Weather and Climate Big Data

기상기후빅데이터

Hands-on Exercise 3: Spatial Interpolation

오승민

Spring 2025

Today’s Topics

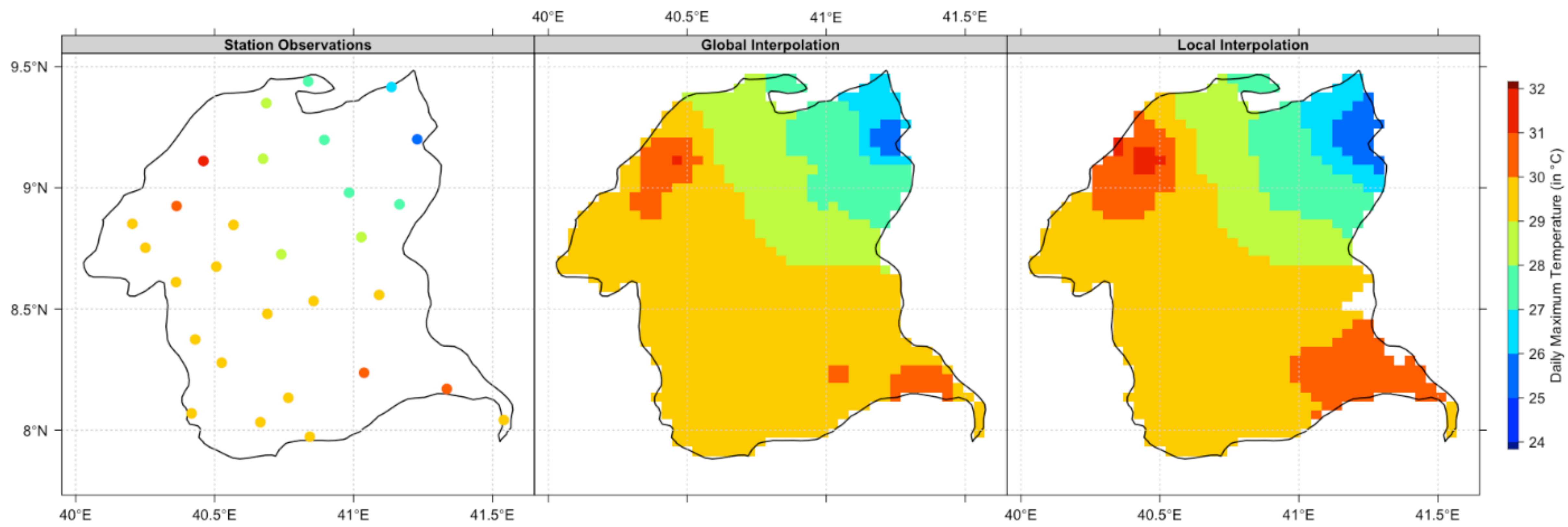
- Spatial Data
- Spatial Interpolation Techniques

What is Spatial Data?

- **Spatial data** refers to data that has a geographic or spatial component — meaning it relates to a specific location on Earth.

- **Examples:**
  - Meteorological station observations (temperature, precipitation at specific points)
  - Satellite imagery (gridded, continuous)
  - Model and reanalysis data (gridded, continuous)

