ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume X-4/W8-2025
The 8th ISPRS Geospatial Conference 2025, 15–17 December 2025, Tehran, Iran

A Comparative Study of OLS and MLP Models for Spatial Downscaling of Air Temperature

Hedieh Zahra Zarkesh, Mohammad Karimi *, Tahereh Ghaemi Rad

GIS Department, Faculty of Geomatics Engineering, Centre of Excellence in Geospatial Information Technology,
K.N. Toosi University of Technology, Tehran, Iran – hediehzz@email.kntu.ac.ir – (mkarimi, t.ghaemirad)@kntu.ac.ir

Keywords: Monthly Temperature Estimation, Spatial Downscaling, Ordinary Least Squares (OLS), Multi-Layer Perceptron (MLP), Remote Sensing, Normalized Difference Vegetation Index (NDVI).

Abstract

High-resolution and accurate air temperature data are critical for climate studies and environmental planning. This study evaluates and compares two spatial downscaling approaches for estimating monthly air temperature in Iran: Ordinary Least Squares (OLS) for its interpretability, and Multilayer Perceptron (MLP) for its predictive power. Both models were trained on data from synoptic stations using predictors such as geographical location, elevation, precipitation, and Normalized Difference Vegetation Index (NDVI). The trained models were then used to generate temperature grids at 5-km spatial resolution for four representative months. The results demonstrate that the MLP model consistently outperforms the OLS model, achieving a higher coefficient of determination (R^2 = 0.946 vs. 0.899 in May) and a significantly lower Root Mean Squared Error (RMSE). Furthermore, visual analysis of the generated maps reveals the MLP's superior ability to capture complex local phenomena, such as the moderating effect of the Caspian Sea and sharp temperature gradients in mountainous regions. Overall, results show that while OLS serves as a simple global linear baseline, non-linear models such as MLP are essential for achieving higher accuracy and capturing geographically realistic temperature patterns.

1. Introduction

Monthly temperature, a key variable in climate and environmental studies, has been widely modeled to support climate change assessment, urban planning, and resource management, particularly in data-scarce regions. The literature indicates that geographic and environmental variables—specifically latitude, longitude, elevation, annual precipitation, and Normalized Difference Vegetation Index (NDVI)—demonstrate strong correlations with temperature (Hopkins, 1968; Khalid et al., 2024; Şevgin and Öztürk, 2024; Zhao et al., 2016).

Limited ground-based observations necessitate high-resolution temperature estimation. To address this, various downscaling techniques have been developed to convert coarse-scale climate data (e.g., Global Climate Model (GCM) outputs) into fine-scale estimates. Downscaling approaches are classified as dynamical, which rely on computationally expensive regional climate models (Soares et al., 2024), and statistical, which empirically relate large-scale predictors to local observations (Ozbuldu and Irvem, 2025; Wang et al., 2019; Groenke et al., 2020).

Among traditional statistical approaches, Ordinary Least Squares (OLS) regression remains widely used due to its simplicity. Studies have shown that OLS can perform competitively with time-series and machine learning models in temperature estimation (Liu et al., 2023; Hadad, 2022; Şan et al., 2023), though its global linear assumption limits performance in heterogeneous landscapes.

Geographically Weighted Regression (GWR) addresses the spatial homogeneity limitation of OLS by allowing parameter variation across space, leading to significant improvements in studies such as those by Khalid et al. (2024) and Siddique et al. (2023).

With the advent of Machine Learning (ML), more advanced statistical downscaling methods have emerged. Algorithms such as Random Forest (RF), Support Vector Regression (SVR), and Artificial Neural Networks (ANN) can effectively capture non-linear relationships between predictors and temperature (Ozbuldu and Irvem, 2025; Zuo et al., 2019). Among ANN architectures, the Multilayer Perceptron (MLP) has been widely recognized for its flexibility and strong predictive performance in environmental modeling. For instance, Zhou et al. (2024) applied an MLP optimized with the Levenberg–Marquardt algorithm to predict air temperature and humidity, achieving an coefficient of determination (R^2) of 0.997 and Root Mean Squared Error (RMSE) of 0.439 °C, highlighting its capability to learn complex, non-linear dependencies between climatic variables. Such models can generalize well even with limited predictors, making them suitable for spatial downscaling tasks. Recent studies have also employed deep learning frameworks such as Convolutional Neural Networks (CNN) and hybrid models for Coupled Model Intercomparison Project Phase 6 (CMIP6) downscaling, achieving high spatial accuracy (Soares et al., 2024; Kumar et al., 2025; Xu et al., 2020).

