COMPARATIVE ANALYSIS OF MULTIPLE LINEAR REGRESSION AND GENERALIZED REGRESSION NEURAL NETWORK FOR WATER TEMPERATURE ESTIMATION OF FONTAINE DES GAZELLES RESERVOIR DAM-BISKRA, ALGERIA

[1]ASSIA MEZIANI

[1] Faculty of Technology, Department of Hydraulic and Civil Engineering, University of El-Oued, PB789, 39000 El-Oued, Algeria.
Email: [1]assia-meziani@univ-eloued.dz

Abstract: Our research centers on estimating the water temperature of Fontaine de Gazelles Reservoir Dam by analyzing air temperature, relative humidity, solar radiation, atmospheric pressure, wind speed, and precipitation. These variables collectively impact water temperature, reflecting the thermal environment, water vapor content, solar energy, air density, wind-induced processes, and precipitation cooling. We employ Multiple Linear Regression (MLR) and Generalized Regression Neural Networks (GRNN) models for accurate estimates. MLR captures linear dependencies among climate variables, while GRNN model complex nonlinear relationships. Trained on historical data and real-time measurements, both MLR and GRNN demonstrate strong capabilities. MLR achieves high Nash-Sutcliffe Efficiency (0.991 to 0.997) and low Root Mean Squared Error (0.406 to 0.625), while GRNN achieves similar values. Both models consistently exceed a coefficient of determination R^2 equal 0.99, indicating a robust correlation, and display low Mean Absolute Error (0.236 to 0.391), affirming their accuracy. This attests to MLR and GRNN's reliability in estimating water temperature for the Fontaine de Gazelles Reservoir Dam.
Keywords: Water Temperature, Multiple Linear Regression, Generalized Regression Neural Network, Reservoir-Dam, Biskra-Algeria.

1. INTRODUCTION

Water temperature plays a vital role in governing the biological processes within aquatic ecosystems, open oceans, coastal waters, lakes, and rivers. Its direct influence on dissolved oxygen concentrations is significant, as temperature increases, the solubility of gases in water, including oxygen, decreases. Therefore, as the water temperature increases, the ability of oxygen to dissolve diminishes, resulting in reduced levels of dissolved oxygen within the water [1]. The thermal characteristics of riverine ecosystems are shaped by various factors, including hydrological conditions, regional attributes, climatic variables, and structural features of the surrounding region, catchment, and specific site [2]. While water temperature is routinely monitored at water sources and treatment facilities, limited attention is given to its monitoring within the drinking water distribution system, despite its well-known impact on physical, chemical, and microbial reactions that affect water quality. Soil temperature, influenced by urban heat island effects, serves as a crucial parameter contributing to variations in drinking water temperature [3]. The application of statistical models for simulating or predicting stream water temperature has become indispensable in managing water resources and aquatic habitats. Parametric models, such as linear and non-linear regression, are commonly employed for shorter time scales, while periodic models offer advantages in capturing seasonal variations. Non-parametric models, such as artificial neural networks, are better suited for analyzing complex relationships between water temperature and environmental variables [4]. Changes in water temperature have wide-ranging consequences, impacting gas solubility, the metabolic rate of aquatic flora and fauna, evaporation rates, and the formation of ice. These influences extend beyond local ecosystems, manifesting as regional and downstream impacts [5, 6] . In light of the global concern regarding the prediction of future water temperatures, modeling studies often rely on climate models. Process-based approaches, which consider the physical mechanisms driving heat exchange in local water bodies, present an opportunity to enhance our confidence in predicting future temperature trends [7]. To measure water temperature in a reservoir dam, several methods can be employed. One common approach is to use a temperature probe or sensor, which can be deployed at various depths within the water column. Another method using remote sensing technologies, such as thermal imaging or infrared cameras, to capture surface temperature variations. These techniques provide a non-contact way of assessing water temperature from a distance. Additionally, manual measurements can be taken using handheld thermometers at specific locations, such as meteorological stations, within the reservoir dam. These measurements can provide valuable data on water temperature at specific points in the reservoir.

