Duke Statistical Science

Hierarchical modeling of annual rainfall data with spatial covariates

STA 702 Fa24 Course Project

Arijit Dey

Department of Statistical Science

Duke University

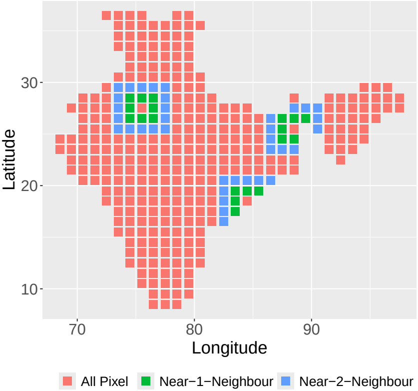
November 26, 2024

Data Set

- Y_{ij}; $i = 1901, \ldots, 2022$ and $j = 1, \ldots, 357$. Annual rainfall at j-th location in i-th year.

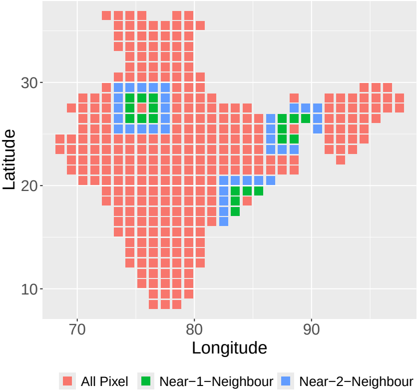
Data Set

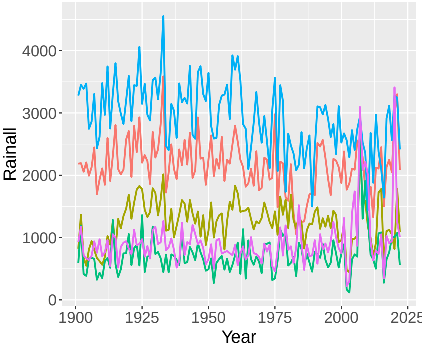
- Y_{ij}; $i = 1901, \ldots, 2022$ and $j = 1, \ldots, 357$. Annual rainfall at j-th location in i-th year.



Data Set

- Y_{ij}; $i = 1901, \ldots, 2022$ and $j = 1, \ldots, 357$. Annual rainfall at j-th location in i-th year.

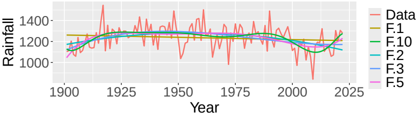




- Hard to model as time series.

Initial Models

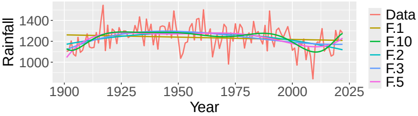
- Simple Linear Regression



F.i is the fitted regression line with time variable taken up to i−th order

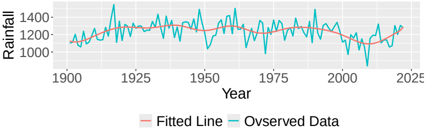
Initial Models

Simple Linear Regression



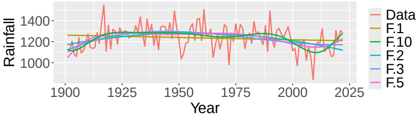
F.i is the fitted regression line with time variable taken up to i−th order

