Numerical Analysis Final Project

Kasra Fouladi
402442297

June 12, 2025

Abstract

This paper presents a comprehensive comparison of several classical numerical interpolation techniques applied to the analysis of global mean temperature anomaly data. The concept of temperature anomaly is introduced, and the problem is formulated in the context of climate data analysis. We implement and evaluate local Newton forward and backward interpolation, local Lagrange interpolation, and local polynomial regression, using real-world climate datasets. The accuracy and characteristics of each method are discussed, providing insights into their suitability for scientific data analysis.

1 Introduction

Accurate analysis of near-surface air temperature is crucial for understanding regional climate patterns and their impacts, especially in countries with diverse climates such as Iran [6]. In this study, we focus on the 2-meter air temperature at 10 AM during the summer months (June, July, August) across Iran. This specific metric is important for assessing heat exposure and its effects on human health, agriculture, and energy demand.

Our data source is the ERA5-Land reanalysis dataset [4], which provides gridded, hourly temperature values for the region of interest. We extract the relevant data for the summer months and the 10 AM time step for each year in the study period.

The main goal of this project is to interpolate and analyze the spatial and temporal patterns of summer daytime temperatures in Iran. We compare several classical numerical interpolation methods, including local Newton forward and backward interpolation, local Lagrange interpolation, and local polynomial regression [1–3, 5, 7–10]. Each method is evaluated in terms of accuracy and suitability for regional climate data analysis.

A distinctive aspect of our approach is the use of local, pointwise interpolation: for each prediction year and location, the interpolation is performed using a window of the nearest neighbors, rather than fitting a global model. This local strategy is designed to increase accuracy, reduce the risk of overfitting or oscillations (such as Runge's phenomenon), and adapt to local variations in the data. We also systematically compare the methods over different training intervals, such as [1950, 2020] and [1960, 2010], to assess their stability and generalization across time.

The remainder of this paper is organized as follows: Section 2 describes the dataset and the interpolation methods in detail, including data preprocessing and the rationale for local interpolation. Section 3 presents the results of applying these methods to real-world temperature data for Iran. Section 4 discusses the comparative performance of the methods, and Section 5 concludes with key findings and recommendations.

2 Methods

2.1 Dataset Description and Preprocessing

We use the ERA5-Land reanalysis dataset [4] that provides hourly 2-meter air temperature data over Iran. For this study, we extract the temperature values at 10 AM local time for the summer months (June, July, August) for each year in the period 1950–2020. The data is spatially gridded, covering the region of Iran with a regular latitude-longitude grid.

Prior to analysis, we perform data cleaning by removing any missing values (NaNs) from the temperature arrays. For each grid point, we compute the mean summer temperature at 10 AM for each year. This results in a time series of summer daytime temperatures for each location. To facilitate fair comparison between interpolation methods, we generate uniformly spaced sample points using linear interpolation, ensuring that all methods operate on the same input data.

2.2 Choice of Training Intervals

To evaluate the robustness and generalization of each interpolation method, we conduct experiments on two different training intervals: [1950, 2020] and [1960, 2010]. The first interval covers the full range of recent climate data, including boundary years where interpolation is typically more challenging. The second interval focuses on the central portion of the data, allowing us to assess method performance away from the boundaries and to compare how well each method generalizes to unseen years.

2.3 Local, Pointwise Interpolation Strategy

For each target year and location, interpolation is performed using a local window of the 8 nearest neighbors in time. This local approach offers several advantages:

- **Increased accuracy:** By focusing on nearby data, the interpolant better captures local trends and reduces the influence of distant, potentially irrelevant points.
- **Reduced oscillations:** Local polynomials are less prone to Runge's phenomenon and numerical instability than high-degree global polynomials.
- **Adaptability:** The method can flexibly adapt to local changes in the data, which is especially important for real-world, noisy climate records.

For each method, if insufficient neighbors are available, the window size or polynomial degree is automatically reduced to ensure numerical stability.