In this paper, we conduct a comparative analysis between multiple linear regression (MLR) and generalized regression neural network (GRNN) model for the estimation of water temperature in Fontaine des Gazelles reservoir dam located in Biskra. We also consider the influence of climatic parameters on water temperature such as air temperature, solar radiation, wind speed, vapour pressure, precipitation, and evaporation, play a significant role in shaping water temperature dynamics in aquatic ecosystems. By comparing these two models, we aim to evaluate their performance in predicting water temperature and determine which model provides more accurate and reliable results for this specific reservoir dam. The Fontaine des Gazelles reservoir-dam is located at geographic coordinates 35°7′20″N and 5°35′0″E, situated in Biskra (**Figure 1**). The climate in Biskra is classified as semi-arid, the region has recorded average annual precipitation of approximately 126.98 mm. The highest average air temperature is recorded in the hottest month, July, at 35.18°C, whereas the lowest average temperature occurs in the coldest month, January, with an average air temperature of 11.86°C. Additionally, the average water temperature varies from 11.21°C in January to 30.38°C in July. The climate in the region is characterized as temperate with hot and dry summers. The winds are classified as weak throughout the year, with speeds ranging from 7-10 knots according to the Beaufort scale. Monthly relative humidity varies from 25.45% to 58.71%. The average monthly atmospheric pressure ranges from 97.15 to 97.94 kPa. The average annual monthly Solar radiation is 5.24 kWh/m^2/day [8].

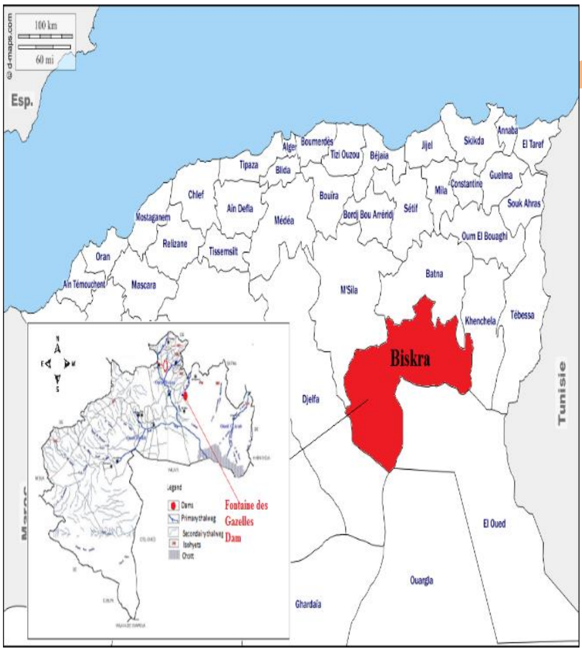

Fig.1. Location of Fontaine des Gazelles reservoir-dam-Biskra
Source : www.d-maps.com, [9]

2. MATERIALS AND METHODS

The utilization of MLR and GRNN models for estimating water temperature based on climate data brings numerous benefits. Firstly, MLR considers multiple climate variables like air temperature, solar radiation, wind speed, Humidity, evaporation and precipitation, to establish a linear relationship with water temperature. This enables us to understand how these climate factors collectively influence water temperature. On the other hand, GRNN models excel in capturing complex nonlinear relationships between climate variables and water temperature. By employing artificial neural networks, GRNN model can effectively analyze and predict water temperature in situations where the relationships are not simply linear. Together, these modeling approaches provide us with valuable water temperature predictions. The monthly statistical measured climatic data for 19 years (From 2000 to 2019) were obtained from Fontaine des Gazelles dam as given in **Table 1**. The climatic elements with the greatest impact on water temperature are air temperature, relative humidity, solar radiation, and atmospheric pressure. These factors significantly shape water temperature. A negative correlation implies that when relative humidity rises, water temperature tends to decrease, and vice versa. This suggests that higher air humidity generally leads to lower water temperatures. Conversely, wind speed and precipitation exert a comparatively milder influence on water temperature in comparison to the other variables.

