STAT 517 (WI 24): HOME WORK 2

SARAYU GUNDLAPALLI
STUDENT NO: 2325381

Exercise 1

Problem 3.4 in Section 3.14 in "Model-based Geostatistics" by P. Diggle and P. J. Ribeiro.[1]

Solution. This exercise considers a method for simulating realizations of a one-dimensional spatial process using a specified set of points and correlation functions.

Key Steps:

- Selection of Points: Choose a set of points $x_i \in \mathbb{R}$ for i = 1,...,n.
- Correlation Matrix: Compute the correlation matrix R of the spatial process at the selected points.
- Singular Value Decomposition: Decompose R into its eigenvectors (D) and eigenvalues (Λ).
- Independent Random Sample: Generate an independent random sample $Y = Y_1, ..., Y_n$ from the standard Gaussian distribution $N(0, 1)$.
- Compute the simulated realization $S = D\Lambda^{1/2}Y$.

The R code implements this method by taking the locations x, correlation function ρ, and the number of simulations *nsim*. It calculates R, its SVD, samples Y and then constructs the realized simulations.

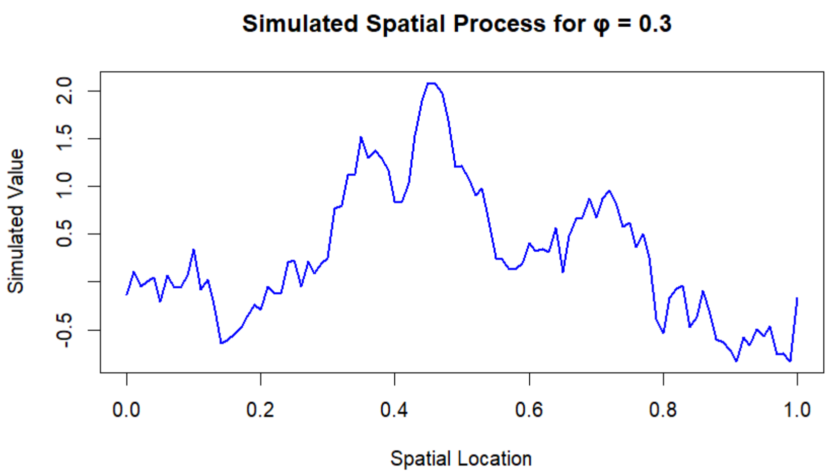
- We set x as 100 equally spaced points between 0 and 1 and use an exponential correlation function with a range parameter of 0.3, 2, 10.1a1b1c

Eigenvalue Truncation Analysis:

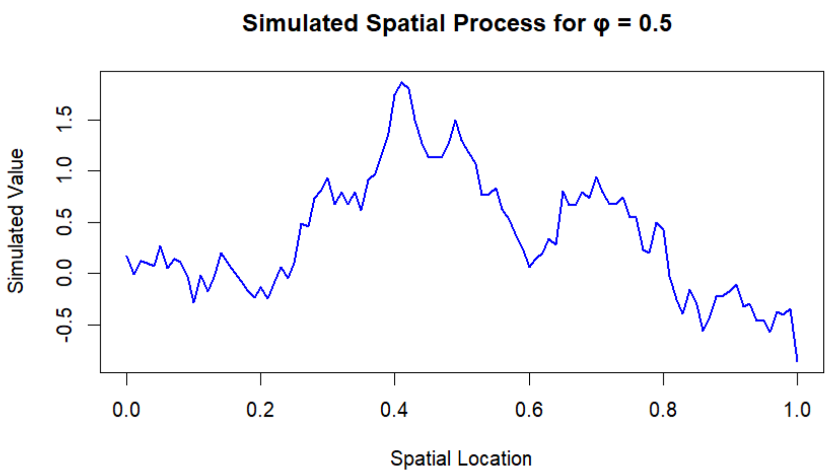
To investigate the impact of eigenvalue truncation, the diagonal matrix Λ is modified by replacing the last k eigenvalues with zeros. We truncate the smallest 10 2b, 50 2c, and 90 2d eigenvalues to 0.

We can see that truncating more eigenvalues leads to " smoother " realizations. Truncating the eigenvalues filters out the finer variation and retains the broader, smooth spatial correlation structure.