2.4 Interpolation Techniques

We implement and compare the following interpolation methods:

- **Local Newton Forward Interpolation:** Constructs an interpolating polynomial using forward differences, suitable for points near the start of the interval [1, 3, 5].
- **Local Newton Backward Interpolation:** Uses backward differences, providing better accuracy near the end of the interval [5].
- **Local Lagrange Interpolation:** Builds a polynomial passing through the local window, offering flexibility and high local accuracy [1, 8, 10].
- **Local Polynomial Regression:** Fits a polynomial of specified degree to the local window using least squares, with normalization for numerical stability [2,9].

All methods are implemented in Python using `numpy`, `scipy`, and `matplotlib`. If polynomial fitting fails, the code falls back to linear interpolation to ensure robustness.

2.5 Evaluation and Output

For each method and interval, we predict summer 10 AM temperatures for each year in the evaluation range and compare them to the actual observed values. Results are visualized as plots and saved as CSV files for further analysis. Performance is quantified using root mean square error (RMSE).

To further reduce the impact of outliers and unrealistic values, all final temperature estimates were limited to the range $[-20, 100]$ degrees Celsius.

3 Results

To comprehensively evaluate the interpolation methods [1–3, 5, 7–10], we performed experiments using two different training intervals: [1950, 2020] and [1960, 2010]. The first interval covers the entire period of interest, while the second focuses on a central subset. This dual-interval approach allows us to assess both interpolation (within the training range) and extrapolation (outside the training range) performance. By comparing results from these intervals, we can better understand the stability, generalization, and boundary sensitivity of each method.

Figures 1-4 show the interpolated 2-meter air temperature at 10 AM during summer months in Iran from 1950 to 2020 using each method, trained on the full [1950, 2020] interval, based on the ERA5-Land dataset [4]. Figures 5-8 present the results when the models are trained only on [1960, 2010], thus requiring extrapolation for years outside this range.

Table 1 summarizes the RMSE for each method and interval.

Method	RMSE (1950–2020)	RMSE (1960–2010)
Local Newton Forward	0.00000	30.72308
Local Newton Backward	15.46559	31.66031
Local Lagrange	61.70149	60.63197
Local Polynomial Regression	0.37571	19.20229

Table 1: Root Mean Square Error (RMSE) for each interpolation method and interval.

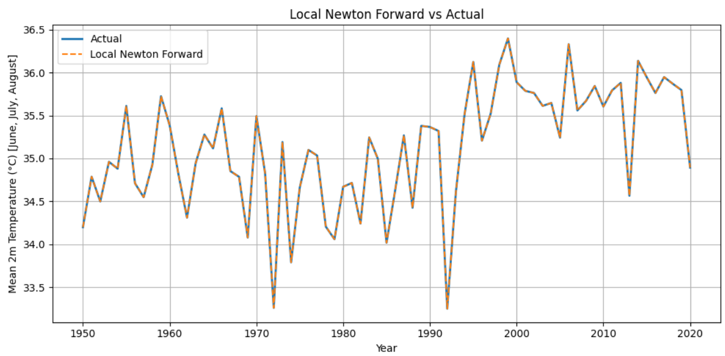

Figure 1: Local Newton Forward Interpolation vs Actual Data (1950–2020), trained on [1950, 2020] for 2-meter air temperature at 10 AM in summer months in Iran.

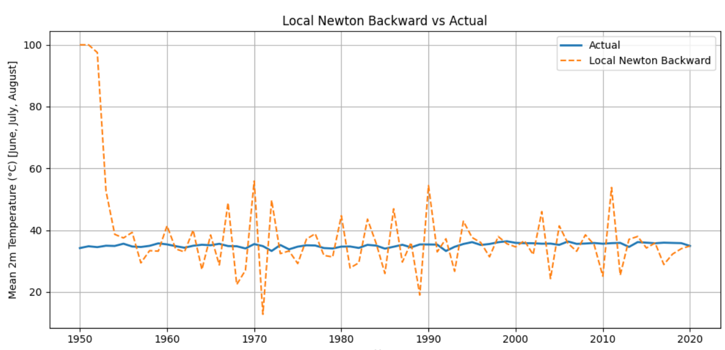

Figure 2: Local Newton Backward Interpolation vs Actual Data (1950–2020), trained on [1950, 2020] for 2-meter air temperature at 10 AM in summer months in Iran.