Despite their accuracy, ML and deep learning models require dense and spatially uniform training data, which are rarely available in complex or mountainous regions. Moreover, recent studies emphasize the importance of input variables such as NDVI, Digital Elevation Model (DEM), and precipitation in improving downscaling accuracy (Wang et al., 2019).

This study introduces a nationwide 5-km grid-based framework for monthly temperature downscaling in Iran, representing a distinct contribution compared to previous regional-scale studies that primarily relied on local or station-based models. It

* Corresponding author

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-4-W8-2025-855-2026 | © Author(s) 2026. CC BY 4.0 License.

leverages key environmental predictors, including geographical location, elevation, precipitation, and NDVI, extracted using Google Earth Engine (GEE), Python, and QGIS. The framework integrates two complementary modeling paradigms, a global linear model (OLS) and a non-linear neural network (MLP), to evaluate the trade-off between simplicity and predictive accuracy under consistent spatial conditions. The performance of both models is rigorously evaluated using statistical metrics and visual analysis of the resulting temperature surfaces. The paper is structured as follows: Section 2 introduces the study area and outlines the research methodology, Section 3 describes the modeling results, while the discussion and conclusion are presented in the final sections.

2. Methodology

2.1 Case study area

The study area of this research encompasses the entire territory of Iran, covering approximately 1,648,000 square kilometres, located between 25° and 40° N latitude and 44° to 63° E longitude. Due to its considerable topographic and climatic diversity, Iran offers a suitable context for exploring spatial relationships between temperature and bioclimatic variables. The elevation range extends from below sea level to over 5,600 meters, and the country includes various climate zones, such as arid, semi-arid, and humid regions, which provide a robust basis for spatial temperature analysis. The geographic location of the spatial distribution of synoptic stations is illustrated in figure 1.

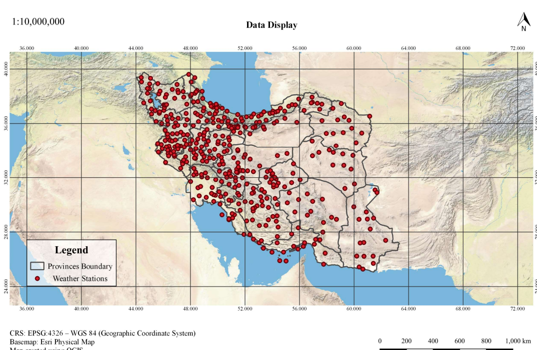


Figure 1. Spatial distribution of synoptic stations

2.2 Data gathering and modeling

Monthly temperature and precipitation data were obtained from 417 synoptic stations distributed across Iran for May, August, and November 2012, and February 2013 (Figure 1, a). This selection corresponds to the four mid-season months of the Iranian calendar year 1391, which was the target year for model development in this study.

A regular grid of artificial points spaced 5-km apart was generated using Python libraries such as numpy and geopandas. A 5-km spacing was selected based on the probability density function of the observed distances between pairs of synoptic stations. (Figure 2) with the aim of increasing the sampling point density and, consequently, improving the accuracy of the interpolation process.

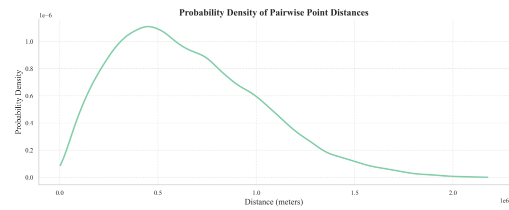


Figure 2. The probability density function of the observed distances between pairs of synoptic stations

The descriptive variables including precipitation, elevation, and NDVI were extracted using the Google Earth Engine (GEE) platform (Table 1).