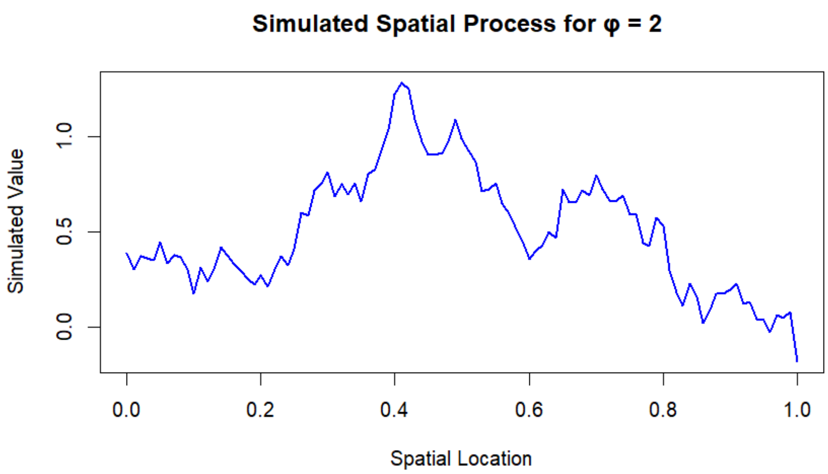
Date: January 27, 2024.



(A) Figure A

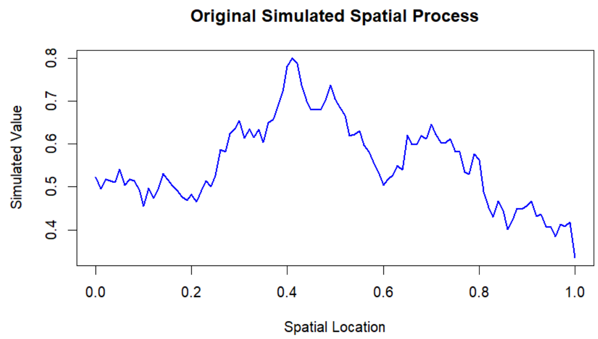


(B) Figure B

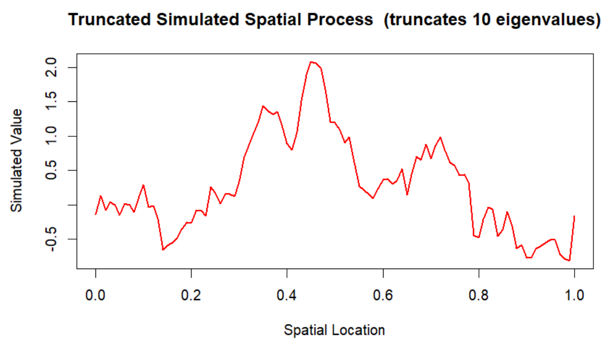


(C) Figure C

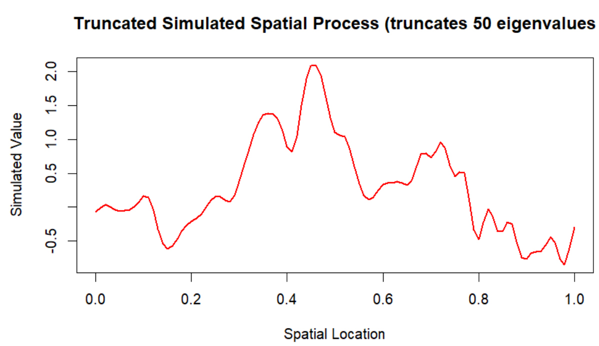
FIGURE 1. Simulated Spatial Process with $\phi = 0.3, 0.5, 2$



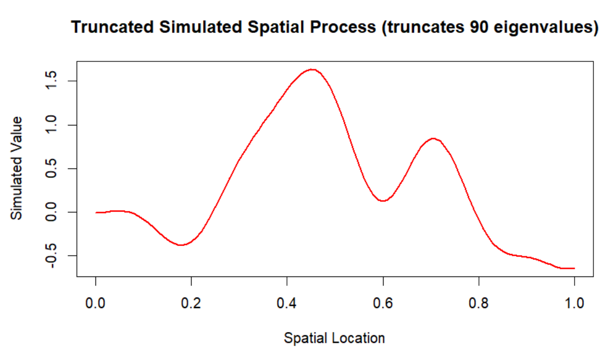
(A) Figure A



(B) Figure B



(C) Figure C



(D) Figure A

FIGURE 2. Original Simulated Spatial Process with $\phi = 0.3$ vs Truncated Simulated Spatial Process

Exercise 2

Part 1.