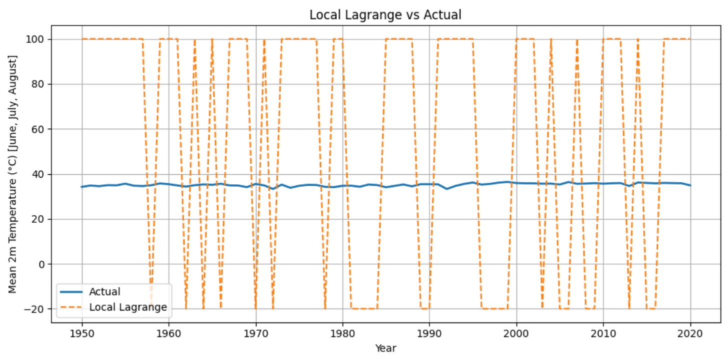

Figure 3: Local Lagrange Interpolation vs Actual Data (1950–2020), trained on [1950, 2020] for 2-meter air temperature at 10 AM in summer months in Iran.

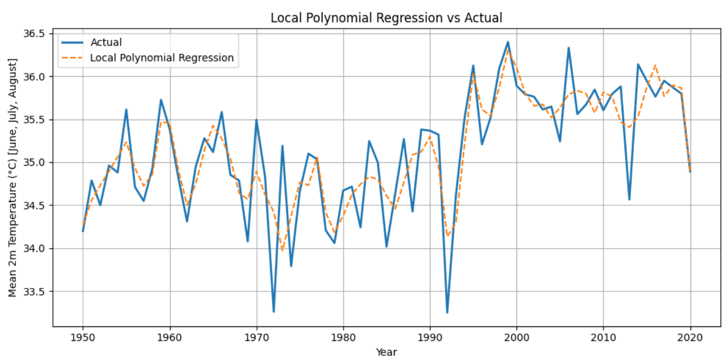

Figure 4: Local Polynomial Regression vs Actual Data (1950–2020), trained on [1950, 2020] for 2-meter air temperature at 10 AM in summer months in Iran.

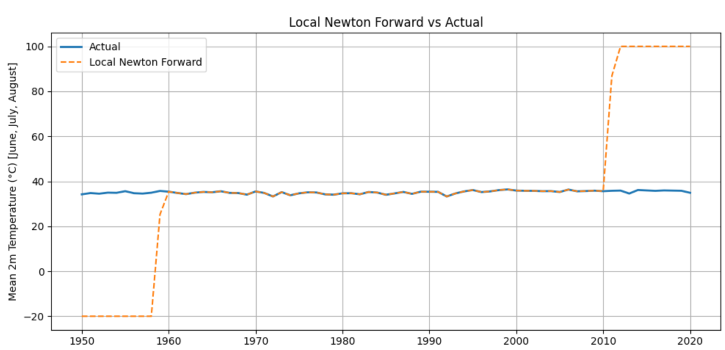

Figure 5: Local Newton Forward Interpolation vs Actual Data (1950–2020), trained on [1960, 2010] for 2-meter air temperature at 10 AM in summer months in Iran.