Table 1: Monthly statistical measured parameters

Data	T_{air}	H_r	R_s	P_{atm}	V	P	T_w
Unit	°C	%	kWh/m^2 /day	kPa	m/s	mm	°C
X_{mean}	22.9	41.4	5.24	92.7	4.0	10.8	20.4
X_{max}	37.7	66.3	7.97	93.1	7.9	67.1	31.6
X_{min}	9.40	21.4	1.98	92.4	0.08	0.00	8.7
S_d	8.2	11.6	1.86	0.23	1.2	16.6	6.80
C_v	0.36	0.28	0.36	0.00	0.3	1.54	0.33
R^2	0.99	-0.83	0.84	0.99	-0.11	-0.14	1.00

Where:
T_{air}: air temperature; H_r: relative humidity; R_s: Solar radiation; P_{atm}: Atmospheric pressure; V: wind speed; P: precipitation; T_w: water temperature, X_{mean}: mean value, X_{max}: maximum value; X_{min}: minimum value; S_d: Standard deviation; C_v: variation coefficient, and R: correlation coefficient.
We will calculate statistical parameters to validate MLR and GRNN models for 70% training, 15% test and 15% validation by the following relations:

-Nash-Sutcliffe Efficiency (NSE):

$$NSE = 1 - \left[\frac{\sum_{i=1}^{n}\left(T_i^{Measured} - T_i^{Model}\right)^2}{\sum_{i=1}^{n}(T_i^{Meausured} - T_{Mean}^{Measurred})^2}\right] \quad (1)$$

- Root Mean Square Error (RMSE):

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(T_i^{Meausured} - T_i^{Model}\right)^2}{N}} \quad (2)$$

- Coefficient of determination:

$$R^2 = \frac{\sum_{i=1}^{n}(T_i^{Measured} - T_{Mean}^{Measured})(T_i^{Model} - T_{Mean}^{Model})}{\sum_{i=1}^{n}(T_i^{Measured} - T_{Mean}^{Measured})^2 \sum_{i=1}^{n}(T_i^{Model} - T_{Mean}^{Model})^2} \quad (3)$$

- Rank Sum Ratio (RSR):

$$RSR = \sqrt{\frac{\sum_{i=1}^{n}\left(T_i^{Meausured} - T_i^{Model}\right)^2}{\sum_{i=1}^{n}\left(T_i^{Meausured} - T_{Mean}^{Measured}\right)^2}} \quad (4)$$

- Willmott Index (WI)

$$WI = 1 - \left(\frac{\Sigma(T_i^{Model} - T_i^{Meausured})}{\Sigma(|T_i^{Model} - T_{Mean}^{Measured}|)}\right) \quad (5)$$

2. 1. MULTIPLE LINEAR REGRESSION AND GENERALIZED REGRESSION NEURAL NETWORK MODELS

Temperature affects dissolved oxygen concentrations, biochemical oxygen demand rates, algae production, and contaminant toxicity [10]. Several water temperature metrics have been identified as important indicators of the "thermal health" of a river and can be useful for the management and conservation of aquatic species [11]. Multiple linear regression (MLR) and Generalized Regression Neural Network (GRNN) are frequently employed models for predicting climate parameters. Multi linear regression (MLR) is a statistical approach utilized to establish the connection between numerous independent variables and a dependent variable. The basic model for multiple linear regression is:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n + \varepsilon \quad (6)$$

Where:

Y is the dependent variable X_1, X_2, ..., X_n are the independent variables. β_0 is the intercept. β_1, β_2, ..., β_n are the coefficients of the independent variables. ε is the error term.

Utilizing the given climatic parameters, the formula for Multiple Linear Regression (MLR) can be expressed as:

$$T_w = \beta_0 + \beta_1 H_r + \beta_2 T_{air} + \beta_3 P_{atm} + \beta_4 V + \beta_5 P + \beta_s R_s + \varepsilon \quad (7)$$

The MLR model aims to find the best-fitting coefficients (β values) that minimize the difference between the predicted water temperature and the measured water temperature based on the given climatic data. The choice of multiple linear regression was motivated by the potential for the accuracy of the regression model could be improved by selecting several independent variables [12, 13]. Furthermore, the results of simulation studies imply that MLR estimates exhibit some robustness, even when observed variable distributions depart from normality [14].