Create and visualize a simulated dataset of a Gaussian process (or field) in 1D (or in 2D) with a quadratic trend. The simulated dataset should have more than 50 observation times (or locations).

Solution:

The key steps are specifying the mean and covariance functions, calculating the covariance matrix, and sampling from the multivariate normal distribution "mvrnorm". The simulated GP is then plotted, showing the observed data and the predicted quadratic trend.

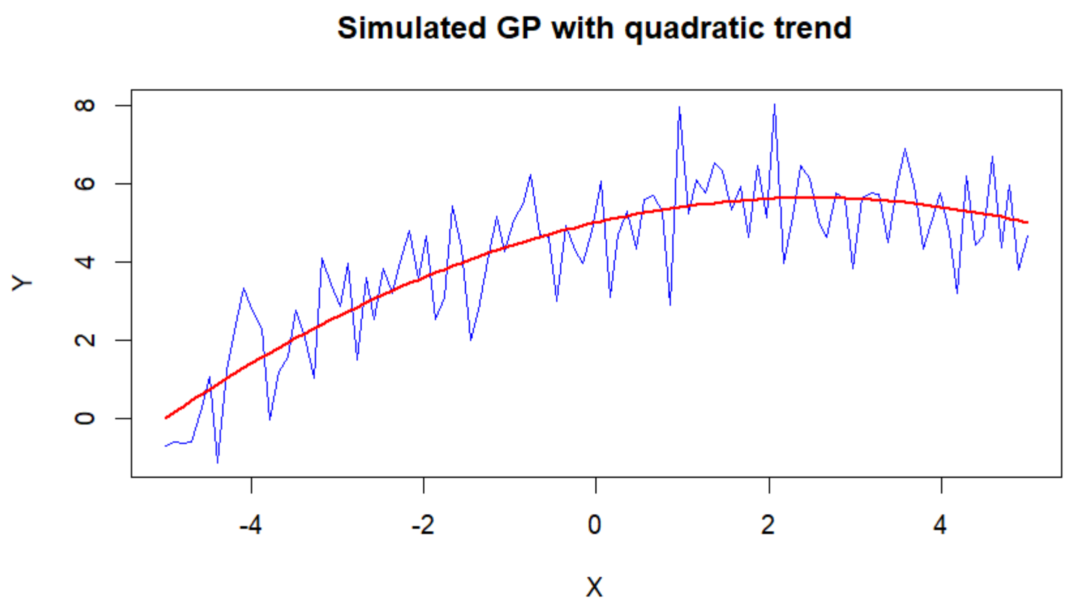


Figure 3. Simulated GP (dataset) with quadratic trend

The generated dataset represents a Gaussian process with a quadratic trend. The plot visually depicts the simulated GP with the added red line for reference.

Part 2.

Pick one of the estimation methods from Lab 1. Fit a statistical Gaussian process (or field) model assuming no trend. You may assume the covariance function parameters are perfectly known.

Solution:

A GP is simulated using a covariance matrix generated by "mvrnorm". The Ordinary Least Squares (OLS) estimation method is used.