Spatial Interpolation



Spatial Interpolation

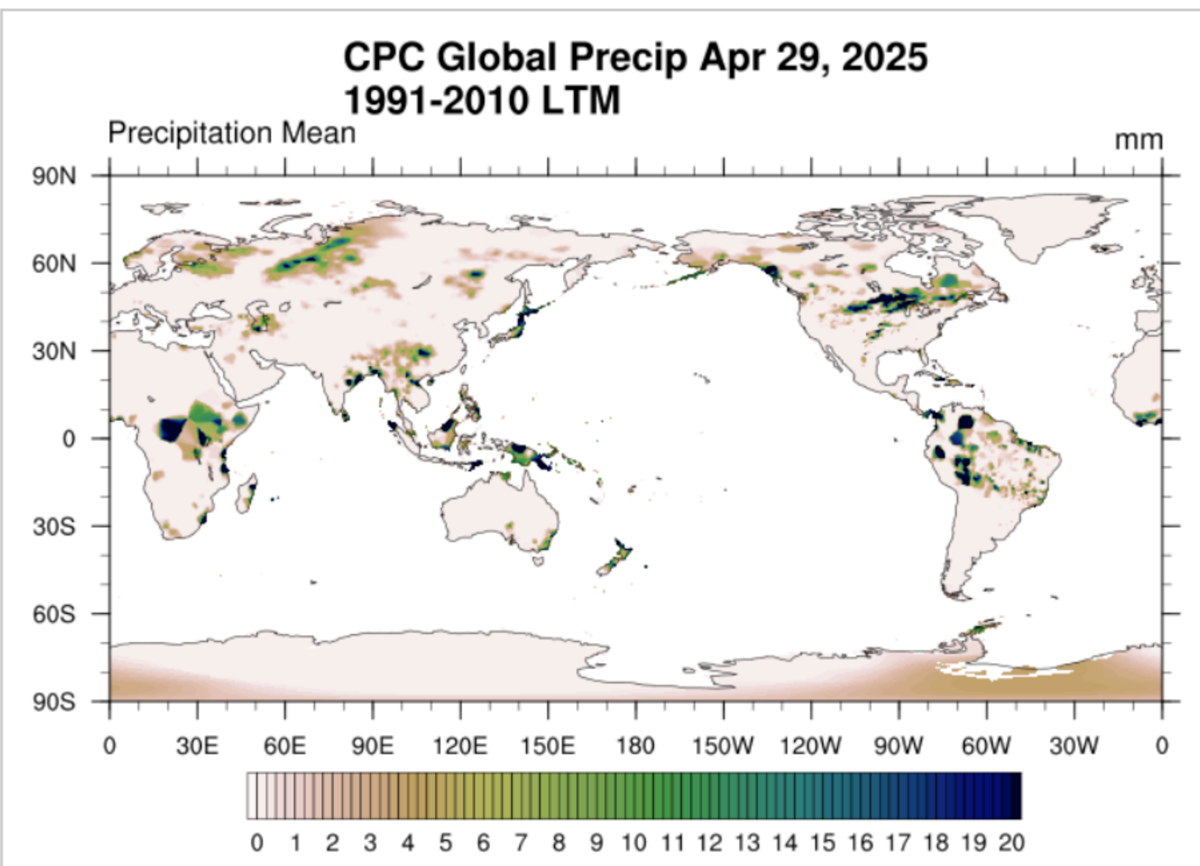
- If we already have **satellite** or **model** gridded data, why do interpolation from station data?
    - **Higher accuracy**: Station data often represents ground truth better at local scales.
    - **Validation**: Ground-based data can validate or correct model/ satellite products.
    - **High-resolution Mapping**: Create finer, customized grids for specific regions.