The Generalized Regression Neural Network (GRNN) architecture [15] subsumes the Radial-Basis Function (RBF) method. It approximates any arbitrary nonlinear function between input and output vectors, drawing the function estimate directly from training data [16, 17]. As the training set expands, the prediction error is reduced to zero [18]. It has the advantages of simple model structure, few parameters to adjust, strong approximation ability, fast learning speed, and high prediction accuracy, which has been widely used in many fields [19-21]. GRNN has a four-layer network structure including input, pattern, summation, and output layers [22-24].

The MLR and GRNN models applied to estimate water temperature by using monthly climatic data as relative humidity (H_r), air temperature (T_{air}), atmospheric pressure (P_{atm}), wind speed (V), precipitation (P) and solar radiation (R_s) from January 2000 to December 2019.

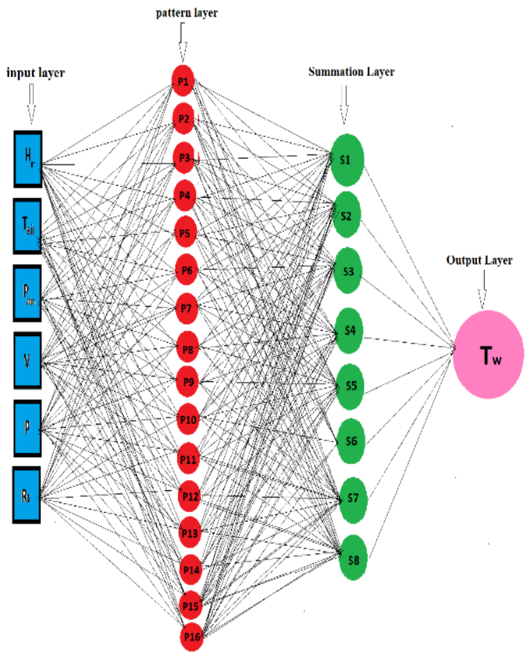

Fig.2. Architecture of GRNN model for modeling monthly water temperature.

As shown in **Figure 2**, The input layer consists of six neurons representing the data's input features, which correspond to available climate parameters. The hidden (pattern) layer is where the neural network learns and captures patterns present in the data. The summation layer, with height neurons, calculates the weighted sum of the inputs and generates activation values based on radial basis functions. The output layer generates the final model prediction, featuring a single neuron dedicated to predicting the water temperature.

3. RESULTS AND DISCUSSION

Both the GRNN and MLR models, trained with 70% of the data, showcase strong predictive abilities with an NSE of 0.991 and R^2 of 0.991(Figure 3. **(i & ii)**), indicating accurate variance explanation. The GRNN model has a slightly higher RMSE (0.630) but a lower MAE (0.387) compared to the MLR model's RMSE of 0.625 and MAE of 0.391. Both models exhibit similar performance in terms of RSR (0.031) and WI (0.998), Overall, the GRNN model displays a marginal advantage in terms of RMSE and MAE (Table 2). In the comparison of validation results at 15%, the GRNN model outperforms the MLR model in terms of R^2 (0.996 vs. 0.996) (Figure 5. **(i & ii)**), but has slightly higher RMSE (0.420 vs. 0.406) and MAE (0.251 vs. 0.236) values. Despite minor differences, both models demonstrate accurate predictive capabilities, with GRNN excelling in variance explanation while MLR shows slightly improved error metrics. The GRNN and MLR models on a 15% test dataset, both demonstrate strong predictive abilities. The MLR model exhibits slightly lower RMSE (0.414 vs. 0.439) and MAE (0.270 vs. 0.297), implying more accurate predictions. Both models excel in R^2 (0.997(Figure 7. **(i & ii)**)), with the MLR model having a marginally better performance. Both models capture residual patterns effectively (GRNN: RSR 0.021, MLR: RSR 0.019) and adhere to randomness assumptions (WI 0.999). Overall, while both models are proficient, the MLR model showcases slightly improved precision in this context. Positive standardized residuals indicate that the observed value is higher than expected, while negative standardized residuals indicate that the observed value is lower than expected (Figure 4. **(i & ii)**).

