in the prediction is reached in May in both Chanlud and El Labrado. In general, the prediction models based on all the climate models present the same behavior. May is the month when the most pronounced wettest season begins to decline.

When analyzing the results corresponding to all the climate models except the assembly, the decrease of performance in August is notorious. This behavior is more pronounced for the El Labrado than Chanlud; for example, it is evident in the results based on MIROC-ESM data in Fig. 3b. August is the driest month (see Fig. 2b). However, it is not possible to present a general conclusion about the relationship between performance and the rainfall amount. In some cases, worse results are shown in October when the second less pronounced wet season appears.

Fig. 4 shows the prediction of rainfall under the RCP 4.5 and 8.5 scenarios for Chanlud (Fig. 4a) and El Labrado (Fig. 4b). Moreover, the figure shows the linear trends for the whole time series and the wettest and driest months. For both Chanlud and El Labrado, rainfall projections in the wettest season under the RCP 8.5 scenario are under the values projected under the RCP 4.5 scenario. This is verified from the trend lines of the wettest month. Likewise, the annual trend lines show that, in general, under RCP 8.5, a decrease in the amount of rainfall water is projected compared to the projection under the RCP 4.5 scenario. Regarding the months with the lowest rainfall, approximately the same rainfall is projected under both RCP scenarios.

The annual trends in Fig. 4 do not show a wide variation in the rainfall amount as the century progresses. However, the wettest months (RCP 4.5 and 8.5) trend shows a slight decrease in rainfall towards 2070, although it cannot be considered significant. This decrease is only slightly more pronounced for El Labrado. In the same way, there exists a very slight progressive increase (not significant) in the rainfall amounts of the driest months. The trend on the driest months has a slightly more pronounced slope in the projections under the RCP 8.5 scenario.

