https://doi.org/10.26682/csjuod.2023.26.2.70
Journal of University of Duhok., Vol. 26, No.2 (Pure and Engineering Sciences), Pp 795-808, 2023
4th International Conference on Recent Innovations in Engineering (ICRIE 2023) (Special issue)

EVALUATION OF M5P DECISION TREE MODEL IN DOWNSCALING CMIP6 CLIMATE OUTPUT FOR ERBIL PLAIN

ABBAS YEGANEH-BAKHTIARY * **HOSSEIN EYVAZOGHLI** **, **SOORKEU A. ATROOSHI***, **SARFRAZ MUNIR***, **ILYAS MASIH***** and **ANNE VAN DAM*****
*School of Science & Engineering, University of Kurdistan Hewler, Kurdistan Region- Iraq
**School of Civil Engineering, Iran University of Science and Technology, Iran
*** Dept. of Water Resources and Ecosystems, IHE Delft Institute, Netherland

(Accepted for Publication: November 27, 2023)

ABSTRACT

Downscaling the effective parameters of Global Climate Models is crucially required to project climate conditions in the future. The capability of machine learning approaches in downscaling Global Climate Models is getting interesting these days. The present study evaluates the M5p Decision Tree (DT) skill in reproducing high resolution monthly precipitation and temperature (predictands) data. To this end, the significant climate-related parameters (predictors) were derived from the General Climate Model of the Coupled Model Intercomparison Project Phase 6 (CMIP6) for the Erbil Plain. Initially, the effective parameters were carefully chosen for developing a downscaling model. Subsequently, multiple models were formulated to accommodate various maximum depths of decision trees (DT). Results obtained from the training process revealed a notably higher correlation between precipitation and temperature predictors in contrast to wind speed and direction. The evaluation of skill indicated enhanced accuracy in downscaling when increasing the maximum depth (MD) of M5p models up to an optimal threshold, with MD = 5 identified as the optimal depth for generating predictive DTs. Finally, it was proved that the M5p model serves as a highly effective tool for downscaling the hydroclimatic parameters in climate change studies.

KEYWORDS: Machine learning, regression, climate change, trend, climatic parameters.

1. INTRODUCTION

Climate change is one of the most important challenges challenge confronting humanities. A recent investigation highlighted a concerning trend that the Euphrates River has experienced a minimum 40 percent decline in its downstream flow since 1972, with projections indicating a further decrease in the river's overall flow in the upcoming decade. Paradoxically, Iraq is endowing with abundant water resources within the region, facing mounting challenges in ensuring water availability (IAUIraq, 2012). The scarcity of water emerges as a direct consequence of climate change (IPCC, 2021), exacerbating the complexities faced by Iraq in managing its water resources.

Global Climate Models (GCMs) have simulated large-scale climate data until the close of century, incorporating the impacts of changes in greenhouse gas (Kamranzad et al., 2014; Swart et al., 2019). Many research works have utilized GCMs to forecast the ramifications of climate

abbas.yeganeh@ukh.edu.krd
hossein.eyvazoghli@gmail.com
soorkeu@gmail.com
munir.sarfraz@gmail.com
i.masih@un-ihe.org
a.vandam@un-ihe.org

change on the precipitation and temperature (Kamranzad, 2014; Zheng et al., 2017). Downscaling is a way to connect the large-scale GCM results to local-scale climate variables (Sachindra et al., 2018; Yeganeh-Bakhtiary et al., 2022).

Three classes of downscaling approaches have been adopted: empirical, semi-empirical, and nesting methods. In the empirical approach, employes the historical climatic data to present local analog scenarios. These studies entail a qualitative conceptual survey, and the obtained results derived from empirical approaches do not generate a climate forecasting model. Conversely, Semi-empirical (statistical) and nested (dynamical) downscaling approaches employ predictions from large-scale GCM models to develop the local climate change scenarios (Sailor et al., 2000). In the dynamical downscaling method, a regional climate model (RCM) operates at the desired mesh resolution utilizing outputs from large-scale GCM models as boundary conditions to generate higher resolution outputs (Fowler et al., 2007). However, the significant drawbacks of dynamical downscaling methods, which restricted their applications in climate change assessments, include their complexity, high computational cost, and performance sensitive to specific cases (Ghosh and Misra, 2010).

Statistical downscaling methods can be divided into three groups (Maraun et al., 2010; Srinivasa Raju and Nagesh Kumar, 2018): the Perfect Prognosis (PP) method, establishes a relationship is between large-scale observational data and locally recorded data (Wilby and Wigley, 2000; Trigo and Palutikof, 2001). In the Model Output Statistics (MOS) method, akin to the PP method, except that in this approach, a relationship is forged between GCM outputs (predictor) and local climate variables (predictands) (Zhang, 2005). The Stochastic Weather Generator (SWG) category develops a relationship by altering probable distribution parameters (Chen et al., 2013). Considering the numerous parameters involved in GCM simulations and the future scenarios intended to simulate trends, the MOS method may be suitable for downscaling when only limited observational data are accessible compared to the PP approach, (see Srinivasa Raju and Nagesh Kumar, 2018; Zhang et al., 2021).

On the other hand, the statistical downscaling methods are developed based on two fundamental assumptions: (i) the establishment of empirical relationships between historical large-scale atmospheric predictors modelled by GCMs and local climate characteristics, and (ii) the validity of these derived empirical relationships under varying climate change scenarios (Sunyer et al., 2012; Kamranzad et al., 2013). The predominant downscaling approach involves devising transfer functions that fit a quantitative relationship between large-scale climate variables and local-scale climate variables. are the most popular statistical downscaling approach. In recent years, machine learning techniques have emerged as a method to determine the required transfer function in statistical downscaling (see among others, Sachindra et al., 2018; Davanlou Tajbakhsh et al., 2019).

Due to the nonlinear nature of climatic processes in predication process, machine learning techniques such as artificial neural networks (ANN) find widespread adoption in modeling and forecasting various characteristics of climatic parameters (e.g., Sailor et al., 2000; Avila et al., 2020). On the other hand, Decision Tree (DT) as one of the most popular and efficient data mining techniques, characterized by its clustering approach, offers the advantage of logical arranging variable choices, thereby leading to results that are both comprehensive and straight forward (Nourani and Molajou, 2017). Comparison assessment between Decision Tree models with other machine learning techniques demonstrate that DT models provide more accurate predictions (Londhe and Dixit, 2012;

Mahtabi and Sattari, 2016).

