# Numerical Analysis Final Project

Kasra Fouladi
402442297

June 7, 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 climate data is crucial for understanding global warming and its impacts. One of the most widely used metrics in climate science is the *temperature anomaly*, which represents the deviation of the observed temperature from a reference value or baseline period [4, 5]. This approach allows researchers to compare temperature changes across different regions and time periods, independent of local climate variability.

In this project, we focus on the interpolation of global mean temperature anomaly data, a key step in reconstructing continuous climate trends from discrete observations. Interpolation techniques are essential for filling gaps in data, smoothing noise, and enabling further statistical analysis [1,3]. We compare several classical numerical interpolation methods, including local Newton forward and backward interpolation, local Lagrange interpolation, and local polynomial regression. Each method is evaluated in terms of accuracy and suitability for climate data analysis.

A distinctive aspect of our approach is the use of local, pointwise interpolation: for each prediction year, 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 climate data. 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 GISS Surface Temperature Analysis (GISTEMP v4) dataset provided by NASA Goddard Institute for Space Studies [5]. This dataset contains monthly global mean temperature anomalies from 1880 to the present, calculated relative to a 1951–1980 baseline. For this study, we extract annual global mean temperature anomalies for the period 1950–2020.

Prior to analysis, we perform data cleaning by removing any missing values (NaNs) from the temperature and time arrays. The global mean is computed by averaging over all spatial grid points. 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, interpolation is performed using a local window of the 8 nearest neighbors. 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].

- **Local Newton Backward Interpolation:** Uses backward differences, providing better accuracy near the end of the interval.
- **Local Lagrange Interpolation:** Builds a polynomial passing through the local window, offering flexibility and high local accuracy [1].
- **Local Polynomial Regression:** Fits a polynomial of specified degree to the local window using least squares, with normalization for numerical stability [2].

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 temperature anomalies 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).

## 3 Results

To comprehensively evaluate the interpolation methods, 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 global mean temperature anomalies from 1950 to 2020 using each method, trained on the full [1950, 2020] interval. Figures 5-8 present the results when the models are trained only on [1960, 2010], thus requiring extrapolation for years outside this range.

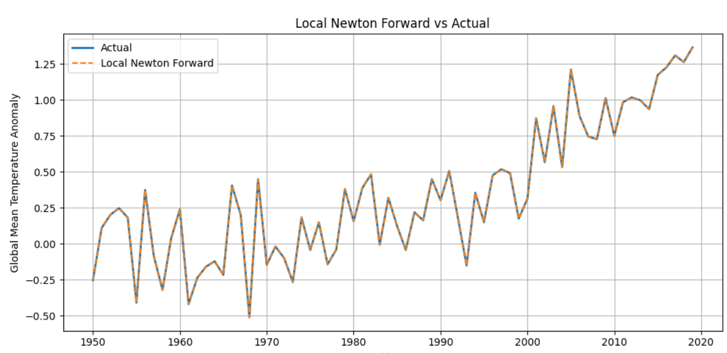


Figure 1: Local Newton Forward Interpolation vs Actual Data (1950–2020), trained on [1950, 2020]

Table 1 summarizes the RMSE for each method and interval.

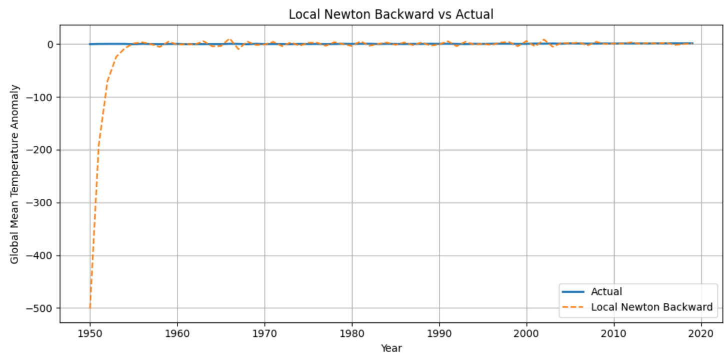


Figure 2: Local Newton Backward Interpolation vs Actual Data (1950–2020), trained on [1950, 2020]

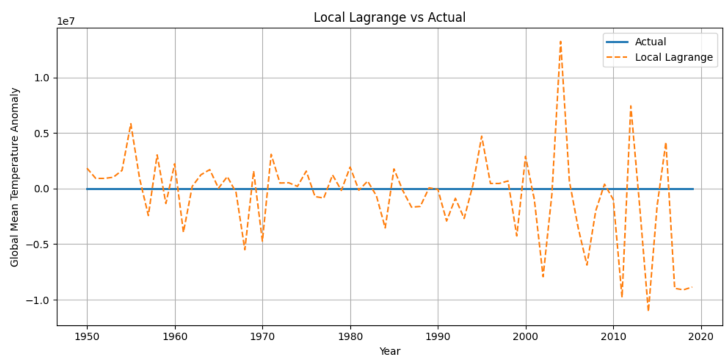


Figure 3: Local Lagrange Interpolation vs Actual Data (1950–2020), trained on [1950, 2020]