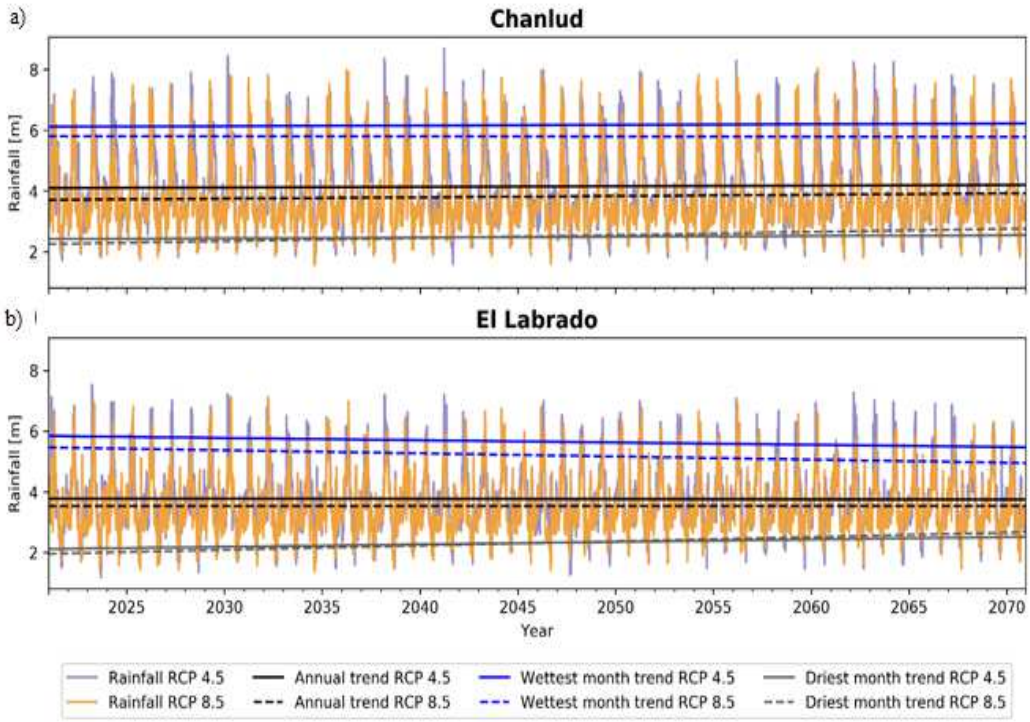

Fig. 4 Prediction of rainfall based on the assembly model data from 2021 to 2070 under the RCP 4.5 and 8.5 scenarios for a) Chanlud and b) El Labrado. The straight lines correspond to linear trends of the annual prediction, the prediction of the wettest months (months of April), and the prediction of the driest months (months of August).

To analyze possible changes in the seasonality of rainfall under the two future scenarios, Fig. 5 shows the differences between the seasonality of the observations (present) and RCP 4.5 and 8.5. In general, rainfall increases during most of the year in both Chanlud (Fig. 5a) and El Labrado (Fig. 5b) stations. The exceptions are February, October, and November, when a decrease is projected in both stations. February is the month when the wettest season starts, and October and November are the months when the second peak of rainfall is present.

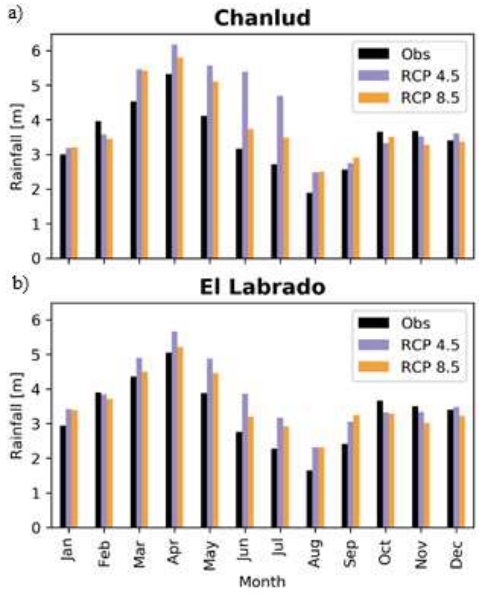

Fig. 5 Seasonality of rainfall observations and predictions under the RCP 4.5 and 8.5 scenarios for a) Chanlud and b) El Labrado. The seasonality of observations is calculated after the aggregation process, explained in section 2.3.

The most accentuated change of the rainfall pattern is shown for Chanlud (Fig. 5a). After the wettest season (Mar-May) peak, a very high increase in rainfall is projected during June and July, especially under the RCP 4.5 scenario. This is particularly interesting because July is one of the two months with the highest performance (Fig. 3a). This fact implies a certain level of confidence in the projection.

In El Labrado, an accentuated increase in June and July as in Chalud is not shown (Fig. 5b), but the same pattern of increase from March to September and decrease from October to December is present. However, it is necessary to consider that the RF-based models had worse performance during October-December (Fig. 3).

IV. Conclusion

Rainfall is essential in many anthropogenic activities and is vital in the water cycle. Projecting rainfall is important for anticipating mitigation measures under possible negative scenarios and leveraging water based on scientific evidence. In this study, the random forest algorithm was used to predict rainfall in two stations located in the Machángara basin. Complex meteorological processes condition rainfall in this tropical Andean basin.

The performance of RF-based models determines that the assembly model of the TCN contains more information to learn patterns of rainfall behavior than the four other models analyzed. The assembly model shows the best performance from January to July, then the performance declines. In Chanlud and El Labrado, May is the month with the best prediction, just after the wettest month when the performance decreases.