This study investigated the potential of the M5p DT as a classification-based machine learning for hindcasting precipitation and temperature in the Erbil Plain. Variations of predicted precipitation and temperature simulated by the GCM, specifically the CanESM5 model (Swart et al., 2019) during the historical period were downscaled by the proposed approach and assessed using reanalyzed data (ERA5-Land). Notably this research marks the first time of studying different designs of M5p DT for the prediction of precipitation and temperature.

2. METHODOLOGY

Long-term rainfall forecasting is very crucial for resource management and policymaking in Erbil plain. This long-term precipitation is affected significantly by climate change, and downscaling of GCMs data provides a robust model to explore the climate change impacts on precipitation and other hydroclimatic parameters. Recently, the application of Decision Trees has increased rapidly to predict the climatic parameters. This section discusses the data and methods that were used in this study.

2.1 Study Area

The study area includes the entire of Erbil province, located within geographical coordinate of latitude of 34°42'-37°22' N and longitudes of 42°25'-46°15' E. Geographically, Erbil Province covers an area of 14,818.1 km^2 and comprises 10 districts: Mergasur, Soran, Choman, Rawanduz, Shaqlawa, Khabat, Erbil, Dashti Hawler, Koysinjaq, and Makhmur. The climate in the IKR is characterized as arid and semi-arid, exhibiting hot and dry conditions in summer and cold and wet in winter. Spring and autumn are relatively short compared to the dominant summer and winter seasons (Gaznayee and Al-Quraishi, 2019). Erbil Province displays a considerable elevation variation ranging between 100 m and 3,565 meters above sea level, as shown in **Error! Reference source not found.**.

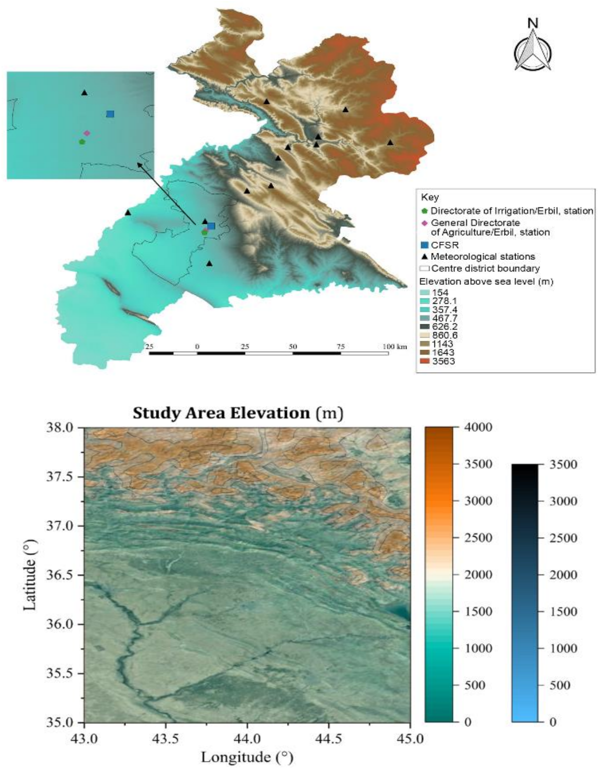


Fig.(1): -Erbil topography and the location of meteorological stations. left: Erbil region (Mustafa et al., 2019), right: considered study area.

Error! Reference source not found. shows the spatial distribution of the mean annual rainfall across the Erbil Plain. As seen a decreases in precipitation amount from the northeast toward the southwest of Erbil Province can be observed; this signifies that the northeastern regions receive higher quantities of precipitation than the southwestern areas of the Erbil city (Mawlood and Al-Quraishi, 2019).

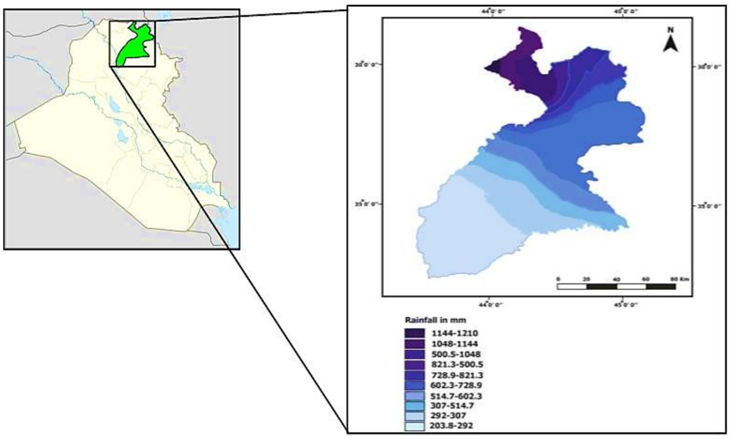


Fig.(2): -Spatial distribution of average annual rainfall in Erbil Plain (Mawlood and Al-Quraishi, 2019).

2.2Global Climate Model (GCM)

Climate change exhibits geographical variation and is not homogenous across the planet, with certain areas are more sensitive to climate change than others. The coupled Atmosphere-Ocean-Land General Circulation Models (GCMs), initially designed for weather forecasting, have evolved into tools for understanding climate dynamics and projecting climate change. The projections concerning climate change using GCMs is based on future scenarios, accounting for the greenhouse gas emissions at each scenario that drive concentrations the greenhouse gas in the atmosphere. As the impacts of climate change are increasingly evident, there is an urgent need to act based on reliable scientific information. The Canadian Centre for Climate Modelling and Analysis (CCCma) is actively engaged in an ongoing effort to improve the modelling of the global Earth system. This aims to advance understanding of climate system’s functionality variability, historical changes, and foster improved quantitative predictions and forecast of future climate. CanESM5, the current version of CCCma's global model, has a pedigree extending back 40 years to the inception of the first atmospheric general circulation model (GCM) developed at CCCma's predecessor, the Canadian Climate Centre.