Variable	Source	Spatial Resolution	Time Period Used
Elevation	SRTM (NASA)	30 m	Static (year 2000)
NDVI	Landsat 5 (NASA/USGS)	30 m	Year 2012
Monthly Precipitation	CHIRPS Daily (UCSB/CHG)	~5 km	May, August, November 2012 and February 2013

Table 1. The descriptive variables extracted from GEE.

To justify the selection of input predictors, an Exploratory Regression Analysis was performed in QGIS. This process examined the statistical relationships between temperature and multiple descriptive variables to identify the combination that best explained spatial temperature variability while minimizing multicollinearity. The variables of latitude, longitude, elevation, precipitation, and NDVI demonstrated the highest explanatory power and were therefore selected for both OLS and MLP modeling to ensure consistency and comparability between the two approaches.

A geographically stratified sampling approach was used to split 417 synoptic stations into training (80%, 334 points) and test (20%, 83 points) sets, ensuring uniform spatial coverage. A comparative modeling framework was then implemented to evaluate the performance of two approaches: OLS, serving as an interpretable global linear baseline, and MLP, designed to capture non-linear spatial patterns. The MLP architecture included two hidden layers (100 and 50 neurons), with ReLU activation and the Adam optimizer.

GWR was excluded due to its local coefficient structure, which limits its applicability for generating continuous prediction surfaces on a regular grid. In contrast, OLS and MLP were selected for their suitability in grid-based estimation and nationwide surface generation. Although other advanced approaches such as Geographically and Temporally Weighted Regression (GTWR) and unsupervised deep learning frameworks can model spatial heterogeneity, they were not included in this study because the main focus was on developing and evaluating a general framework for large-scale air temperature estimation. Compared with other methods or algorithms, the combination of OLS and MLP provides a balanced trade-off between computational simplicity, predictive accuracy, and spatial generalizability across Iran. Future studies may extend this framework by incorporating GTWR or unsupervised learning models for broader comparative analysis.

Model performance was rigorously assessed using the held-out test set, applying statistical metrics including R², Mean Absolute Error (MAE), and RMSE. Finally, both trained models were applied to a 5-km resolution artificial grid,

https://doi.org/10.5194/isprs-annals-X-4-W8-2025-855-2026

producing monthly high-resolution temperature surfaces that enabled quantitative and visual comparison of the linear and non-linear approaches.

3. Results and Discussion

The regression coefficients, presented in Table 2, were estimated using the OLS method based on the training dataset.

Month	Intercept	Lat	Lon	Elev	Percip	NDVI
February	49.098	-0.92	-0.03	-0.005	0.016	-0.894
May	49.449	-0.62	0.055	-0.006	-0.034	-7.956
August	63.4	-0.46	-0.23	-0.005	-0.034	-6.167
November	84.419	-1.14	-0.43	-0.005	-0.046	4.992

Table 2. Regression coefficients estimated using OLS method.

The OLS coefficient analysis provides an interpretable baseline for understanding the fundamental relationships between the predictors and temperature. As expected, latitude and elevation consistently show significant negative effects on temperature. However, the variable seasonal influence of predictors like NDVI and precipitation suggests that the underlying relationships are not purely linear. To test if a non-linear model could better capture these complexities and provide higher predictive accuracy, an MLP model was also developed for a direct comparison. The performance of both the OLS and MLP models was rigorously evaluated and compared using the held-out test set. The primary evaluation metrics included the R^2, MAE, and RMSE to assess the predictive accuracy of each model. The results, summarized in Table 3 demonstrate the MLP model's consistent outperformance of the OLS model across all four months.

Month	OLS			MLP		
	R^2 Score	RMSE	MAE	R^2 Score	RMSE	MAE
February	0.706	3.069	1.903	0.722	2.986	2.247
May	0.899	1.99	1.62	0.946	1.458	1.144
August	0.788	2.3175	1.9135	0.841	2.014	1.564
November	0.744	3.279	2.757	0.854	2.474	1.994

Table 3. Comparative Performance Evaluation of OLS and MLP Models

Month	P-value	Significance (α = 0.05)
February	0.0099	Significant
May	< 0.0001	Significant
August	0.023	Significant
November	0.0225	Significant

Table 4. Statistical Significance of Performance Difference between OLS and MLP

The results in Table 3 clearly demonstrate the superior predictive accuracy of the MLP model, which achieves higher R^2 and lower RMSE/MAE values across all months compared to OLS. To assess whether these performance differences were statistically significant, a Wilcoxon signed-rank test was conducted on the absolute errors. The results, presented in Table 4, confirm that the MLP's improved accuracy is statistically significant ($p < 0.05$) for all evaluated months.