For a d-dimensional process, the *ordinary least squares* estimator of the parameter vector $\theta \in \Theta \subset R^d$ is

$$\widehat{\theta} = \arg\min_{\theta\in\Theta} \sum_{k=1}^{K} [\widehat{\gamma}(u_k) - \gamma(u_k, \theta)]^2 \tag{1}$$

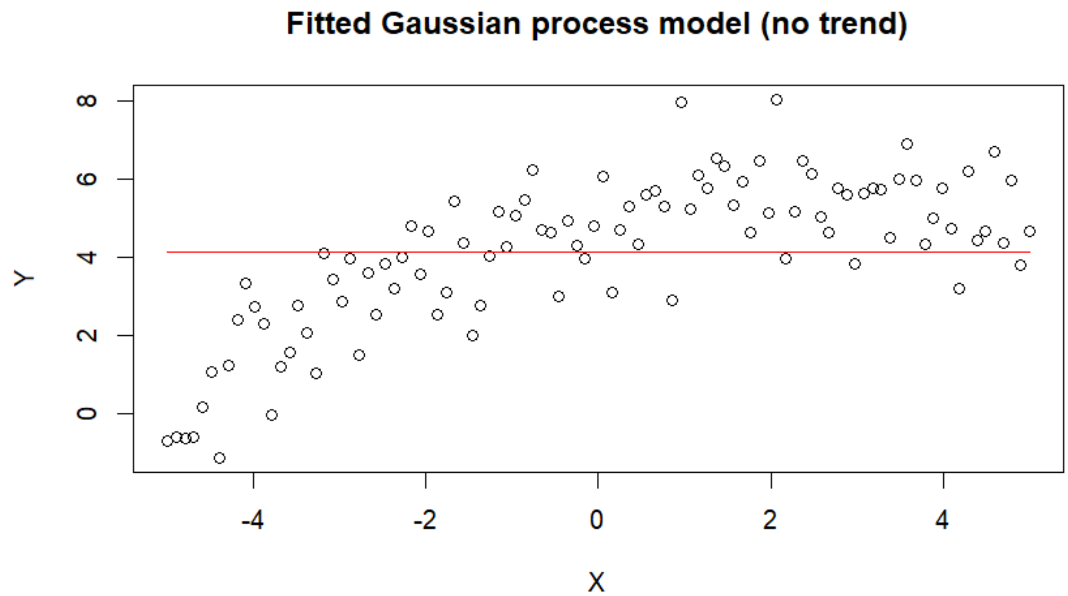


FIGURE 4. OLS fitted GP model (no trend)

The OLS fitting is applied to estimate the parameters for the GP without a trend. The resulting GP model and the original data are then plotted, providing a comparison between the simulated GP and the OLS-fitted GP.

Part 3.

Implement yourself from scratch a leave-one-out estimator of the prediction error. Apply this prediction error estimator to estimate the prediction error of your statistical model on your simulated dataset.

Solution:

The objective is to assess the predictive performance of the model by leaving out one observation at a time, refitting the model, and calculating the prediction error.

- a loop iterates over each observation, leaving it out in each iteration.
- The model is refitted with the remaining data, and the prediction error for the left-out observation is calculated.
- The average prediction error across all observations is then computed, which is essentially the LOO prediction error for the GP without a trend.

For no trend GP, LOO error was 3.857611

0.1. **Part 4.**

Redo the previous steps now assuming a linear trend, and then assuming a quadratic trend.

(1) Linear Trend:

Data with a linear trend is simulated, and OLS fitting is applied to estimate the linear trend parameters. LOO is then performed, leaving out one observation at a time and evaluating the prediction error. The observed data, predicted values, and LOO errors are visualized through plots.

```
y <- 0.5 * x + rnorm(n, 0, 1)
```