CanESM5 represents a major upgrade from CanESM2, featuring incremental improvements across the atmosphere, land surface, and terrestrial ecosystem models. The major changes relative to CanESM2 encompass the introduction of completely new models for the ocean, sea ice, and marine ecosystems, and a novel coupler. Developers of the model face decisions on allocating limited computational resources among augmentation in model resolution, model complexity, and model throughput (i.e., number of simulated years). maintain a resolution akin to CanESM5 (T63 or ∼ 2.8˚ in the atmosphere and ∼ 1˚ in the ocean), situating it at the lower end of the spectrum of CMIP6 models. The advantage of this coarse resolution lies in its relatively high model throughput given the model’s complexity, enabling simulation of many years within available computational resources.

Among the multiple parameters provided by the GCM, three specific parameters- total precipitation, near-surface temperature, and both U and V components of wind speed- were selected and prepared as inputs (predictors) for the machine learning model. Table (**1** summarizes the mentioned variables along with their description. It should be noted that, akin to the control period, all three parameters were prepared for the forecasting period under three distinct scenarios: SSP1-1.9, SSP5-8.5, and SSP4-6.0. These scenarios represent three global warming scenarios: minimum, medium, and maximum. The most recent set of scenarios, used for CMIP6 (2016-2021) and IPCC Sixth Assessment Report (AR6) (2021) are known as the Shared Socio-economic Pathways (SSPs). These SSPs scenarios represent the most complex scenarios created to date and encompass a spectrum from highly ambitious mitigation strategies to ongoing emissions growth.

Table (1): -The predictor variables and their description.

Variables name	Unit	Description
Precipitation	kg m^{-2} s^{-1}	The sum of liquid and frozen water, comprising rain and snow, that falls to the Earth's surface. It is the sum of large-scale precipitation and convective precipitation. This variable represents amount of water per unit area and time.
Near-Surface air temperature	K	Temperature of air at 2m above the surface of land, sea or inland waters. 2m temperature is calculated by interpolating between the lowest model level and the Earth's surface, taking account of the atmospheric conditions.
Northward near-surface wind	m s^{-1}	Magnitude of the northward component of the two-dimensional horizontal air velocity 10m above the surface.
Eastward near-surface wind	m s^{-1}	Magnitude of the eastward component of the two-dimensional horizontal air velocity 10m above the surface.

2.3 Observed Data

In regions with a limited instrumental coverage, reanalysis products are used as a viable alternative to observational datasets. Reanalysis datasets offer a distinct advantage owing to their comprehensive global coverage and extensive temporal scope. To facilitate the production of contemporary reanalysis data, a fusion of data from European Centre for Medium-Range Weather Forecasts (ECMWF), ERA5 satellite, and gauge measurements were assimilated. Notably the latest generation of the ECMWF reanalysis, ERA5, replaced the previously successful ERA-Interim reanalysis. ERA5 operates on a function of four-dimensional variational (4D-Var) data assimilation, integrating a soil model and an ocean wave model into its framework.

Atmospheric data from ERA5 spans 137 hybrid vertical levels and is available on the Climate Data Store (CDS), interpolated to 37 pressure levels ranging from 1000 hPa (near the surface) to 1 hPa (at approximately 80 km altitude). An enhanced version focusing on the land component, termed as a ERA5-land, has been developed using a horizontal resolution model and uses the ERA5 data as a primary input. This enhancement enables a horizontal resolution of 9 km at hourly intervals. Here in this study, ERA5-Land data served as the observed data set, presented on a regular 0.1° × 0.1° grid, specifically covering terrestrial surfaces. The annual average precipitation recorded fell in the range of 1-2.5 *mm*, with temperature values approximating 15 °*C*. These metrics delineate the environmental parameter essential for the analysis conducted in this study.

2.4 M5p Decision Tree

Recently, Decision Tree (DT) algorithm have garnered significant attention as a robust machine learning technique for predicting of hydroclimatic parameters (see among others Abolfathi et al., 2016 and Avila et al., 2020; Yeganeh-Bakhtiary et al., 2022). Decision Trees are versatile, supporting both classification and regression problems by creating a tree structure to evaluate data instances. This tree is inverted, commencing from the root and progressing down to the leaves until a prediction can be established. Predictions are made iteratively, refining the tree it reaches to a tree with a fixed depth. Post- construction, pruning techniques may be applied to improve the model's ability to generalize to new data.

In general, a typical Decision Tree is comprised of four main parts: root, branch, node, and leaves. The root (or initial node) is at the top of the tree's apex, while the leaves (or the last nodes) mark the end of the chain of branches and nodes. **Error! Reference source not found.** depicts the flow chart of the M5p DT model used as a predictive model in this study. As Decision Trees can be considered as a graphical method, the interpretations of DT model outputs offer a more visible and understandable representation compared to the other machine learning methods (Abolfathi et al., 2016). This adaptability and the interpretability of Decision Trees make them a compelling choice for hydroclimatic parameter prediction, facilitating both effective modeling and comprehensible presentation of results.

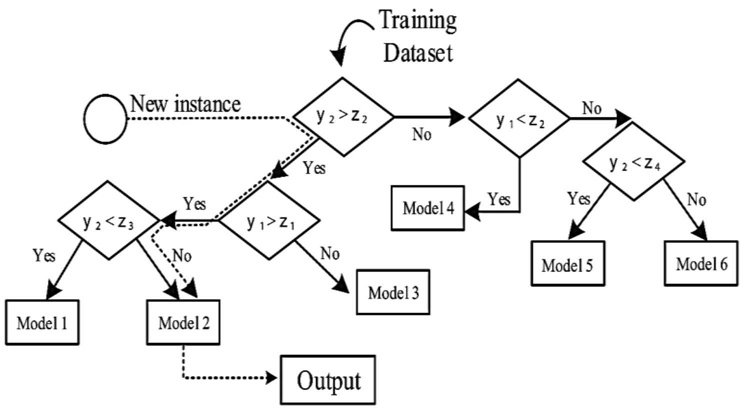


Fig.(3): -Flow chart of M5p model prediction procedures (Abolfathi et al., 2016)

Considering the primary objective of categorization is to develop a predictive model, it is suggested to use the M5p DT as a robust machine learning approach for forecasting climatic parameters. Unlike neural networks, Decision Tree yields to the interpretable rules. The prediction derived from the tree is explained through a series of rules, providing a transparent insight into the reasoning behind the predications. Whereas neural networks solely present the predicted results, concealing the process through which the prediction was generated within the network. This distinction highlights the advantage of Decision Trees in offering transparent and understandable rules for prediction, enhancing interpretability, and facilitating a clearer understanding of the model's decision-making process.