Table 2: Performances Metrics for MLR-GRNN models i) training ii) validation iii) testing

PERFORMANCE METRIC MLR			
	Training 70%	Validation 15%	Test 15%
NSE	0,991	0,996	0,997
RMSE	0,625	0,406	0,414
MAE	0,391	0,236	0,270
R^2	0,991	0,996	0,997
RSR	0,031	0,020	0,019
WI	0,998	0,999	0,999
PERFORMANCE METRIC GRNN			
	Training 70%	Validation 15%	Test 15%
NSE	0,991	0,996	0,996
RMSE	0,630	0,420	0,439
MAE	0,387	0,251	0,297
R^2	0,991	0,996	0,996
RSR	0,031	0,021	0,021
WI	0,998	0,999	0,999

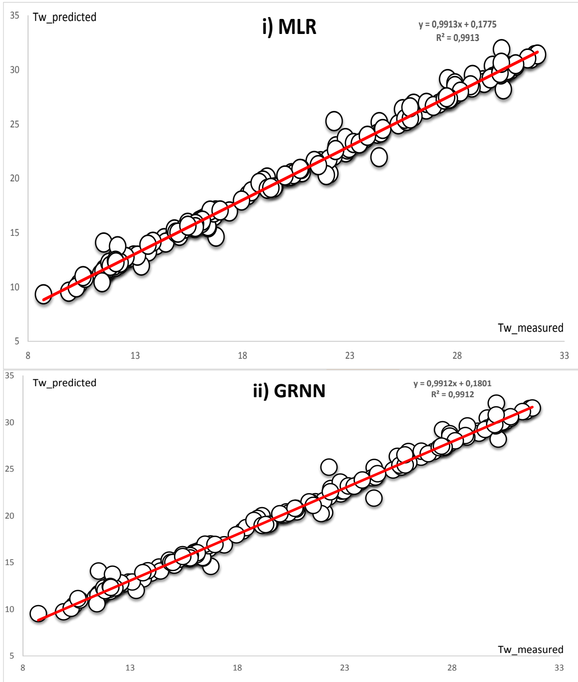

Fig.3. (i & ii) Scatter plots of 70% training for MLR and GRNN models.

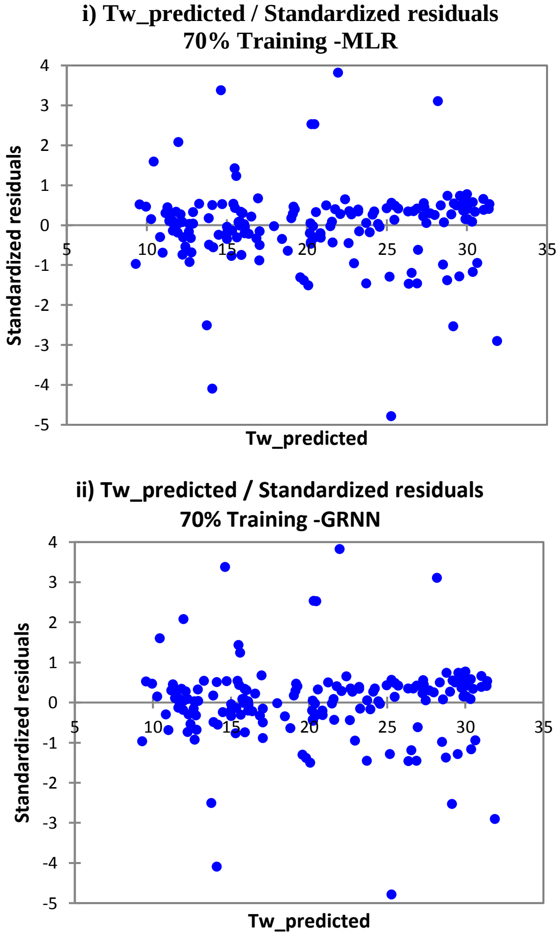

Fig.4. (i & ii) Standardized residuals ($T_{w\_predicted}$) of 70% training for MLR and GRNN models.

We mention that the majority of standard residuals are close to zero and follow a normal distribution, it indicates that the models predict water temperature effectively.