To visually confirm these findings and further analyze the models' behavior, Figure 3 presents the scatter plots of observed versus predicted temperatures for both models. In these plots, the MLP predictions are visibly more concentrated around the 1:1 line of perfect agreement compared to the OLS predictions, indicating less scatter and a better overall fit. Notably, while both models exhibit weaker performance in February, suggesting greater complexity in modeling winter temperatures, the MLP model still maintains a statistically significant higher accuracy.

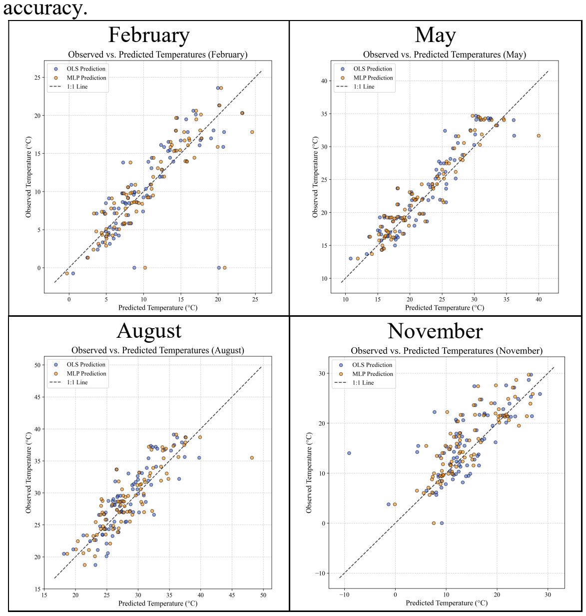


Figure 3. Scatter plots showing the observed versus predicted temperatures using the OLS and MLP methods

To visually assess and compare the spatial performance of the models, the predicted temperature values on the 5-km grid were interpolated using Kriging to generate continuous surfaces. Figure 4 presents a comparative visualization of these final interpolated surfaces for OLS and MLP for two representative months, May and November, including a difference map (MLP - OLS) to highlight areas of disagreement.

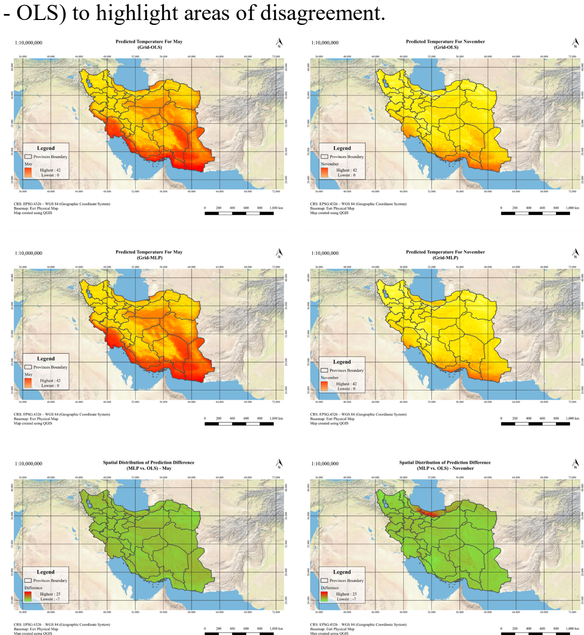


Figure 4. Predicted Temperature Surfaces for OLS and MLP Models, and Their Difference, for May and November

The visual analysis reveals the MLP's superior ability to capture complex, local geographic patterns. The MLP model produces a more detailed and realistic surface, particularly in defining the cooling effect of the mountain ranges. Most notably, in November, the difference map highlights the MLP's capacity to model the moderating effect of the Caspian Sea, predicting significantly warmer temperatures along the coastline a complex local phenomenon that the global OLS model fails to capture.
To provide a more detailed statistical insight into the error distributions, Kernel Density Estimate (KDE) plots of the absolute errors were generated for each month, as shown in Figure 5.