2.5 Study Procedure

Data preparation, model development, evaluation, and prediction are the main steps of downscaling employing a machine learning approach. In the data preparation, we needed to generate a calculation network. All the models' developments, verifications, and predictions will be done on the nodes of the network. Since the observation was presented in a 0.1° × 0.1° resolute grid, the calculational network was generated in the same resolution.

Collecting observed and GCM data is the next step in data preparation. In this study, the GCM's simulated precipitation (*Pr*), temperature (*T*), and wind characteristics (wind speed and direction) are selected as the primary predictors (forecasting model's inputs). Besides the climatic variables, topographic terms, including the longitude, latitude, and altitude were also employed to predict *Pr* and *T*. It is noteworthy that the altitude data are collected from GEBCO web-based application and gridded to the generated network by averaging the gathered data. On the other hand, observed precipitation and temperature are used to develop and verify the downscaling and prediction models, which are called the predictands.

A preliminary examination of the correlation between the chosen predictors and predictands is necessary to select the possible effective variables. As the M5'-based models are multilinear rules, it can be expected that the predictands correlated higher than linear correlation and lower than non-linear. Accordingly, both the Spearman and Pearson correlation coefficients are investigated for prediction models inputs.

For the next step, considered models were developed under the several limited max depths of the tree condition based on the first part of the data (the train section) to predict precipitation and temperature. The developed models' performance was evaluated using several error detection criteria. The results had shown the best max depth of the tree, and the model was employed to predict future precipitation and temperature. **Error! Reference source not found.** illustrates the downscaling and predicting procedure.

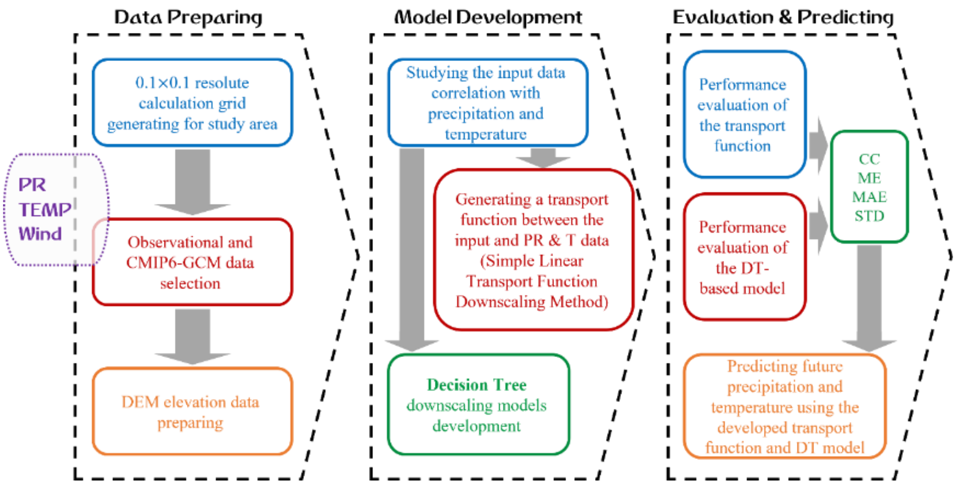


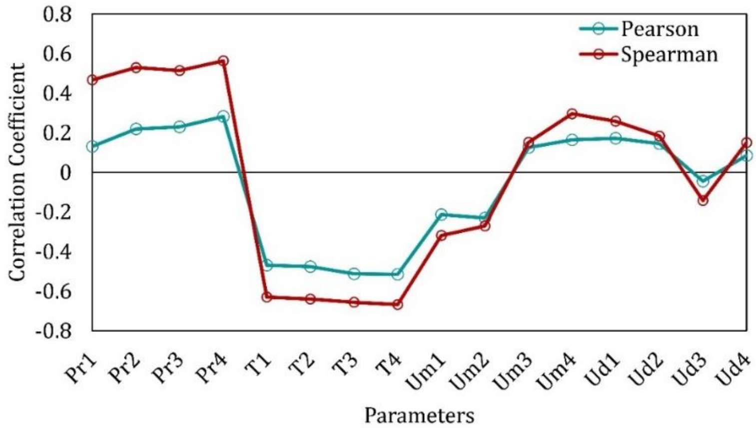
Fig.(4): -Schematic of the downscaling methodology and predictive modelling processes

3. RESULTS AND DISCUSSION

To develop a robust downscaling and forecasting model, a comprehensive collection of variables that potentially influencing the results were collected from the CMIP6 GCM (CanESM5) simulations. But prior to constructing the models based on these variables, an assessment of the correlation between these variables and predictands (*Pr* and *T*) was imperative. A strong correlation can lead to a more robust model development and better prediction of *Pr* and T. Conversely, the prediction models built upon weak correlated inputs can yield less accurate results. Precipitation, temperature, wind speed in both north and east directions, and topographical specifications (longitude, latitude, and altitude) were selected as the input data to derive the prediction model. The correlation analyses were conducted between these input parameters and predictand to identify the most influential parameters.

Error! Reference source not found. shows the results of correlation evaluation between the selected GCM outputs and ERA5 for both precipitation and temperature. It is evident from **Error! Reference source not found.** that the input wind specifications exhibited limited alignment with the predictand precipitation, while the input temperatures displayed a significant inverse correlation with precipitation. The results indicate that a decrease in temperature corresponds to an increase in precipitation, inlining with the rising temperature enhance the potential of rainfall.