The prediction of rainfall under RCP 4.5 and 8.5 scenarios, using the data of the assembly model, is presented. A slightly

progressive decrease of rainfall in the wettest months is projected, although not significant, more accentuated for the El Labrado station. Likewise, a slight increase of rainfall is projected for the driest months, a little more accentuated (and not significant) under the RCP 8.5 scenario. When comparing the seasonality of predicted rainfall and observations, there is a systematic increase of rainfall during January and March-September in Chanlud and El Labrado. However, in Chanlud there is a considerable increase in June and July under the RCP 4.5 scenario, contrary to the El Labrado. On the other hand, February, October, and November show a rainfall decrease. The results presented here could be leveraged to choose a climatic model from the TCN to study climate in Ecuador. The results show that the methodology used in the TCN to deploy the assembly model boosts the information of individual models to learn rainfall patterns.

ACKNOWLEDGMENT

This work was carried out in the context of the Project "Andes Adaptation to the Impact of Climate Change on Water Resources" (AICCA). The authors are grateful for the support of the Global Environment Facility, Banco de Desarrollo de América Latina, CONDESAN, Ministerio del Ambiente del Ecuador, ELECAUSTRO, and Universidad de Cuenca. This manuscript is an outcome of the Doctoral Program in Water Resources, offered by Universidad de Cuenca, Escuela Politécnica Nacional, and Universidad Técnica Particular de Loja.

REFERENCES

[1] P. D. Williams *et al.*, "A Census of Atmospheric Variability From Seconds to Decades," *Geophys. Res. Lett.*, vol. 44, no. 21, Nov. 2017, doi: 10.1002/2017GL075483.

[2] S. Trzaska and E. Schnarr, "A Review of Downscaling Methods for Climate Change Projections," Center for International Earth Science Information Network (CIESIN), Burlington, Vermont, USA, 2014. [Online]. Available: http://www.ciesin.org/documents/Downscaling_CLEARED_000.pdf.

[3] Ministerio del Ambiente del Ecuador, "Third National Communication of Ecuador on Climate Change (in Spanish)," MAE, Quito, Ecuador, 2017.

[4] J. F. Farfán, K. Palacios, J. Ulloa, and A. Avilés, "A hybrid neural network-based technique to improve the flow forecasting of physical and data-driven models: Methodology and case studies in Andean watersheds," *J. Hydrol. Reg. Stud.*, vol. 27, p. 100652, Feb. 2020, doi: 10.1016/j.ejrh.2019.100652.

[5] A. Avilés, A. Solera Solera, J. Paredes-Arquiola, and M. Pedro-Monzonís, "Integrated Methodological Framework for Assessing the Risk of Failure in Water Supply Incorporating Drought Forecasts. Case Study: Andean Regulated River Basin," *Water Resour. Manag.*, vol. 32, no. 4, pp. 1209–1223, Mar. 2018, doi: 10.1007/s11269-017-1863-7.

[6] A. Avilés, K. Palacios, J. Pacheco, S. Jiménez, D. Zhiña, and O. Delgado, "Sensitivity exploration of water balance in scenarios of future changes: a case study in an Andean regulated river basin," *Theor. Appl. Climatol.*, vol. 141, no. 3–4, pp. 921–934, Aug. 2020, doi: 10.1007/s00704-020-03219-y.

[7] K. E. Taylor, R. J. Stouffer, and G. A. Meehl, "An Overview of CMIP5 and the Experiment Design," *Bull. Am. Meteorol. Soc.*, vol. 93, no. 4, pp. 485–498, Apr. 2012, doi: 10.1175/BAMS-D-11-00094.1.

[8] J. G. Powers *et al.*, "The Weather Research and Forecasting Model: Overview, System Efforts, and Future Directions," *Bull. Am. Meteorol. Soc.*, vol. 98, no. 8, pp. 1717–1737, Aug. 2017, doi: 10.1175/BAMS-D-15-00308.1.

[9] G. E. Armenta Porras, J. L. Villa Cedeño, and P. Jácome, "Proyecciones climáticas de precipitación y temperatura para Ecuador, bajo distintos escenarios de cambio climático," Ministerio del Ambiente, Quito, Ecuador, 2016.