LOO Prediction Error (Linear Trend): 0.9882892

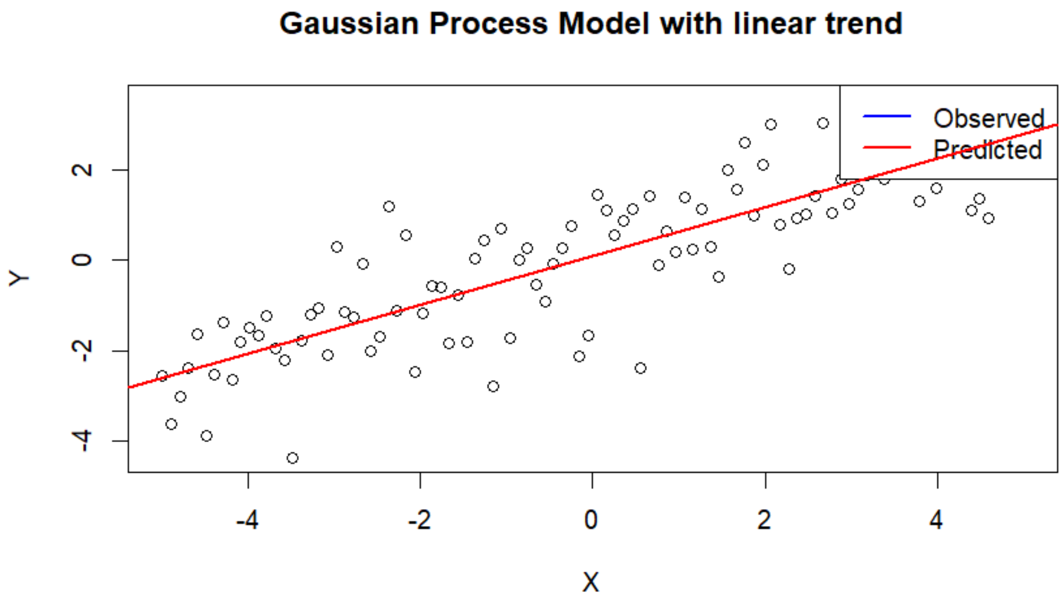


FIGURE 5. GP model with linear trend

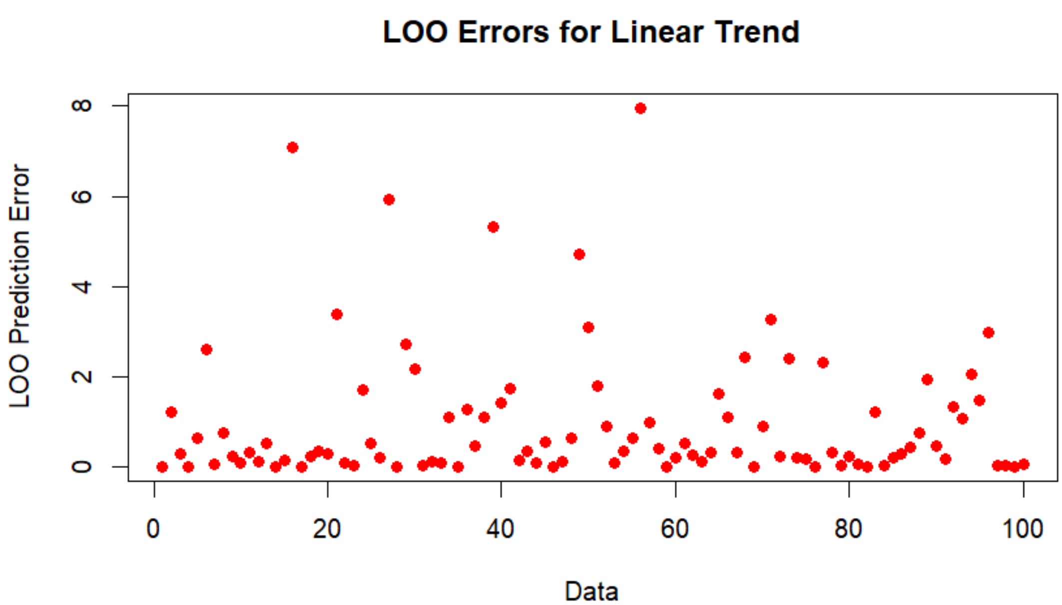


FIGURE 6. LOO Errors for linear trend

(2) Quadratic Trend:

Data with a quadratic trend is simulated, and the same method above is applied.

```
y <- 0.5 * x + 0.1 * x^2 + rnorm(n, 0, 1)
```