- Data can be created by interpolation or prediction of models

Spatial Interpolation

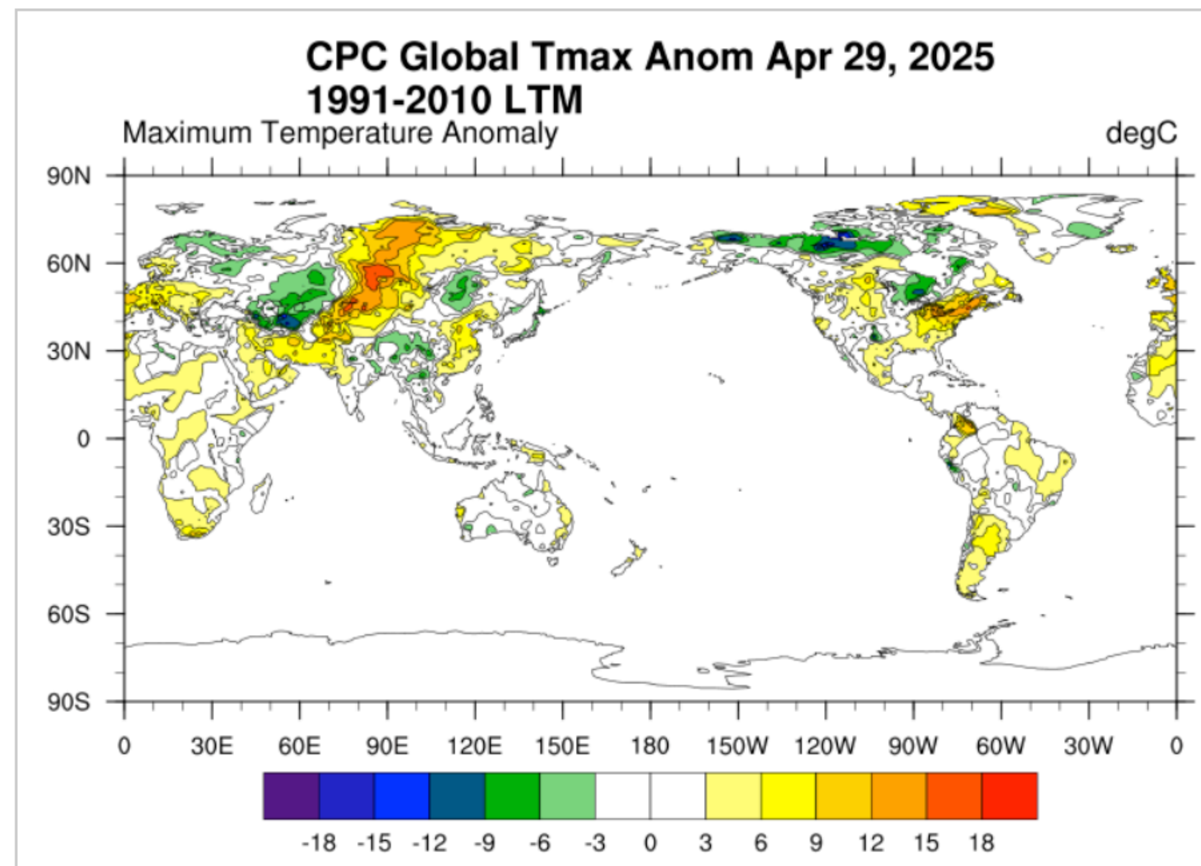
CPC Global Unified Gauge-Based /

CPC 0.5x0.5 Global Daily Unified Gauge-Based Analysis of Precipita



CPC Global Unified Temperature

CPC 0.5x0.5 Global Daily Gridded Temperature



Inverse Distance Weighting (IDW)

- A geostatistical technique used to estimate values at unmeasured locations based on the distances to measured locations

$$\hat{Z}(x_0) = \frac{\sum_{i=1}^{N} w_i \cdot Z(x_i)}{\sum_{i=1}^{N} w_i} \quad \text{where} \quad w_i = \frac{1}{d(x_0, x_i)^p}$$

  - $Z(x_0)$: Estimated value at target location
  - $Z(x_i)$: Observed value at known location
  - w_i: Weight based on the inverse of the distance
  - $d(x_0,x_i)$: Distance between x0 and xi
  - p: Power parameter controlling influence of distance (commonly p = 2)

Inverse Distance Weighting (IDW)

- A geostatistical technique used to estimate values at unmeasured locations based on the distances to measured locations
- Pros:
  - Simple and intuitive
  - Easy to implement and fast to compute
  - No need to model spatial correlation explicitly
- Cons:
  - Sensitive to choice of power (p) and number of neighbors (N)
  - Tends to produce smooth, “bullseye”-shaped patterns
  - Less accurate in sparse or irregular data regions

Kriging

- Kriging is a geostatistical method used for spatial interpolation, estimating values at unmeasured locations based on known values at nearby locations

- **Observed data points**: Locations where values are measured.
- **Spatial correlation model:**
  - Kriging relies on modeling how related nearby points are.
  - This relationship is captured through a **variogram**.
- **Kriging system:**
  - Solves for optimal weights based on distance and spatial structure.
- **Outputs:** Provides both predicted values and associated uncertainty.

Kriging Workflow Overview

1. Gather observed data points.
2. Calculate empirical variogram using the data.
3. Fit a theoretical variogram model.
4. Solve Kriging equations using variogram.
5. Predict new values and estimate uncertainty.