[10] G. Esquivel-Hernández *et al.*, "Moisture transport and seasonal variations in the stable isotopic composition of rainfall in Central American and Andean Páramo during El Niño conditions (2015–2016)," *Hydrol. Process.*, p. hyp.13438, Apr. 2019, doi: 10.1002/hyp.13438.

[11] A. Vázquez-Patiño, L. Campozano, D. Ballari, M. Córdova, and E. Samaniego, "Virtual Control Volume Approach to the Study of Climate Causal Flows: Identification of Humidity and Wind Pathways of Influence on Rainfall in Ecuador," *Atmosphere (Basel).*, vol. 11, no. 8, p. 848, Aug. 2020, doi: 10.3390/atmos11080848.

[12] L. Breiman, "Random Forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Jul. 2001, doi: 10.1023/A:1010933404324.

[13] G. Karalis, "Decision Trees and Applications," 2020, pp. 239–242.

[14] P. Contreras, J. Orellana-Alvear, P. Muñoz, J. Bendix, and R. Célleri, "Influence of Random Forest Hyperparameterization on Short-Term Runoff Forecasting in an Andean Mountain Catchment," *Atmosphere (Basel).*, vol. 12, no. 2, p. 238, Feb. 2021, doi: 10.3390/atmos12020238.

[15] P. Muñoz, J. Orellana-Alvear, and R. Célleri, "Application of a Machine Learning Technique for Developing Short-Term Flood and Drought Forecasting Models in Tropical Mountainous Catchments," 2021, pp. 11–35.

[16] J. Orellana-Alvear, R. Célleri, R. Rollenbeck, and J. Bendix, "Optimization of X-Band Radar Rainfall Retrieval in the Southern Andes of Ecuador Using a Random Forest Model," *Remote Sens.*, vol. 11, no. 14, p. 1632, Jul. 2019, doi: 10.3390/rs11141632.

[17] J. L. Speiser, M. E. Miller, J. Tooze, and E. Ip, "A comparison of random forest variable selection methods for classification prediction modeling," *Expert Syst. Appl.*, vol. 134, pp. 93–101, Nov. 2019, doi: 10.1016/j.eswa.2019.05.028.

[18] P. Probst, M. N. Wright, and A. Boulesteix, "Hyperparameters and tuning strategies for random forest," *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, vol. 9, no. 3, May 2019, doi: 10.1002/widm.1301.

[19] P. Probst, A.-L. Boulesteix, and B. Bischl, "Tunability: importance of hyperparameters of machine learning algorithms," *J. Mach. Learn. Res.*, vol. 20, no. 1, pp. 1934–1965, 2019.

[20] J. Bergstra and Y. Bengio, "Random Search for Hyper-Parameter Optimization," *J. Mach. Learn. Res.*, vol. 13, no. 10, pp. 281–305, 2012, [Online]. Available: http://jmlr.org/papers/v13/bergstra12a.html.

[21] W. J. M. Knoben, J. E. Freer, and R. A. Woods, "Technical note: Inherent benchmark or not? Comparing Nash–Sutcliffe and Kling–Gupta efficiency scores," *Hydrol. Earth Syst. Sci.*, vol. 23, no. 10, pp. 4323–4331, Oct. 2019, doi: 10.5194/hess-23-4323-2019.

[22] M. Peña, A. Vázquez-Patiño, D. Zhiña, M. Montenegro, and A. Avilés, "Improved Rainfall Prediction through Nonlinear Autoregressive Network with Exogenous Variables: A Case Study in Andes High Mountain Region," *Adv. Meteorol.*, vol. 2020, pp. 1–17, Sep. 2020, doi: 10.1155/2020/1828319.