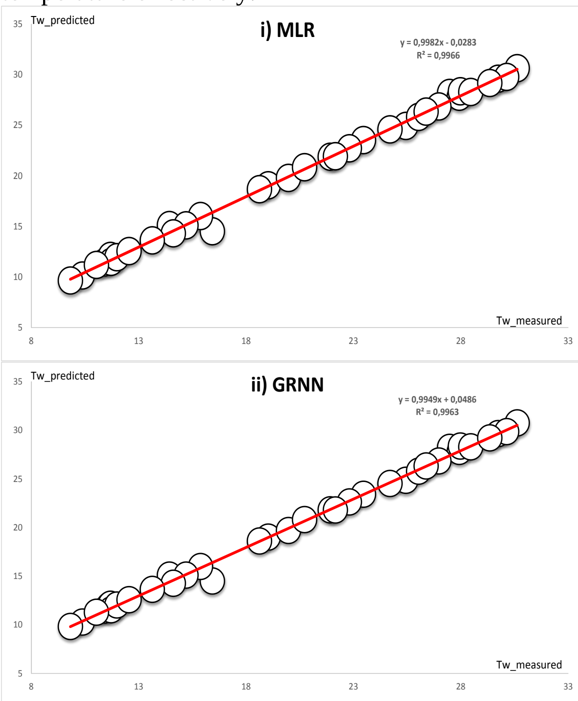

Fig. 5. (i & ii) Scatter plots of 15% validation for MLR and GRNN models.

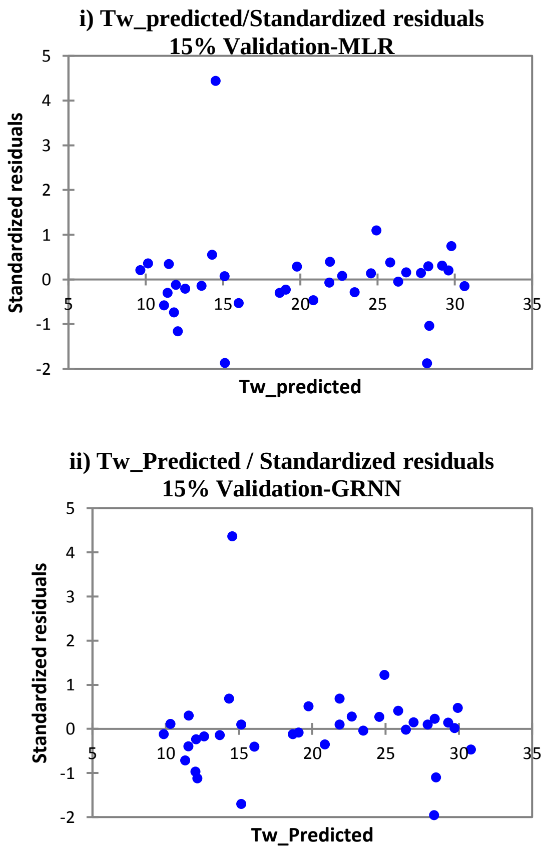

Fig. 6. (i & ii) Standardized residuals ($T_{w\_predicted}$) of 15% validation for MLR and GRNN models

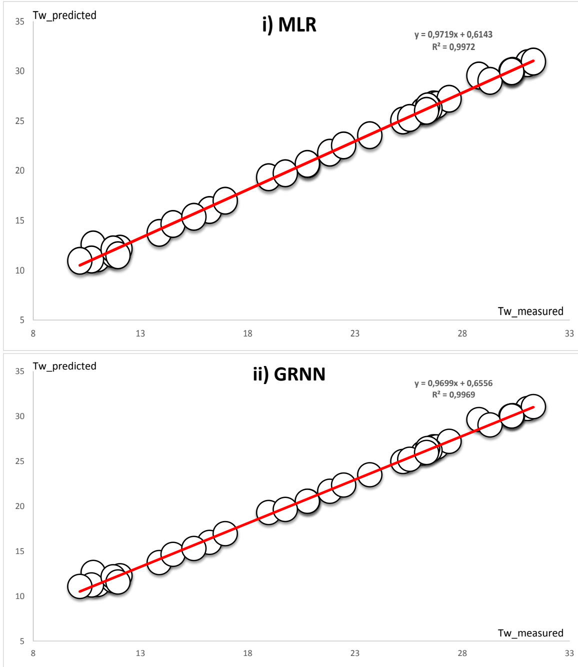

Fig. 7. (i & ii) Scatter plots of 15% test for MLR and GRNN models.