Variogram

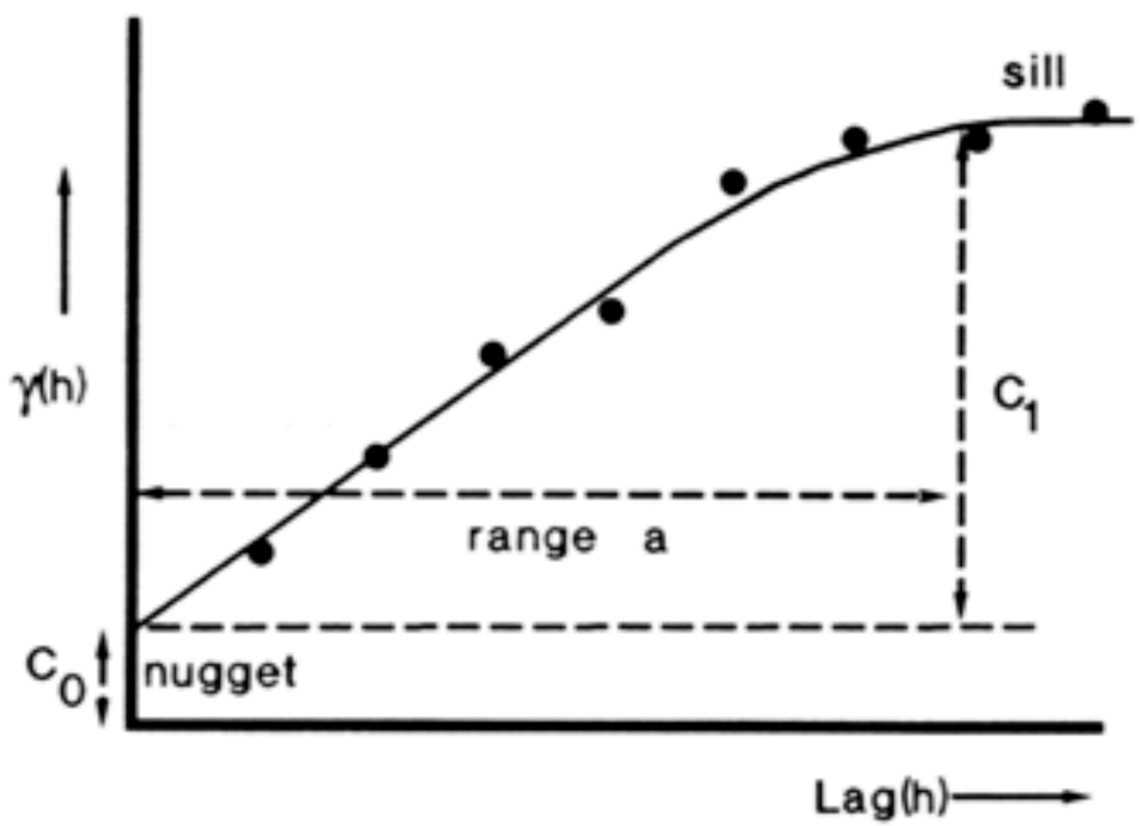
- A **variogram** quantifies how the similarity between data points **decreases as the distance between them increases.**

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(x_i) - Z(x_i + h)]^2$$

- **h**: Distance between two locations
- **γ(h)**: Semi-variance, i.e. the average squared difference between values at distance h