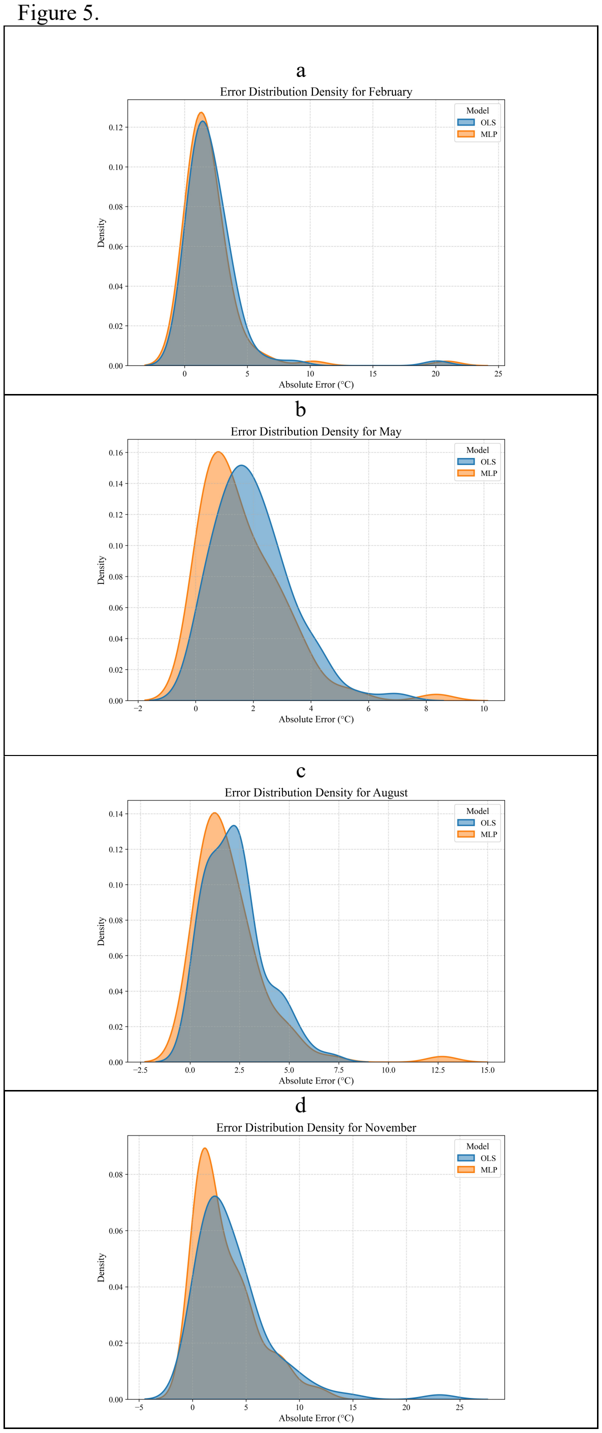
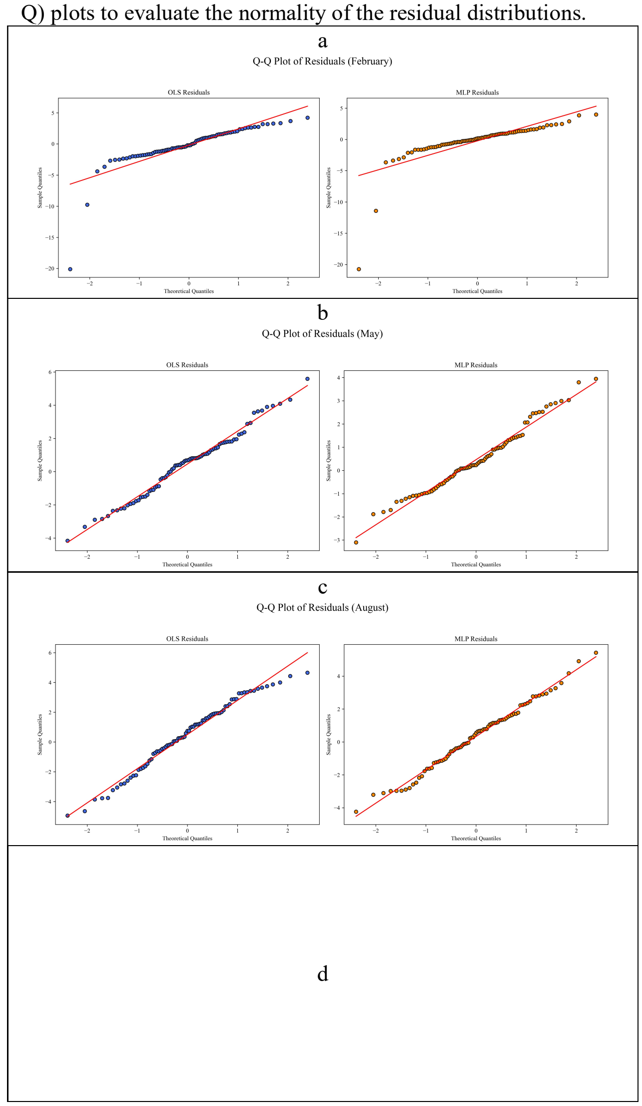