(a) Precipitation



(b) Temperature

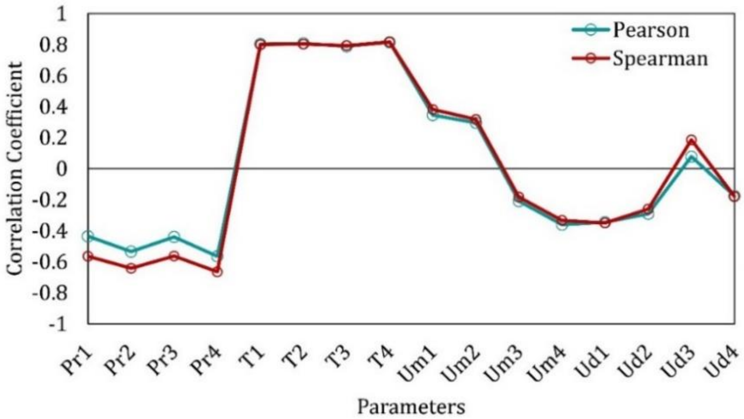


Fig.(5): - Correlation coefficient of the selected GCM outputs and ERA5: (a) Precipitation, (b) Temperature.

Input wind specifications exhibited limited alignment with the predictand precipitation, while input temperatures showed a noteworthy inverse correlation with precipitation. A decrease in temperature corresponds to an increase in precipitation, suggesting a discernible relation between these parameters. Further analyses indicated a positive correlation between input precipitations and the predictand *Pr*. The graphical illustration (Figure 5) demonstrates the strong correlation between GCM precipitation and temperature, particularly in relation to ERA5 temperature. Additionally, the close alignment between linear and non-linear indexes implies the predictability of this phenomenon using a linear model. Therefore, it is anticipated that the M5p based models may exhibit greater accuracy in predicting temperature compared to precipitation. On the other hand, it is important to note that the equations (leaves) within the models do not unilaterally ensure improving prediction accuracy. Instead, there exists an optimal value, where the models' performance reaches its peak. To ascertain this optimal value, we evaluate the effect of the prediction models' maximum depth on their performance. The maximum depth (MD) value was systematically adjusted to 1 to 10, 12, 15, and 20. **Error! Reference source not found.**6 illustrates the variations in the linear correlation coefficient between the predicted precipitation and temperature against ERA5 records versus to change in the tree's maximum depth.

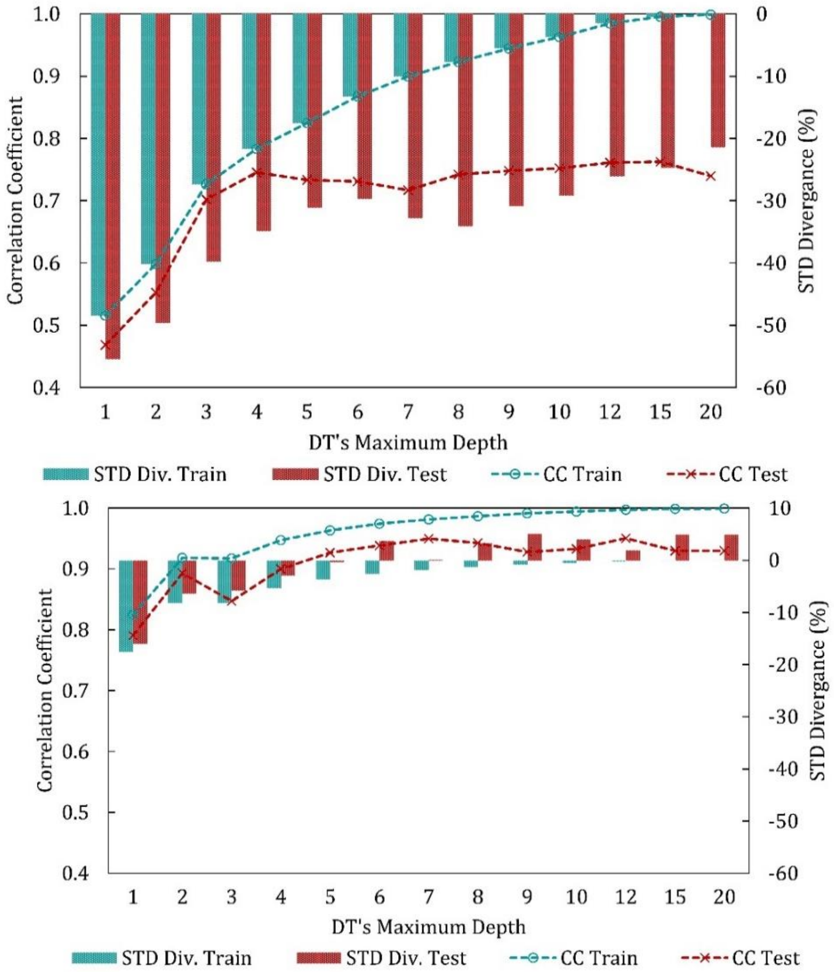


Fig.(6): -Correlation coefficient (lines) and STD divergence (column bars) of the predicted precipitation (up) and temperature (down) with ERA5 values versus DT's max depth variation

As seen from Figure 6, a notable trend in the correlation coefficient as the tree depth increases during the training period, reaching its zenith at MD = 20, where it closely reaches *CC*= 1.0. However, during the test period, the correlation coefficient shows an upper limit, with values around *CC*≃ 0.75 for precipitation and *CC*≃ 0.9 for temperature. This indicates that evaluating the maximum depth of the prediction model does not uniformly argument the correlation index between predicted and actual values. Similar behavior was observed in the *STD* divergence trends. While in the training phase, increasing the tree depth leads to decreased *STD* divergence, approaching minimal differences between *STD*s and reaching close to zero. During the verification phase, higher MD values indicated a limited divergence, around 5% for temperature and approximately 30% for precipitation predictions. This analysis sheds light on the intricate relationship between model complexity, represented by maximum depth values, and the resulting predictive accuracy.

A Comparison of precipitation and temperature curves reveals disparities in the models' performance. The temperature predictor exhibited superior forecasting capabilities compared to the precipitation model, as evidenced by higher *CC* and reduced *STD* divergence measures. **Error! Reference source not found.** illustrates this discrepancy, with the precipitation model demonstrating a minimum *CC* of 0.45 in the training phase, while the weakest performance for temperature was *CC*≃ 0.75.