APPENDIX

TABLE III

PERFORMANCE OF THE RF-BASED MODELS TO PREDICT RAINFALL EACH MONTH IN CHANLUD BASED ON THE CLIMATE MODELS DATA OF THE TNC. THE METRIC VALUES IN BOLD CORRESPOND TO THE MODEL WHOSE DATA ALLOWED RF TO OBTAIN THE HIGHEST PERFORMANCE EACH MONTH (ONE VALUE IN BOLD FOR EACH MONTH). THE SHADED VALUES CORRESPOND TO THE MONTH IN WHICH RF OBTAINED THE BEST PERFORMANCE BASED ON THE RESULTS OF EACH CLIMATE MODEL (ONE SHADED VALUE FOR EACH CLIMATE MODEL).

		Jan	**Feb**	**Mar**	**Apr**	**May**	**Jun**	**Jul**	**Aug**	**Sep**	**Oct**	**Nov**	**Dec**
NSE	CSIRO-Mk3-6-0	0.88	0.87	0.86	0.86	0.87	0.86	0.88	0.87	0.87	0.88	0.88	0.86
	Assembly	0.90	0.91	0.90	0.89	0.89	0.87	0.89	0.89	0.88	0.86	0.89	0.87
	GISS-E2-R	0.89	0.88	0.86	0.86	0.87	0.86	0.88	0.88	0.87	0.88	0.86	0.87
	IPSL-CM5A-MR	0.88	0.91	0.88	0.88	0.89	0.86	0.88	0.88	0.87	0.88	0.88	0.86
	MIROC-ESM	0.88	0.87	0.85	0.85	0.88	0.87	0.87	0.84	0.87	0.86	0.87	0.86
KGE	CSIRO-Mk3-6-0	0.75	0.72	0.73	0.72	0.78	0.76	0.77	0.73	0.77	0.75	0.76	0.73
	Assembly	0.77	0.79	0.80	0.77	0.82	0.77	0.81	0.75	0.78	0.77	0.74	0.74
	GISS-E2-R	0.77	0.75	0.72	0.71	0.77	0.73	0.76	0.75	0.77	0.72	0.71	0.74
	IPSL-CM5A-MR	0.76	0.79	0.79	0.76	0.79	0.74	0.77	0.78	0.75	0.75	0.74	0.71
	MIROC-ESM	0.73	0.73	0.73	0.72	0.79	0.73	0.79	0.72	0.75	0.73	0.75	0.74
PBIAS	CSIRO-Mk3-6-0	-0.88	0.81	0.42	1.44	0.87	1.73	0.61	1.76	1.00	-0.88	0.04	-0.76
	Assembly	-0.44	0.70	0.37	1.67	-0.01	0.85	0.02	0.99	0.70	-2.91	0.58	-0.30
	GISS-E2-R	-1.67	0.61	-0.48	1.96	0.08	1.00	0.98	1.82	0.33	-0.94	0.27	-0.67
	IPSL-CM5A-MR	-1.33	0.42	0.45	1.05	-0.22	0.65	0.20	1.68	0.17	-1.84	-0.20	-0.38