Figure 5. Comparative density of absolute prediction errors for OLS and MLP models across a) May, b) August, c) November, and d) February.

The analysis of the error distribution densities (Figure 5) indicates that the MLP model's performance is superior in most months. For May, August, and November, the MLP's error curve is taller, narrower, and its peak is shifted towards lower error values compared to the OLS curve. This demonstrates that the MLP consistently produces smaller and less dispersed errors. The longer tail in the OLS distribution, particularly in August, suggests a higher frequency of large errors. In February, the performance of the two models is nearly identical, with only a slight advantage for the MLP in the density of smaller errors. Overall, the KDE plots confirm that the MLP model offers greater stability and accuracy in temperature prediction.
To further investigate the statistical behavior of the errors and assess model fit quality, diagnostic plots of the residuals (observed minus predicted values) were examined. Figure 7 presents the residuals plotted against the predicted values for both models for May and November, allowing for an assessment of error randomness and homoscedasticity (constant variance). Additionally, Figure 6 displays Quantile-Quantile (Q-Q) plots to evaluate the normality of the residual distributions.



https://doi.org/10.5194/isprs-annals-X-4-W8-2025-855-2026

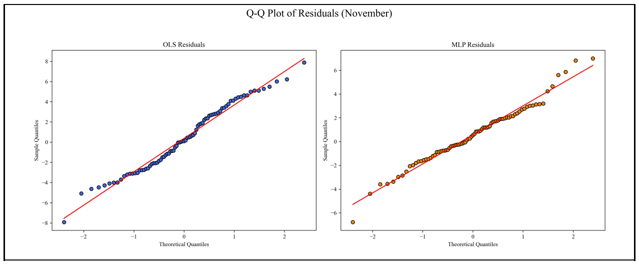


Figure 6. Quantile-Quantile (Q-Q) Plots comparing residual distributions to a theoretical normal distribution for OLS and MLP models across a) February, b) May, c) August, and d) November.