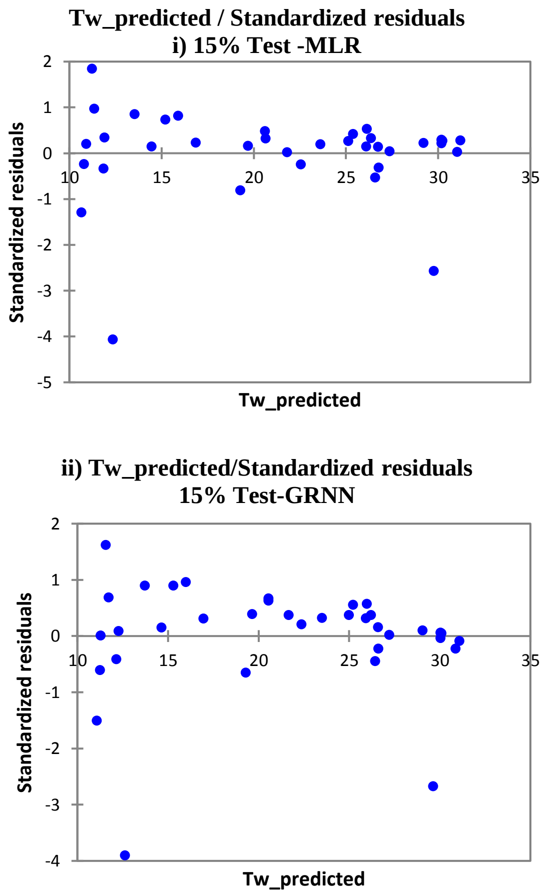

Fig. 8. (i & ii) Standardized residuals ($T_{w\_predicted}$) of 15% test for MLR and GRNN models

In case of a standardized residual is greater than 2 (in absolute value), it means that the predicted value deviates significantly from what would be expected based on the model. We observe several cases for the MLR and GRNN models **Figure. 4. (i & ii), Figure. 6. (i & ii) and Figure. 8. (i & ii)**. But if the standardized residual is less than 1, it indicates that the observed value is closer to the predicted value, and there isn't a significant deviation between the two.

4. CONCLUSIONS

After examining the performance metrics of the Multiple Linear Regression (MLR) and Generalized Regression Neural Network (GRNN) models, serval important observations come to light:

1. Both MLR and GRNN models exhibit excellent predictive capabilities across the training, validation, and testing sets.
2. The MLR and GRNN models demonstrate outstanding predictive capabilities across various performance metrics.
3. The consistently high-performance metrics across the training, validation, and testing datasets underscore their proficiency in estimating water temperature based on the given climatic parameters.
4. The models' accuracy and robustness, as revealed by the diverse metrics, affirm their suitability for applications in water temperature prediction

ACKNOWLEDGMENTS

We extend our sincere gratitude to the engineers of Agence Nationale de Barrages et Transfers (ANBT) for graciously providing us with the data and essential climatic parameters. Their contributions have been truly priceless, propelling our research forward and enabling us to make a tangible difference. We also wish to express our heartfelt appreciation to the dedicated team at Fontaine des Gazelles for their collaborative effort in sharing the indispensable data and technical insights that underpinned our study.