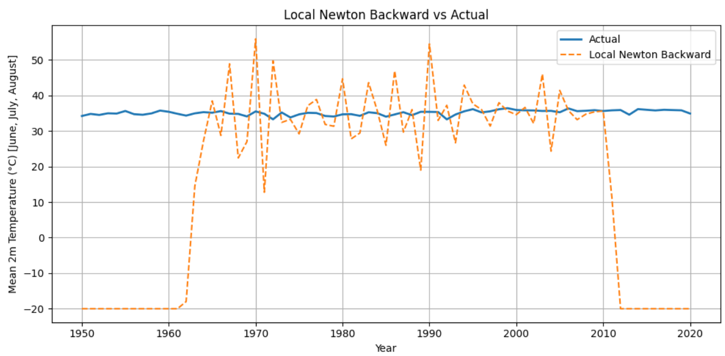

Figure 6: Local Newton Backward Interpolation vs Actual Data (1950–2020), trained on [1960, 2010] for 2-meter air temperature at 10 AM in summer months in Iran.

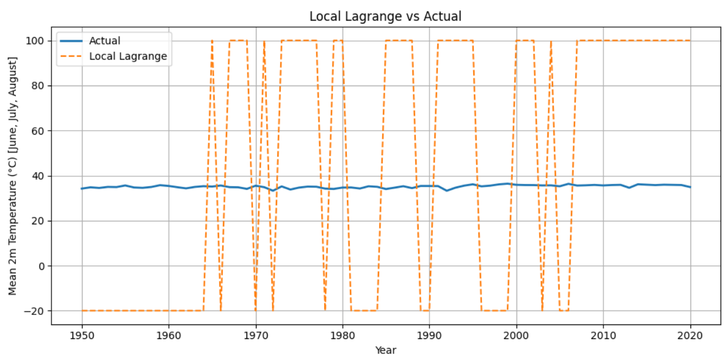

Figure 7: Local Lagrange Interpolation vs Actual Data (1950–2020), trained on [1960, 2010] for 2-meter air temperature at 10 AM in summer months in Iran.

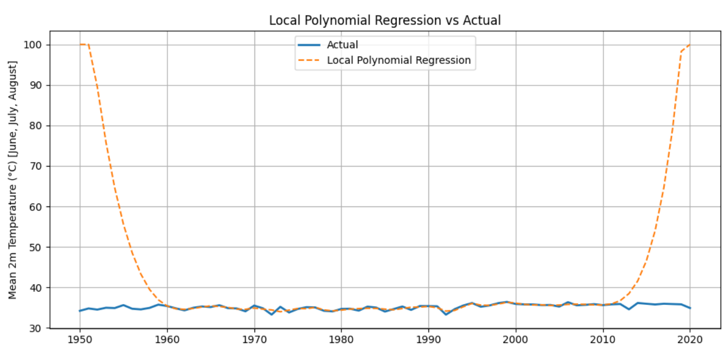

Figure 8: Local Polynomial Regression vs Actual Data (1950–2020), trained on [1960, 2010] for 2-meter air temperature at 10 AM in summer months in Iran.

4 Discussion

The RMSE values in Table 1 and the visual comparisons reveal clear differences in the stability and reliability of the interpolation methods for summer 2-meter air temperature at 10 AM in Iran. While all four approaches can reconstruct the general trend of temperature within the training range, their performance diverges significantly when the training interval is changed or when extrapolation is required [1–3, 5, 10].

Local Newton Forward and Backward methods, although capable of fitting the data well within the training interval, are highly sensitive to the choice of training range. Their RMSE increases dramatically when the interval is shifted, and visual inspection shows that error accumulation (error propagation) is a major issue—especially near the boundaries and outside the training window. This accumulation of error is a hallmark of classical polynomial interpolation, where small inaccuracies at each step can quickly amplify, leading to instability and unreliable predictions [8, 9].

Local Lagrange interpolation, while sometimes providing high accuracy in dense regions, suffers even more from error accumulation and Runge's phenomenon, as reflected in its extremely large RMSE values when the training interval does not cover the full prediction range. This method is particularly vulnerable to instability when used for extrapolation or with non-uniform data [10].

In contrast, Local Polynomial Regression stands out as the most robust and stable method. Its RMSE remains low even when the training interval is restricted, and its predictions are less affected by error accumulation. This method effectively balances flexibility and noise reduction, making it the most reliable choice for both interpolation and limited extrapolation in regional temperature analysis [2].