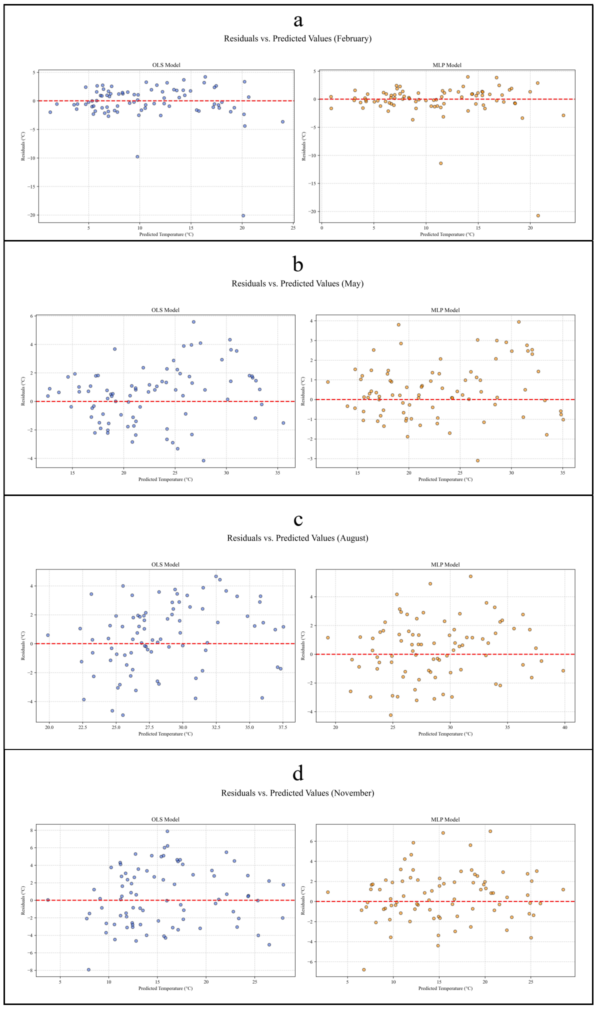


Figure 7. Residuals versus Predicted Temperature Values for OLS and MLP models across a) February, b) May, c) August, and d) November.

Analysis of the diagnostic plots (Figures 6 and 7) provides further evidence of the MLP model's robustness. The Residuals vs. Predicted plots (Figure 7) show that the MLP residuals generally exhibit a more random scatter around the zero line compared to the OLS residuals, which occasionally show subtle patterns or clustering, particularly potential heteroscedasticity (non-constant variance, e.g., funnel shape in November OLS). This suggests the MLP model captures the underlying data structure more effectively.

Furthermore, the Q-Q plots (Figure 6) indicate that the distribution of MLP residuals aligns more closely with the theoretical normal distribution line compared to the OLS residuals. The OLS residuals consistently show greater deviations from normality, especially in the tails. While neither model perfectly meets all classical assumptions, the MLP demonstrates residuals with more desirable statistical properties (greater randomness, better homoscedasticity, closer to normality), reinforcing its superior fit to the complex temperature data.

4. Conclusion

This study conducted a comparative evaluation of a linear (OLS) and a non-linear (MLP) model for the spatial downscaling of monthly air temperature in Iran. The findings affirmed established climatic patterns within the region. The OLS model, serving as an interpretable baseline, explicitly quantified the fundamental negative relationships between temperature and variables such as latitude and elevation through its regression coefficients.

The non-linear MLP model consistently achieved higher predictive accuracy than the OLS model across all evaluated months, with higher R² scores and lower RMSE and MAE values. Both models, however, showed relatively lower performance in February, which can be attributed to the increased spatial variability and stronger climatic contrasts of the winter season. These conditions—such as temperature inversions, snow cover, and reduced solar radiation—introduce additional complexity to temperature estimation. Despite this, the MLP model maintained a statistically significant improvement over OLS, reflecting its superior ability to learn non-linear spatial relationships even under challenging winter conditions. Moreover, visual analysis of the downscaled maps confirmed the MLP's advantage in capturing complex geographical patterns, including sharp temperature gradients in mountainous areas and the moderating influence of the Caspian Sea.

In conclusion, this research highlights the critical trade-off between model interpretability and predictive accuracy in environmental modeling. While OLS provides a simple and understandable baseline, the results prove that non-linear models like MLP are essential for developing high-fidelity, geographically realistic temperature surfaces. The spatial downscaling approach, powered by a robust machine learning model, is thereby validated as a powerful strategy for enhancing the spatial detail and utility of climate data in regions with complex topography and sparse observational networks.

Although this study focused on two representative and computationally efficient approaches (OLS and MLP), future research could extend this framework by incorporating additional machine learning or hybrid models to enable a more comprehensive quantitative comparison.

This study provides a novel national-scale demonstration of grid-based temperature downscaling over Iran, showing that a simple interpretable baseline (OLS) and a non-linear neural model (MLP) can be effectively integrated for generating accurate, high-resolution climatic datasets.

References

Groenke, B., Madaus, L., Monteleoni, C., 2020: ClimAlign: Unsupervised statistical downscaling of climate variables via normalizing flows. Proceedings of the 10th International Conference on Climate Informatics. https://doi.org/10.1145/3429309.3429318

Hadad, A., 2022: Downscaling climatic variables at a river basin scale: Statistical validation and ensemble projection under climate change scenarios. Doctoral dissertation.