REFERENCES

1. Pivato, M., et al., *Remote Sensing for Optimal Estimation of Water Temperature Dynamics in Shallow Tidal Environments.* Remote Sensing, **12**(1): p. 51, 2020.
2. Dallas, H., *Water temperature and riverine ecosystems: An overview of knowledge and approaches for assessing biotic responses, with special reference to South Africa.* Water SA. **34**: p. 393-404, 2008.
3. Agudelo-Vera, C., et al., *Drinking Water Temperature around the Globe: Understanding, Policies, Challenges and Opportunities.* Water. **12**(4): p. 1049, 2020.
4. Benyahya, L., et al., *A Review of Statistical Water Temperature Models.* Canadian Water Resources Journal / Revue canadienne des ressources hydriques. **32**(3): p. 179-192, 2007.
5. Olden, J.D. and R.J. Naiman, *Incorporating thermal regimes into environmental flows assessments: modifying dam operations to restore freshwater ecosystem integrity.* Freshwater Biology. **55**(1): p. 86-107, 2010.
6. Wanders, N., et al., *High-Resolution Global Water Temperature Modeling.* Water Resources Research, **55**(4): p. 2760-2778, 2019.
7. Miller, R.L., *Modeling response of water temperature to channelization in a coastal river network.* River Research and Applications. **37**(3): p. 433-447, 2021.
8. Meziani, A., *Etude de l'evaporation des eaux de barrages reservoirs en regions arides.* Université de Ouargla. p. 335, 2021.
9. Sedrati, N., *Origines et caractéristiques physic-chimiques des eaux de la wilaya de Biskra*, *Faculté des sciences de la Terre*, *Département de Géologie*. Badji mokhtar-annaba university. p. 18, 2011.
10. John C. Risley, U.S.G.S., Edwin A. Roehl, Jr., Advanced Data Mining, LLC, and U.S.G.S. Paul A. Conrads, *Estimating Water Temperatures in Small Streams in Western Oregon Using Neural Network Models*, in *Water-Resources Investigations*. p 1. Portland, Oregon, 2003.
11. Ouarda, T.B.M.J., C. Charron, and A. St-Hilaire, *Regional estimation of river water temperature at ungauged locations.* Journal of Hydrology X, **17**: p. 100133, 2022.
12. Kim, S.W., D. Jung, and Y.-J. Choung *Development of a Multiple Linear Regression Model for Meteorological Drought Index Estimation Based on Landsat Satellite Imagery*. Water. **12**, , 2020. DOI: 10.3390/w12123393.
13. Yun, H.; Um, M.; Cho, W.; Heo, J.H. *Orographic Orographic Precipitation Analysis with Regional Frequency Analysis and Multiple Linear Regression*. J. Korea Water Resour. Assoc. **42**, 465–480.2009.
14. Li, C.-H., *Statistical estimation of structural equation models with a mixture of continuous and categorical observed variables.* Behavior Research Methods. **53**(5): p. 2191-2213, 2021.
15. Specht, D.F., *A general regression neural network.* IEEE Transactions on Neural Networks. **2**(6): p. 568-576, 1991.
16. Kişi, Ö., *Generalized regression neural networks for evapotranspiration modelling.* Hydrological Sciences Journal. **51**(6): p. 1092-1105, 2006.
17. Kumar, D. and R.K. Bhattacharjya, *GRNN Model for prediction of groundwater fluctuation in the state of Uttarakhand of India using GRACE data under limited bore well data.* Journal of Hydroinformatics. **23**(3): p. 567-588, 2021.

18. Üneş, F., et al., *Modeling of dam reservoir volume using generalized regression neural network, support vector machines and m5 decision tree models.* Applied ecology and environmental research. **17**((3)): p. 7043-7055, 2019.
19. Feng, Y., et al., *Evaluation of random forests and generalized regression neural networks for daily reference evapotranspiration modelling.* Agricultural Water Management. **193**: p. 163-173, 2017.
20. Li, J., L. Wang, X. Wang, and P. Gao, *Chaos-generalized regression neural network prediction model of mine water inflow.* SN Applied Sciences **3**(12): p. 861, 2021.
21. Xie, Y., C. Li, Y. Lv, and C. Yu, *Predicting lightning outages of transmission lines using generalized regression neural network.* Applied Soft Computing. **78**: p. 438-446, 2019.
22. Cai, Q.-C., T.-H. Hsu, and J.-Y. Lin *Using the General Regression Neural Network Method to Calibrate the Parameters of a Sub-Catchment*. Water. **13**, , 2021. DOI: 10.3390/w13081089.
23. Ghritlahre, H.K. and R.K. Prasad, *Exergetic performance prediction of solar air heater using MLP, GRNN and RBF models of artificial neural network technique.* Journal of Environmental Management. **223**: p. 566-575, 2018.
24. Izonin, I., R. Tkachenko, V. Verhun, and K. Zub, *An approach towards missing data management using improved GRNN-SGTM ensemble method.* Engineering Science and Technology, an International Journal. **24**(3): p. 749-759, 2021.