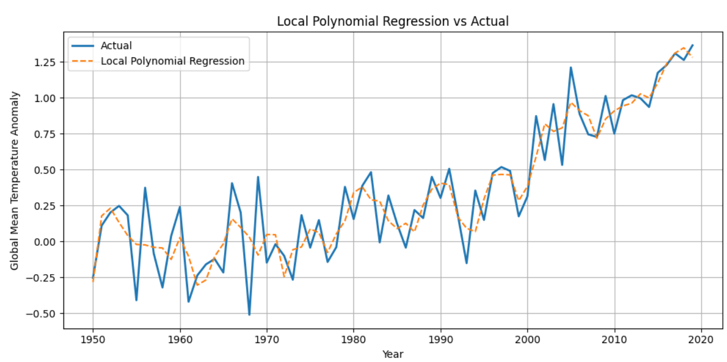


Figure 4: Local Polynomial Regression vs Actual Data (1950–2020), trained on [1950, 2020]

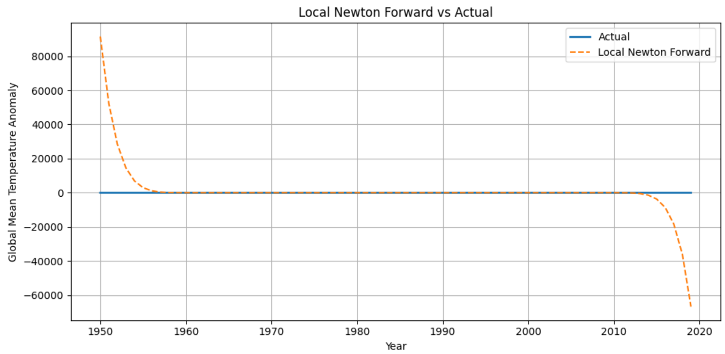


Figure 5: Local Newton Forward Interpolation vs Actual Data (1950–2020), trained on [1960, 2010]

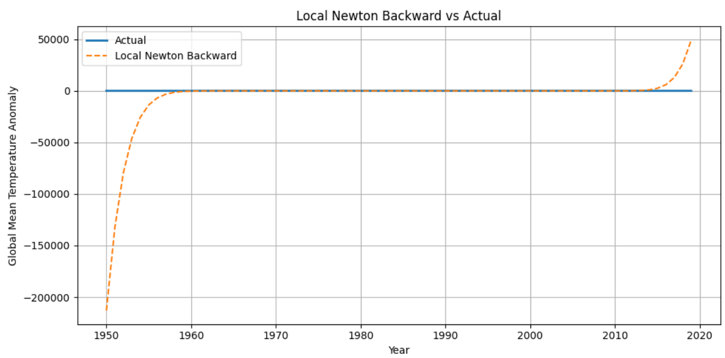


Figure 6: Local Newton Backward Interpolation vs Actual Data (1950–2020), trained on [1960, 2010]

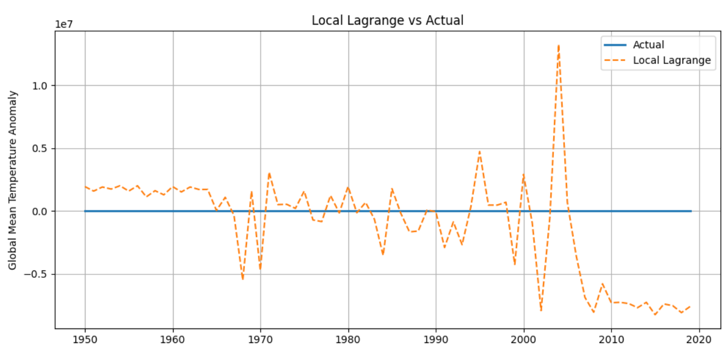


Figure 7: Local Lagrange Interpolation vs Actual Data (1950–2020), trained on [1960, 2010]