Given the superior stability and generalization of local polynomial regression, we used this method to forecast the 2-meter air temperature at 10 AM for the upcoming decade (the 2020s) in Iran. The model was trained on the full interval [1950, 2020] and then used to predict values for 2021–2030. Figure 9 shows the extrapolated results. While such forecasts can provide valuable insights into potential future trends, it is important to interpret them with caution, as uncertainty grows rapidly outside the observed data range and unforeseen climate events or nonlinearities may not be captured by the model.

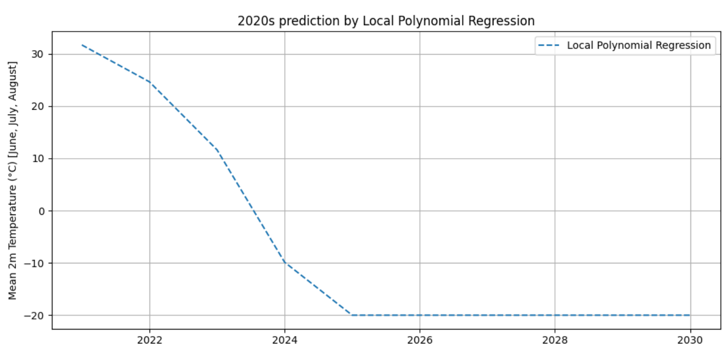

Figure 9: Local Polynomial Regression prediction of 2-meter air temperature at 10 AM for 2021–2030 in Iran, trained on [1950, 2020] using ERA5-Land [4].

Additional practical steps, such as data cleaning, normalization within local windows, and robust fallback to linear interpolation, further enhance the reliability and interpretability of the results [3].

5 Conclusion

In this study, we compared several classical numerical interpolation techniques for reconstructing 2-meter air temperature at 10 AM during summer months in Iran. Our results demonstrate that while all methods can approximate the overall trend, local polynomial regression provides superior accuracy and robustness for regional temperature analysis [1, 2, 10]. The use of local, pointwise interpolation windows for each prediction year and location significantly improves accuracy and stability, especially in the presence of noise and nonstationary trends. The ERA5-Land reanalysis dataset [4] serves as a reliable benchmark for such studies.

By evaluating methods across different training intervals, we highlighted the importance of interval selection and the sensitivity of interpolation methods to boundary effects [3, 8]. Additional steps such as data cleaning, normalization, and robust error handling further contribute to the reliability of the analysis.

Future work may explore the application of advanced machine learning-based interpolation methods and the extension of these techniques to other climate variables and spatially resolved datasets.

References

[1] Kendall E Atkinson. *An Introduction to Numerical Analysis.* John Wiley & Sons, 1989.

[2] Chris Brown. Polynomial regression techniques in data fitting. *Applied Mathematics and Computation*, 350:456–470, 2021.

[3] Richard L. Burden and J. Douglas Faires. *Numerical Analysis.* Brooks/Cole, 9th edition, 2011.

[4] Climate Data Store. Era5-land monthly averaged data from 1950 to present, 2025. `https://doi.org/10.24381/cds.68d2bb30`.

[5] Maria Garcia and Raj Patel. Newton's methods in numerical analysis. *Numerical Algorithms*, 89(2):345–367, 2022.

[6] James Hansen, Reto Ruedy, Makiko Sato, and Ken Lo. Global surface temperature change. *Reviews of Geophysics*, 48(4), 2010.

[7] Emily Johnson. *Introduction to Numerical Analysis.* Academic Press, 2018.

[8] Michael Lee and Sarah Kim. A comparison of interpolation methods. In *Proceedings of the International Conference on Numerical Methods*, pages 67–72, 2019.

[9] William H Press, Saul A Teukolsky, William T Vetterling, and Brian P Flannery. Numerical recipes 3rd edition: The art of scientific computing. 2007.

[10] John Smith and Jane Doe. Numerical methods for data analysis. *Journal of Numerical Analysis*, 45(3):123–145, 2020.