	MIROC-ESM	-0.93	1.11	-0.07	1.53	-0.40	0.74	-0.58	2.02	0.68	-0.19	0.86	-0.31
MARE	CSIRO-Mk3-6-0	0.16	0.16	0.14	0.13	0.14	0.13	0.15	0.15	0.16	0.17	0.16	0.18
	Assembly	0.14	0.13	0.12	0.12	0.12	0.13	0.14	0.14	0.16	0.17	0.16	0.17
	GISS-E2-R	0.16	0.14	0.14	0.14	0.14	0.14	0.15	0.14	0.16	0.17	0.18	0.17
	IPSL-CM5A-MR	0.15	0.12	0.12	0.13	0.13	0.14	0.15	0.15	0.16	0.16	0.16	0.18
	MIROC-ESM	0.16	0.15	0.15	0.14	0.13	0.14	0.15	0.16	0.16	0.18	0.17	0.17
RMSE	CSIRO-Mk3-6-0	0.69	0.80	0.86	0.88	0.78	0.69	0.52	0.51	0.58	0.72	0.77	0.77
	Assembly	0.64	0.66	0.73	0.79	0.72	0.66	0.50	0.48	0.56	0.78	0.75	0.72
	GISS-E2-R	0.68	0.75	0.86	0.88	0.77	0.68	0.53	0.50	0.58	0.74	0.85	0.72
	IPSL-CM5A-MR	0.68	0.64	0.78	0.81	0.73	0.70	0.54	0.50	0.58	0.71	0.79	0.77
	MIROC-ESM	0.69	0.78	0.87	0.91	0.74	0.67	0.55	0.57	0.59	0.78	0.81	0.74
MAE	CSIRO-Mk3-6-0	0.54	0.61	0.65	0.65	0.56	0.45	0.38	0.37	0.43	0.54	0.59	0.58
	Assembly	0.49	0.50	0.55	0.57	0.51	0.44	0.36	0.35	0.43	0.56	0.58	0.54
	GISS-E2-R	0.53	0.55	0.65	0.67	0.56	0.47	0.39	0.35	0.43	0.56	0.64	0.54
	IPSL-CM5A-MR	0.51	0.48	0.56	0.61	0.54	0.48	0.40	0.37	0.44	0.52	0.59	0.57
	MIROC-ESM	0.54	0.60	0.67	0.67	0.56	0.46	0.40	0.40	0.44	0.59	0.61	0.55
MCM	CSIRO-Mk3-6-0	0.55	0.52	0.53	0.53	0.58	0.56	0.58	0.53	0.56	**0.55**	**0.56**	0.51
	Assembly	**0.59**	0.62	**0.63**	**0.59**	**0.64**	**0.58**	**0.62**	0.57	**0.57**	0.54	0.55	0.53
	GISS-E2-R	0.57	0.57	0.53	0.52	0.58	0.54	0.56	0.55	0.56	0.52	0.50	**0.53**
	IPSL-CM5A-MR	0.56	**0.63**	0.61	0.58	0.60	0.54	0.57	**0.57**	0.55	0.54	0.54	0.50
	MIROC-ESM	0.54	0.53	0.53	0.52	0.60	0.55	0.58	0.50	0.55	0.52	0.54	0.53