Variogram

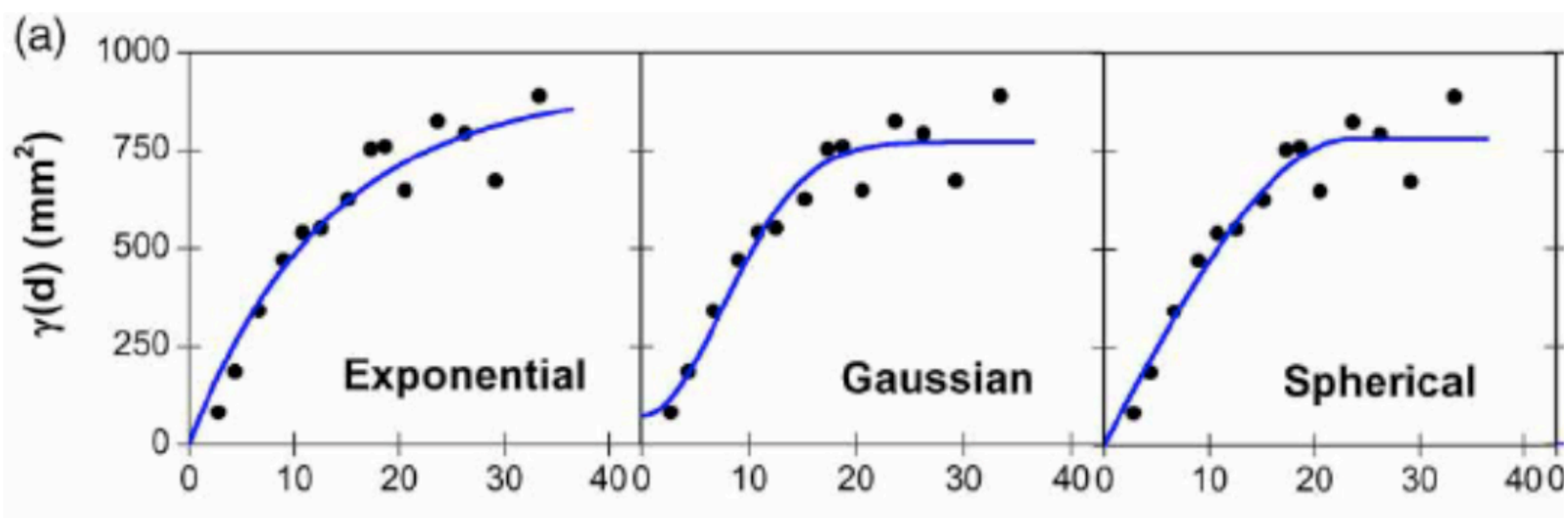
- A **variogram** quantifies how the similarity between data points **decreases as the distance between them increases.**



- Nugget: Variation at zero distance (e.g. measurement error)
- Sill: Maximum variance level — where correlation levels off
- Range: Distance at which points are no longer correlated

Variogram

- We fit a smooth variogram model to noisy empirical data for use in kriging.
  - **Spherical**: Increases up to the range, then levels off sharply.
  - **Exponential**: Gradually approaches the sill, with a long tail.
  - **Gaussian**: Very smooth near the origin; best for high-resolution data.



Kriging Workflow Overview

1. Gather observed data points.
2. Calculate empirical variogram.
3. Fit a theoretical variogram model.
4. Solve Kriging equations using variogram.
5. Predict new values and estimate uncertainty.

Kriging

- Kriging is not a single algorithm, but a family of spatial statistical models.
  - Ordinary Kriging: Assumes an unknown but constant mean across the region.
  - Simple Kriging: Assumes a known and fixed mean throughout the domain.
  - Universal Kriging: Accounts for spatial trend in the data.
  - Co-Kriging: Uses multiple correlated variables to improve prediction accuracy.

Ordinary Kriging

$$Z^*(x_0) = \sum_{i=1}^{N} w_i \cdot Z(x_i)$$

- w_i: Weight assigned to the observed point x_i
- $Z(x_i)$: Observed value at location
- N: Number of neighboring points used in the prediction

Ordinary Kriging 시스템 (행렬식):

$$\begin{bmatrix} \gamma(x_1, x_1) & \gamma(x_1, x_2) & \dots & \gamma(x_1, x_N) & 1 \\ \gamma(x_2, x_1) & \gamma(x_2, x_2) & \dots & \gamma(x_2, x_N) & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \gamma(x_N, x_1) & \gamma(x_N, x_2) & \dots & \gamma(x_N, x_N) & 1 \\ 1 & 1 & \dots & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_N \\ \lambda \end{bmatrix} = \begin{bmatrix} \gamma(x_1, x_0) \\ \gamma(x_2, x_0) \\ \vdots \\ \gamma(x_N, x_0) \\ 1 \end{bmatrix}$$

Other Advanced Methods

- **Spline interpolation**:
  - Fits a smooth surface minimizing overall curvature.
  - Good for smoothly varying fields.
- **Co-Kriging**:
  - Uses secondary variables (e.g. elevation) to improve prediction.
- **Machine Learning models**:
  - Random Forests, Gradient Boosting, and Deep Learning can be used for spatial prediction when many auxiliary datasets are available.

Summary

- **Spatial interpolation** helps convert irregular point data into continuous fields.
- **IDW**: Simple, fast, distance-only.
- **Kriging**: Statistically optimal, models spatial structure.

=> *Let's compare the IDW VS Kriging with Meteorological data!*