LOO Prediction Error (Quadratic Trend): 0.9182262

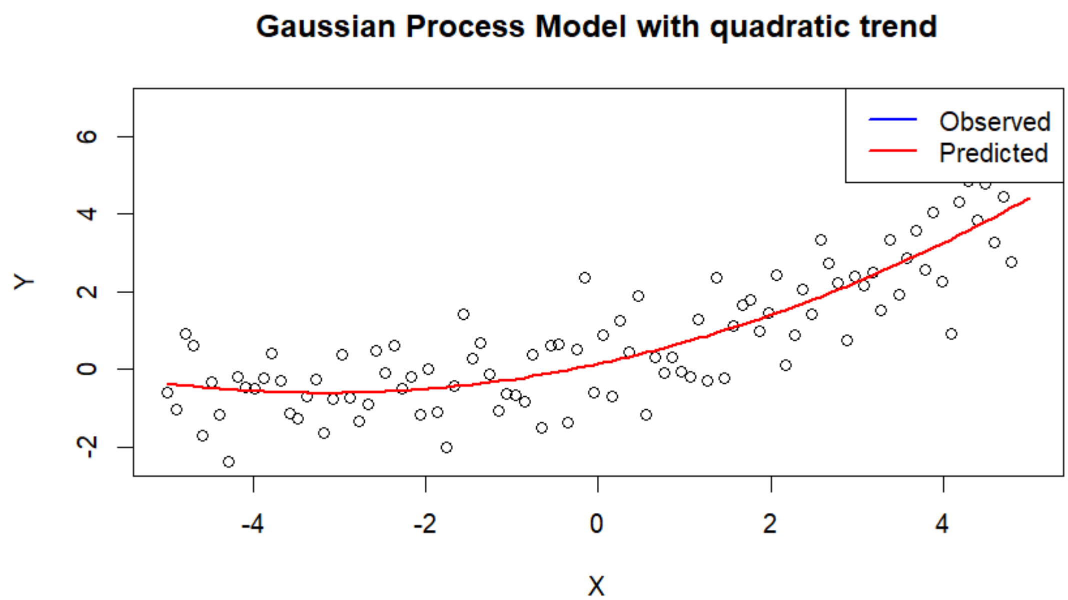


FIGURE 7. GP model with quadratic trend

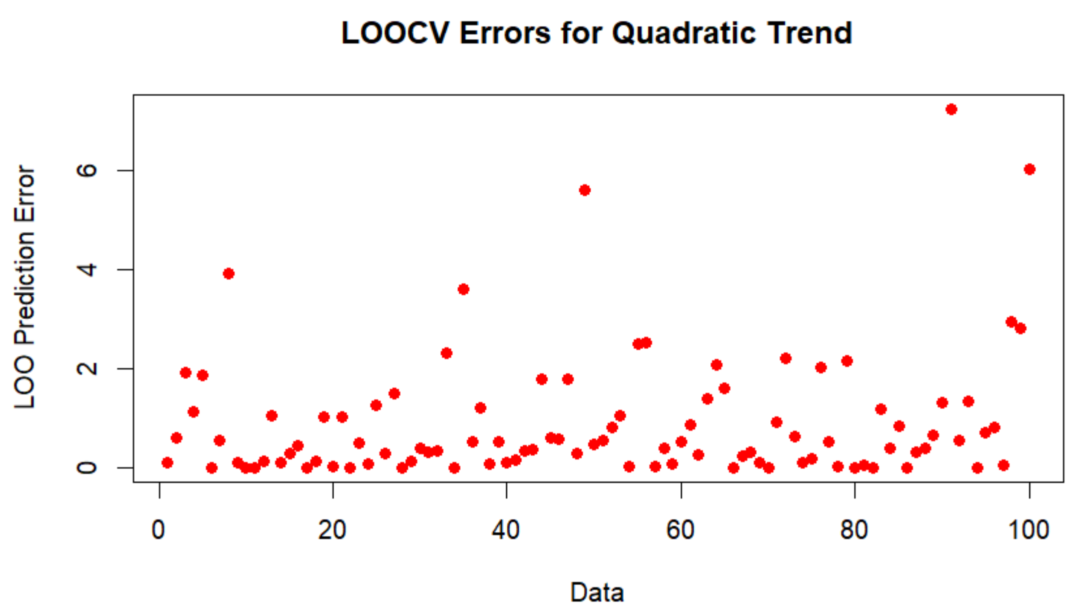


FIGURE 8. GP model with quadratic trend

Acknowledgments

I am thankful to Parkes Kendrick for the fruitful discussion about the exercises and implementation.

References

[1] P. Diggle and P. Ribeiro. *Model-based Geostatistics.* Springer Series in Statistics. Springer, Mar. 2007.