TABLE IV
THE SAME AS IN TABLE III BUT FOR EL LABRADO.

	El Labrado	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
NSE	CSIRO-Mk3-6-0	0.88	0.87	0.87	0.87	0.87	0.90	0.89	0.86	0.88	0.89	0.87	0.87
	Assembly	0.90	0.91	0.90	0.90	0.91	0.90	0.89	0.88	0.88	0.88	0.88	0.88
	GISS-E2-R	0.88	0.87	0.86	0.86	0.88	0.88	0.89	0.87	0.88	0.88	0.86	0.88
	IPSL-CM5A-MR	0.90	0.90	0.89	0.89	0.90	0.89	0.86	0.86	0.89	0.88	0.87	0.88
	MIROC-ESM	0.88	0.88	0.88	0.88	0.90	0.88	0.87	0.83	0.87	0.86	0.86	0.86
KGE	CSIRO-Mk3-6-0	0.76	0.74	0.73	0.73	0.79	0.81	0.78	0.75	0.78	0.75	0.75	0.74
	Assembly	0.79	0.79	0.82	0.77	0.82	0.80	0.79	0.76	0.79	0.77	0.73	0.73
	GISS-E2-R	0.75	0.75	0.73	0.71	0.77	0.76	0.77	0.78	0.79	0.73	0.71	0.74
	IPSL-CM5A-MR	0.78	0.77	0.79	0.76	0.82	0.79	0.76	0.75	0.78	0.75	0.75	0.74
	MIROC-ESM	0.75	0.74	0.76	0.74	0.81	0.77	0.76	0.70	0.76	0.72	0.73	0.75
PBIAS	CSIRO-Mk3-6-0	-0.48	0.09	-0.25	1.31	0.53	0.56	-0.13	1.64	0.97	-0.65	-0.11	-1.38
	Assembly	-0.62	0.14	-0.55	1.11	-0.10	0.38	-0.33	1.34	0.45	-2.10	-0.09	-1.34
	GISS-E2-R	-0.72	-0.17	-0.73	1.75	-0.03	0.49	0.43	1.44	0.23	-0.63	-0.18	-1.60
	IPSL-CM5A-MR	-0.97	-0.07	-0.98	0.65	-0.42	0.16	0.16	1.50	0.02	-1.36	-1.00	-1.87
	MIROC-ESM	-0.53	0.40	-0.77	0.87	-0.89	0.17	-0.68	2.01	0.31	-0.40	0.22	-1.25
MARE	CSIRO-Mk3-6-0	0.15	0.15	0.14	0.14	0.14	0.13	0.15	0.15	0.16	0.16	0.16	0.17
	Assembly	0.14	0.13	0.12	0.12	0.12	0.13	0.14	0.14	0.16	0.16	0.16	0.16
	GISS-E2-R	0.16	0.15	0.14	0.14	0.14	0.14	0.15	0.15	0.16	0.17	0.17	0.16
	IPSL-CM5A-MR	0.14	0.14	0.12	0.13	0.13	0.14	0.16	0.16	0.17	0.16	0.16	0.17
	MIROC-ESM	0.16	0.15	0.14	0.13	0.13	0.14	0.16	0.17	0.17	0.18	0.17	0.17
RMSE	CSIRO-Mk3-6-0	0.67	0.76	0.78	0.80	0.73	0.54	0.44	0.47	0.58	0.69	0.75	0.71
	Assembly	0.61	0.63	0.66	0.69	0.62	0.54	0.43	0.43	0.58	0.71	0.73	0.69
	GISS-E2-R	0.68	0.76	0.81	0.81	0.71	0.60	0.44	0.45	0.59	0.72	0.79	0.67
	IPSL-CM5A-MR	0.62	0.68	0.70	0.72	0.64	0.55	0.48	0.47	0.57	0.70	0.75	0.69
	MIROC-ESM	0.67	0.74	0.75	0.77	0.66	0.58	0.48	0.52	0.62	0.76	0.79	0.73
MAE	CSIRO-Mk3-6-0	0.51	0.57	0.59	0.62	0.52	0.38	0.33	0.32	0.42	0.52	0.56	0.54
	Assembly	0.46	0.48	0.50	0.55	0.46	0.38	0.31	0.31	0.41	0.52	0.56	0.51
	GISS-E2-R	0.53	0.56	0.62	0.63	0.53	0.42	0.33	0.31	0.42	0.54	0.62	0.51
	IPSL-CM5A-MR	0.46	0.51	0.52	0.58	0.49	0.41	0.36	0.33	0.42	0.51	0.57	0.53
	MIROC-ESM	0.52	0.58	0.59	0.60	0.50	0.41	0.35	0.37	0.44	0.57	0.61	0.52
MCM	CSIRO-Mk3-6-0	0.57	0.55	0.55	0.54	0.59	**0.64**	0.59	0.54	0.57	0.55	**0.55**	0.52
	Assembly	**0.61**	**0.63**	**0.65**	**0.61**	**0.65**	0.63	**0.60**	0.57	**0.58**	**0.55**	0.54	0.53
	GISS-E2-R	0.55	0.56	0.53	0.52	0.59	0.57	0.58	**0.57**	0.58	0.53	0.50	**0.54**
	IPSL-CM5A-MR	0.60	0.60	0.62	0.59	0.64	0.61	0.55	0.53	0.57	0.55	0.54	0.53
	MIROC-ESM	0.55	0.55	0.57	0.56	0.62	0.59	0.55	0.47	0.55	0.51	0.52	0.53