Hopkins, J.W., 1968: Correlation of air temperature normals for the Canadian Great Plains with latitude, longitude, and altitude. Canadian Journal of Earth Sciences, 5(2), 199–210.

Khalid, W., Shamim, S.K., Ahmad, A., 2024: Exploring urban land surface temperature with geospatial and regression modelling techniques in Uttarakhand using SVM, OLS and GWR models. Evolving Earth, 2, 100038. https://doi.org/10.1016/j.eve.2024.100038

Kumar, B., Yadav, B.K., Mukhopadhyay, S., Rohan, R., Singh, B.B., Chattopadhyay, R., Chilukoti, N., Sahai, A.K., 2025: Towards location-specific precipitation projections using deep neural networks. arXiv preprint, arXiv:2503.14095. https://doi.org/10.48550/arXiv.2503.14095

Liu, K., Zheng, B., Ding, X., 2023: Comparative model study of global temperature forecasting based on the ARIMA–OLS model. Highlights in Science, Engineering and Technology. https://www.semanticscholar.org/paper/607bffdd237100cb1c9fa7be98c89f30b0c9ec41

Ozbuldu, M., Irvem, A., 2025: Projecting and downscaling future temperature and precipitation based on CMIP6 models using machine learning in Hatay Province, Türkiye. Pure and Applied Geophysics, 182(4), 1825–1842. https://doi.org/10.1007/s00024-024-03656-0

Şan, M., Nacar, S., Kankal, M., Bayram, A., 2023: Daily precipitation performances of regression-based statistical downscaling models in a basin with mountain and semi-arid climates. Stochastic Environmental Research and Risk Assessment, 37(4), 1431–1455.

Şevgin, F., Öztürk, A., 2024: Variation of temperature increase rate in the Northern Hemisphere according to latitude, longitude and altitude: The Turkey example. Scientific Reports, 14(1). https://doi.org/10.1038/s41598-024-68164-6

Shen, Z., Shi, C., Shen, R., Tie, R., Ge, L., 2023: Spatial downscaling of near-surface air temperature based on deep learning cross-attention mechanism. Remote Sensing, 15(21), 5084.

Siddique, M.A., Boqing, F., Dongyun, L., 2023: Modeling the impact and risk assessment of urbanization on urban heat island and thermal comfort level of Beijing City, China (2005–2020). Sustainability, 15(7), 6043. https://doi.org/10.3390/su15076043

Soares, P.M.M., Johannsen, F., Lima, D.C.A., Lemos, G., Bento, V.A., Bushenkova, A., 2024: High-resolution downscaling of CMIP6 Earth system and global climate models using deep learning for Iberia. Geoscientific Model Development, 17(1), 229–259. https://doi.org/10.5194/gmd-17-229-2024

Wang, L., Chen, R., Han, C., Yang, Y., Liu, J., Liu, Z., Wang, X., Liu, G., Guo, S., 2019: An improved spatial–temporal downscaling method for TRMM precipitation datasets in alpine regions: A case study in northwestern China's Qilian Mountains. Remote Sensing, 11(7), 870. https://doi.org/10.3390/rs11070870

Xu, J., Zhang, F., Jiang, H., Hu, H., Zhong, K., Jing, W., Yang, J., Jia, B., 2020: Downscaling Aster Land Surface Temperature over Urban Areas with Machine Learning-Based Area-To-Point Regression Kriging. Remote Sensing, 12(7), 1082. https://doi.org/10.3390/rs12071082

Zhao, N., Yue, T., Zhou, X., 2016: Statistical downscaling of precipitation using local regression and high accuracy surface modeling method. Theoretical and Applied Climatology, 129(1–2), 281–292. https://doi.org/10.1007/s00704-016-1776-z

Zhou, L., Wang, H., Li, Y., Zhang, C., Sun, F., 2024: Prediction of air temperature and humidity in greenhouses via artificial neural network. PLOS ONE, 19(6), e0325650. https://doi.org/10.1371/journal.pone.0325650

Zuo, J., Xu, J., Chen, Y., Wang, C., 2019: Downscaling precipitation in the data-scarce inland river basin of Northwest China based on Earth system data products. Atmosphere, 10(10), 613. https://doi.org/10.3390/atmos10100613