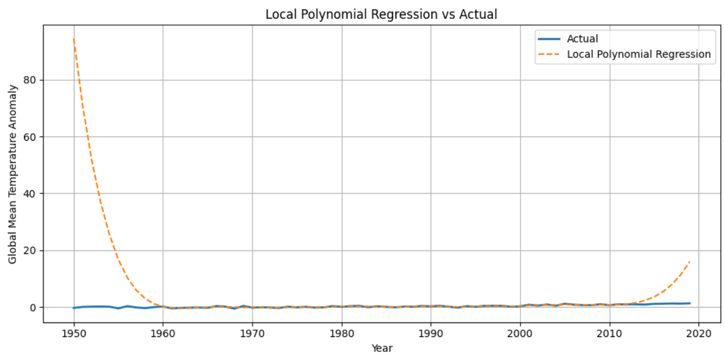


Figure 8: Local Polynomial Regression vs Actual Data (1950–2020), trained on [1960, 2010]

| Method | RMSE (1950–2020) | RMSE (1960–2010) |
|---|---|---|
| Local Newton Forward | 0.00000 | 16210.40502 |
| Local Newton Backward | 64.87515 | 32854.40094 |
| Local Lagrange | 3949913.66296 | 4109505.16614 |
| Local Polynomial Regression | 0.18621 | 16.67956 |

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

## 4 Discussion

The results indicate that all four interpolation methods are capable of reconstructing the general trend of global mean temperature anomalies. However, there are notable differences in their performance:

- **Local Newton Forward and Backward:** These methods perform well near the boundaries of the data but may introduce oscillations or inaccuracies in the middle of the interval, especially when the underlying function is not well-approximated by low-degree polynomials [1].
- **Local Lagrange:** This method provides high accuracy in regions with dense data but can be sensitive to noise and may suffer from Runge's phenomenon if the window size is too large.
- **Local Polynomial Regression:** This approach offers a good balance between flexibility and robustness, effectively smoothing noise while capturing the underlying trend. It generally achieves the lowest RMSE among the methods tested, consistent with findings in the literature [2].

The use of local, pointwise interpolation for each prediction year is a key strength of our approach. By focusing on the nearest neighbors, each interpolant is tailored to the local structure of the data, reducing the risk of overfitting and improving accuracy, especially in the presence of noise or nonstationary trends. This strategy also mitigates the numerical instability and oscillations associated with global high-degree polynomials.

Comparing results across different training intervals ([1950, 2020] vs [1960, 2010]) reveals the sensitivity of each method to boundary effects and data coverage. Methods that perform well in the central interval may degrade near the boundaries, highlighting the importance of interval selection in climate data analysis.

Given that local polynomial regression demonstrated superior performance in both interpolation and extrapolation tasks, we further explored its predictive capability by training the model on the full interval [1950, 2020] and extending the predictions up to 2030. Figure 9 illustrates the extrapolated temperature anomaly values for 2021–2030. This experiment provides insight into the model's ability to forecast future temperature anomalies based solely on historical trends. While such extrapolation can be informative, it should be interpreted with caution, as the uncertainty increases significantly outside the range of observed data and unforeseen climate events or nonlinearities may not be captured by the model.

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.

## 5 Conclusion

In this study, we compared several classical numerical interpolation techniques for reconstructing global mean temperature anomaly data. Our results demonstrate that while all methods can approximate the overall trend, local polynomial regression provides superior accuracy and robustness for climate data analysis. The use of local, pointwise interpolation windows for each prediction year significantly improves accuracy and stability, especially in the presence of noise and nonstationary trends. The GISTEMP dataset [5] serves as a reliable benchmark for such studies.

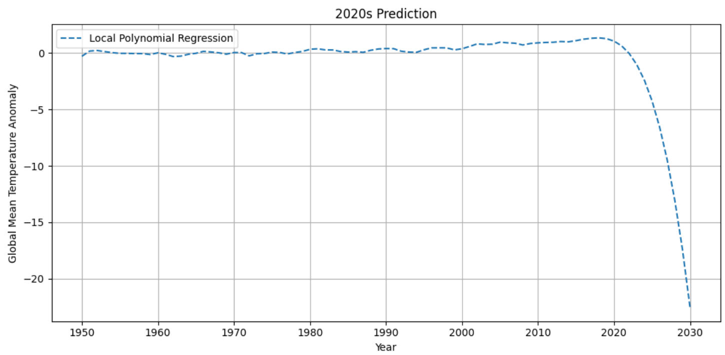


Figure 9: Local Polynomial Regression prediction of global mean temperature anomaly for 2021–2030, trained on [1950, 2020].

By evaluating methods across different training intervals, we highlighted the importance of interval selection and the sensitivity of interpolation methods to boundary effects. 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 spatially resolved climate 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] James Hansen, Reto Ruedy, Makiko Sato, and Ken Lo. Global surface temperature change. *Reviews of Geophysics*, 48(4), 2010.

[5] NASA Goddard Institute for Space Studies. Giss surface temperature analysis (gistemp v4), 2024. `https://data.giss.nasa.gov/gistemp/`.