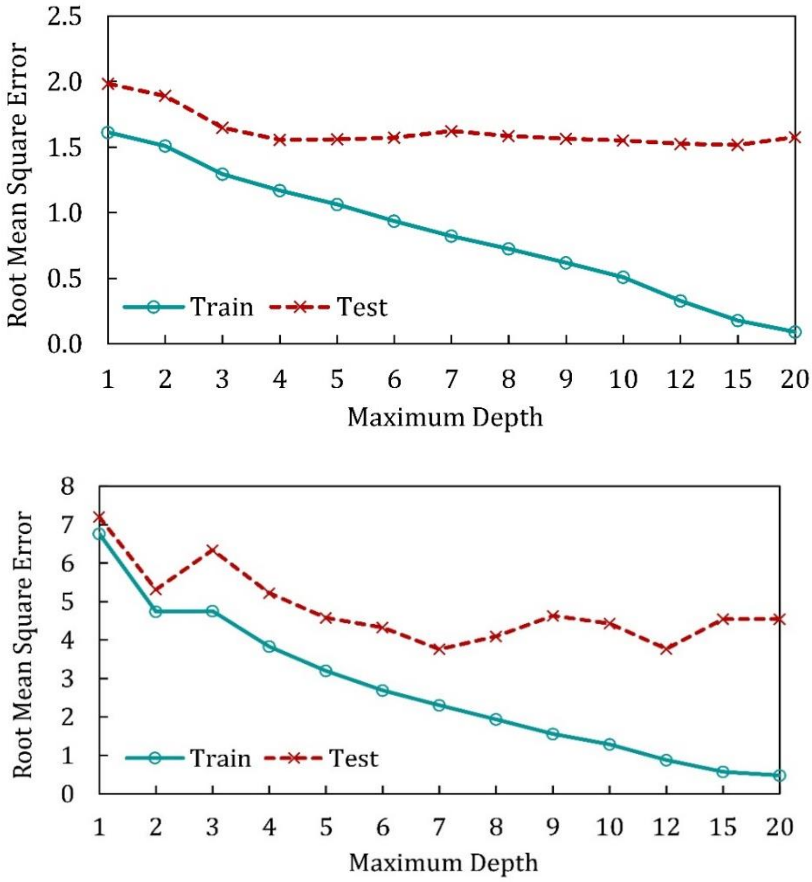


Fig.(7): -Prediction RMSE of the generated models by changing the DT's max depth: precipitation (up) and temperature (down)

Ultimately the *CC* value converged to approximately 0.75 and 0.95 for precipitation and temperature, respectively. Furthermore, the difference between the *STD* of ERA5 and predicted values are approximately 30% for precipitation and only 5% for temperature. The superior performance of the temperature predictor suggests its higher predictability compared to precipitation within the studied context, indicating the need for nuanced modeling approaches to accurately capture and forecast precipitation dynamics.

For a comprehensive evaluation, assessing prediction errors alongside *CC* and *STD* divergence is essential. For this purpose, the RMSE values were calculated for Precipitation and Temperature prediction. Figure 7 illustrates the trend where RMSE decreases as the prediction tree size increases in the training phase for precipitation and temperature. But this reduction reaches a plateau during the test period. From the **Error! Reference source not found.** and 7 collectively suggest that while larger tree size enhance the performance of the predication models during training phase, the models' capability becomes constrained even with trees of *MD* = 20. Consequently, determining the optimum model size is imperative based on observation. From the depicted trends, *MD* = 5 was identified as the optimum maximum depth for future evaluations and predictions, given that the since the model's performance did not improve beyond *MD* = 5.

Evaluating the spatial distribution of the prediction errors across the study area is crucial for a detailed assessment of the selected model (*MD* = 5). Figurer 8 depicts the mean absolute error (*MAE*) of precipitation and temperature predictions across the study area for both the training and test phases.

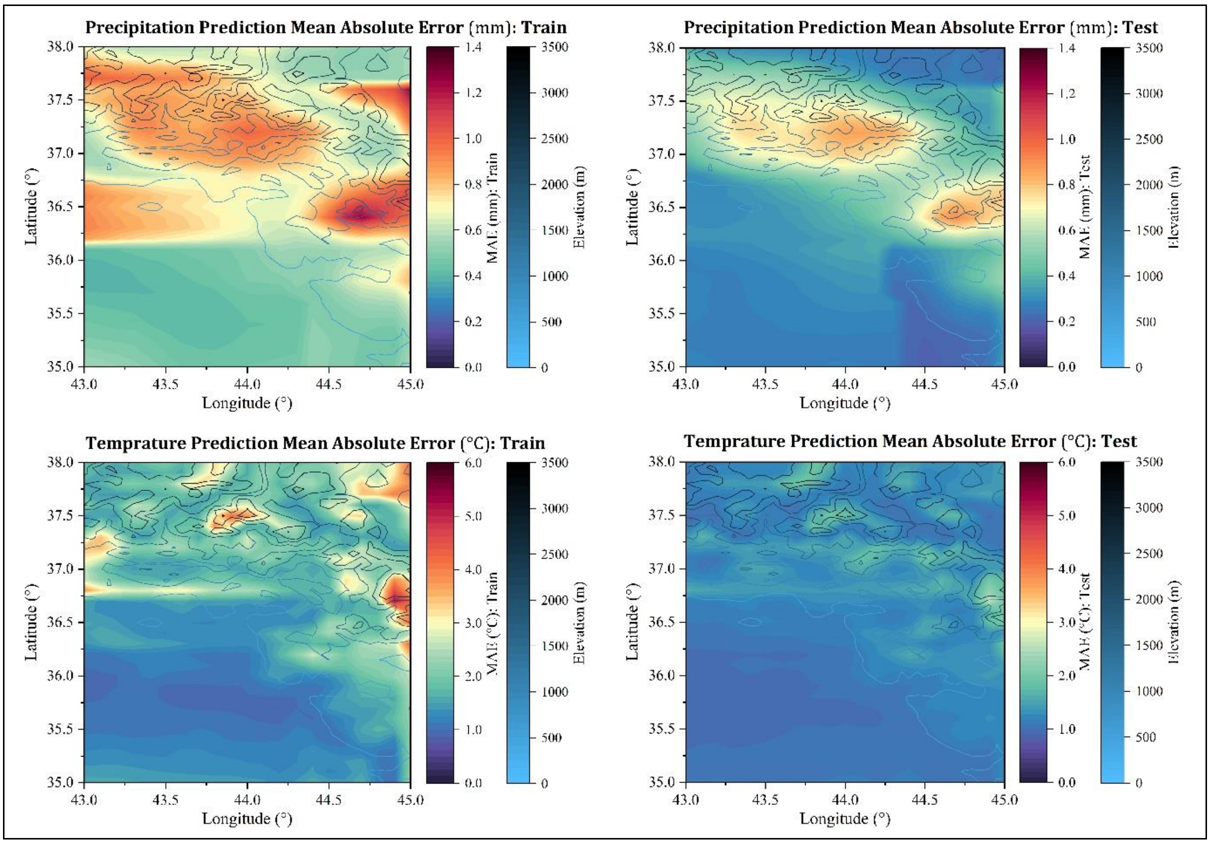


Fig.(8): -Spatial distribution of the mean absolute error of predicted precipitation (up row) and temperature (down row) during the train (left column) and verification (right column).

It is evident from Figure 8 that, while the prediction RMSE in the test phase was consistently larger than that of the training phase across the entire calculation region, the spatial distribution reveals an increase in prediction error during verification phase compared to the training phase. Further analysis of the prediction values with respect to the grid's nodes altitudes in Fig.(8 indicates that MAE escalated with higher elevations of topography. This trend may be attributed to the variation of data density across lower altitudes. Given that a significant portion of the training data belongs to lower altitudes, the predictions trend to be more accurate in these regions due to errors, highlighted the influence of data distribution across varying topographic area on model accuracy.

4. CONCLUSION

This study focused on assessing the effectiveness of the M5p DT model in downscaling CMIP6 climate output for the Erbil Plain. The evaluation of downscaled historical data was performed using *CC*, *MAE*, and *RMSE* metrics, highlighting the suitability of the M5p DT model in downscaling hydroclimatic parameters. The main findings can be summarized as follows:

- The correlation analysis indicated that wind speed and direction exhibited a weaker correlation with *Pr* and *T* compared to stronger correlation observed between precipitation and temperature predictors.
- Both linear and non-linear correlation indexes showed a closer relation between the predictors and temperature predictand. Notably, the predictive accuracy of temperature suppressed that of the precipitation.
- The obtained results showed that increasing the depth of DT led to more accurate predictions within the training dataset and procedural phases.
- The skill assessment results demonstrated a marked enhancement in downscaling accuracy by escalating the maximum depth of M5p models, reaching an optimum point at MD = 5; thereby generating highly predictive DTs.

Overall, the obtained results contribute to the growing knowledge of downscaling techniques and emphasize on the potential of employing the M5p DT model as a valuable tool in projection future climatic trends in Erbil Plain and other regions with comparable climate setting.

REFERENCES

Abolfathi, S., Yeganeh-Bakhtiary, A., Hamze-Ziabari, S.M., Borzooei, S., 2016. Wave runup prediction using M5′ model tree algorithm. Ocean Engineering 112, 76-81. https://doi.org/10.1016/j.oceaneng.2015.12.016

Avila, D., Marichal, G.N., Padrón, I., Quiza, R., Hernández, Á., 2020. Forecasting of wave energy in Canary Islands based on Artificial Intelligence. Applied Ocean Research 101, 102189. https://doi.org/10.1016/j.apor.2020.102189

Chen Y, Zhang Q, Xiao M, Singh V., 2013. Evaluation of risk of hydrological droughts by the trivariate Plackett copula in the East River basin (China). Natural Hazards: Journal of the International Society for the Prevention and Mitigation of Natural Hazards, 68:529-47.

Davanlou Tajbakhsh, A., Nourani, V., Molajou, A., 2019. Hybrid wavelet-M5 model application in rainfall-runoff process forecast (Case study: Aji Chay catchment). Iran-Water Resources Research, 15 (2 #l00929), 1-10.

Etemad-Shahidi, A., Taghipour, M., 2012. Predicting longitudinal dispersion coefficient in natural streams using M5p model tree. Journal of Hydraulic Engineering 138 (6), 542-554. https://10.1061/(ASCE)HY.1943-7900.0000550

Fatli, T., 2018. Water scarcity and population displacement in southern Iraq: Perceptions and reality, Master's Thesis, American University in Cairo, AUC Knowledge Fountain.

Fowler HJ, Blenkinsop S, Tebaldi C., 2007. Linking climate change modelling to impacts studies: recent advances in downscaling techniques for hydrological modelling. International Journal of Climatology, 27:1547-78.

Gaznayee, H., Al-Quraishi, A., 2019. Analysis of Agricultural Drought’s Severity and Impacts in Erbil Province, the Iraqi Kurdistan Region based on Time Series NDVI and TCI Indices for 1998 through 2017. Journal of Advanced Research in Dynamical and Control Systems. 10.5373/JARDCS/V11I11/20193198

Ghosh S, Misra C., 2010. Assessing hydrological impacts of climate change: Modeling techniques and challenges. The Open Hydrology Journal, 410:115-21.

IAUIraq, 2012. Inter-Agency Information and Analysis Unit: Climate change in Iraq, https://reliefweb.int/sites/reliefweb.int/files/resources/Climate%20change%20In%20Iraq%20Fact%20sheet%20-%20English.pdf,

IPCC, 2007. Intergovernmental Panel on Climate Change, Climate Change 2007: The Physical Science Basis. Contribution of Working Group

I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge

IPCC, 2014. Intergovernmental Panel on Climate Change, Climate Change 2013: The Physical Science Basis. Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge. https://10.1017/CBO9781107415324

IPCC, 2021. Intergovernmental Panel on Climate Change, Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge

Kamranzad, B., Etemad-Shahidi, A., Chegini, V., Hadadpour, S., 2013. Assessment of CGCM 3.1 wind field in the Persian Gulf. Journal of Coastal Research 65 (sp1), 249-253. https://10.2112/SI65-043.1

Londhe, S.N., Dixit, P., 2012. Forecasting stream flow using support vector regression and M5 model trees. International Journal of Engineering Research and Development 2 (5), 1-12. https://www.idc-online.com/technical_references/pdfs/civil_engineering/Forecasting%20Stream.pdf

Mahjoobi, J., Etemad-Shahidi, A., 2008. An alternative approach for the prediction of significant wave heights based on classification and regression trees. Applied Ocean Research 30 (3), 172-177. https://doi.org/10.1016/j.apor.2008.11.001

Mahtabi, G., Sattari, M., 2016. Investigation of hydraulic jump characteristics in rough beds using M5 model tree. Jordan Journal of Agricultural Sciences 12, 631-648. http://dx.doi.org/10.12816/0030042

Makkeasorn, A., Chang, N.B., Zhou, X., 2008. Short-term streamflow forecasting with global climate change implications – A comparative study between genetic programming and neural network models. Journal of Hydrology 352 (3), 336-354. https://doi.org/10.1016/j.jhydrol.2008.01.023

Maraun D, Wetterhall F, Ireson AM, Chandler RE, Kendon EJ, Widmann M., 2010. Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user. Reviews of Geophysics. 2010;48.

Mawlood, D.K., Al-Quraishi, A., 2019. Sustainability of aquifer and ground water condition in Erbil Basin-Iraq. ZANCO Journal of Pure and Applied Sciences 31 (6), 51-60. http://dx.doi.org/10.21271/ZJPAS.31.6.6

Mustafa, A., Muhammed, H.H., SzydBowski, M., 2019. Extreme rainfalls as a cause of urban flash floods; a case study of the Erbil-Kurdistan Region of Iraq. Acta Scientiarum Polonorum Formatio Circumiectus

Nourani, V., Davanlou Tajbakhsh, A., Molajou, A., Gokcekus, H., 2019a. Hybrid wavelet-M5 model tree for rainfall-runoff modeling. Journal of Hydrologic Engineering 24 (5), 04019012. https://10.1061/(ASCE)HE.1943-5584.0001777

Nourani, V., Molajou, A., 2017. Application of a hybrid association rules/Decision Tree model for drought monitoring. Global and Planetary Change 159, 37-45. https://doi.org/10.1016/j.gloplacha.2017.10.008

Nourani, V., Molajou, A., Uzelaltinbulat, S., Sadikoglu, F., 2019b. Emotional artificial neural networks (EANNs) for multi-step ahead prediction of monthly precipitation; case study: northern Cyprus. Theoretical and Applied Climatology 138 (3), 1419-1434. https://10.1007/s00704-019-02904-x

Sachindra, D.A., Ahmed, K., Rashid, M.M., Shahid, S., Perera, B.J.C., 2018. Statistical downscaling of precipitation using machine learning techniques. Atmospheric Research 212, 240-258. https://doi.org/10.1016/j.atmosres.2018.05.022

Sailor, D.J., Hu, T., Li, X., Rosen, J.N., 2000. A neural network approach to local downscaling of GCM output for assessing wind power implications of climate change. Renewable Energy 19 (3), 359-378. https://doi.org/10.1016/S0960-1481(99)00056-7

Sattari, M.T., Pal, M., Apaydin, H., Ozturk, F., 2013. M5 model tree application in daily river flow forecasting in Sohu Stream, Turkey. Water Resources 40 (3), 233-242. https://10.1134/S0097807813030123

Solomatine, D.P., Dulal, K.N., 2003. Model trees as an alternative to neural networks in rainfall—runoff modelling. Hydrological Sciences Journal 48 (3), 399-411. https://10.1623/hysj.48.3.399.45291

Solomatine, D.P., Xue, Y., 2004. M5 Model Trees and Neural Networks: Application to Flood Forecasting in the Upper Reach of the Huai River in China. Journal of Hydrologic Engineering 9 (6), 491-501. https://10.1061/(ASCE)1084-0699(2004)9:6(491)

Sunyer MA, Madsen H, Ang PH., 2012. A comparison of different regional climate models and statistical downscaling methods for extreme rainfall estimation under climate change. Atmospheric Research,103:119-28.

Swart, N.C., Cole, J.N.S., Kharin, V.V., Lazare, M., Scinocca, J.F., Gillett, N.P., Anstey, J., Arora, V., Christian, J.R., Hanna, S., Jiao, Y., Lee, W.G., Majaess, F., Saenko, O.A., Seiler, C., Seinen, C., Shao, A., Sigmond, M., Solheim, L., von Salzen, K., Yang, D., Winter, B., 2019. The Canadian Earth System Model version 5 (CanESM5.0.3). Geoscientific Model Development 12, 4823-4873. https://10.5194/gmd-12-4823-2019

Trafalis, T.B., Santosa, B., Richman, M.B., 2005. Learning networks in rainfall estimation. Computational Management Science 2 (3), 229-251. https://10.1007/s10287-005-0026-0

Trigo RM, Palutikof JP., 2001. Precipitation Scenarios over Iberia: A Comparison between Direct GCM Output and Different Downscaling Techniques. Journal of Climate, 14:4422-46.

Vakili, S.H., 2018. Forecasting of monthly precipitation using M5 model tree and classic statistical methods (case study: Oroumieh synoptic station) (Technical note). Iran Water Resources Research 13, 179-183. https://www.iwrr.ir/article_45176.html?lang=en

Wilby RL, Wigley TML., 1997. Downscaling general circulation model output: a review of methods and limitations. Progress in Physical Geography: Earth and Environment. 21:530-48.

Yeganeh-Bakhtiary, A., EyvazOghli, H., Shabakhty, N., Kamranzad, B., Abolfathi, S., 2022. Machine Learning as a Downscaling Approach for Prediction of Wind Characteristics under Future Climate Change Scenarios. Complexity 2022, 8451812. https://10.1155/2022/8451812

Yeganeh-Bakhtiary, A., EyvazOghli, H., Shabakhty, N., Abolfathi, S., 2023. Machine Learning Prediction of Wave Characteristics: Comparison between Semi-Empirical Approaches and DT Model. Ocean Engineering 286(2). https://doi.org/10.1016/j.oceaneng.2023.115583

Zhang XC. Spatial downscaling of global climate model output for site-specific assessment of crop production and soil erosion. Agricultural and Forest Meteorology. 2005; 135:215-29.

Zhang X, Shen M, Chen J, Homan JW, Busteed PR. Evaluation of statistical downscaling methods for simulating daily precipitation distribution, frequency, and temporal sequence. Transactions of the ASABE. 2021; 64:771-84.