Basis Spline Regression



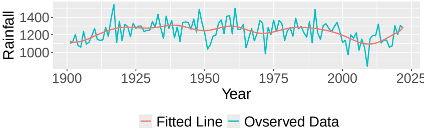
Initial Models

Simple Linear Regression

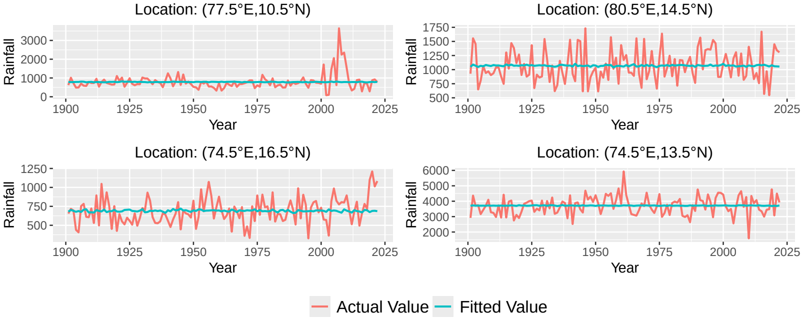


F.i is the fitted regression line with time variable taken up to i–th order

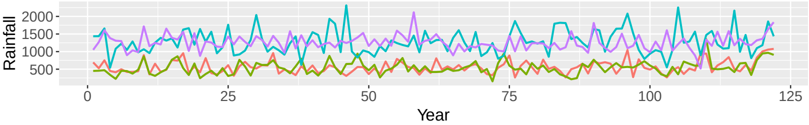
Basis Spline Regression



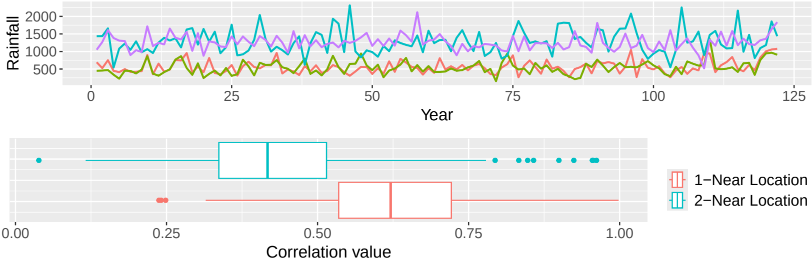
Gaussian Process



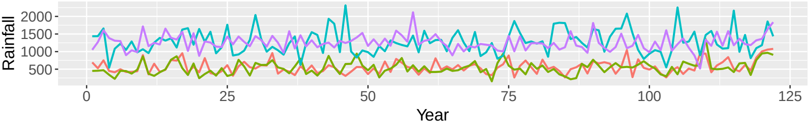
EDA



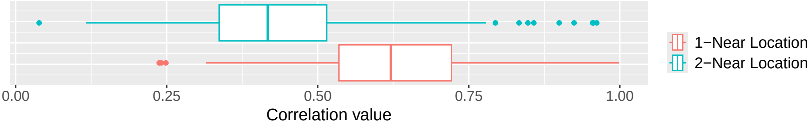
EDA


Rainfall
2000
1500
1000
500
0
25
50
75
100
125
Year
0.00
0.25
0.50
0.75
1.00
Correlation value
1−Near Location
2−Near Location

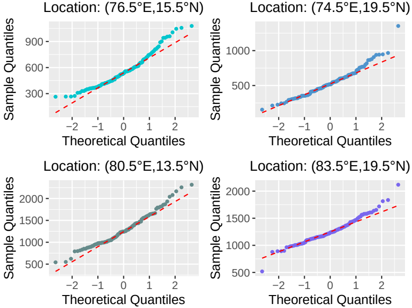
EDA



Rainfall
2000
1500
1000
500
0
25
50
75
100
125
Year

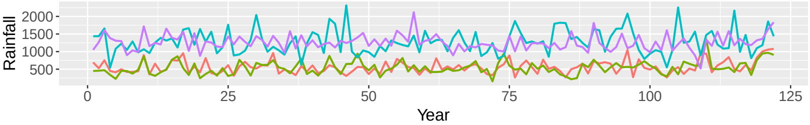


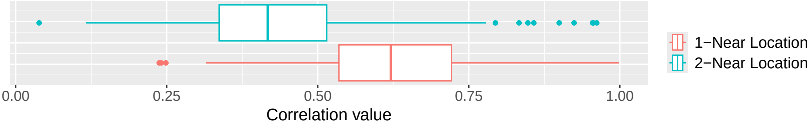
1−Near Location
2−Near Location
0.00
0.25
0.50
0.75
1.00
Correlation value

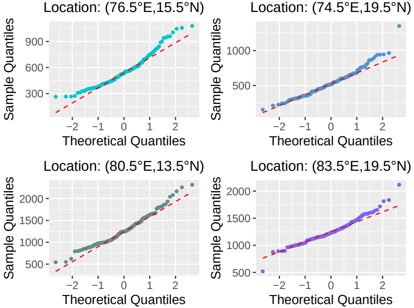


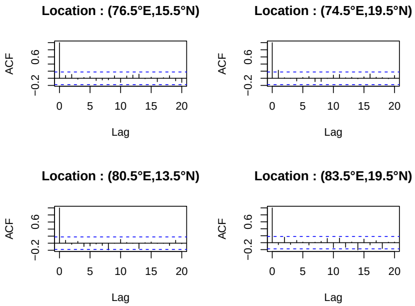
Location: (76.5°E,15.5°N)
Location: (74.5°E,19.5°N)
Location: (80.5°E,13.5°N)
Location: (83.5°E,19.5°N)
Sample Quantiles
Theoretical Quantiles

EDA


Rainfall
2000
1500
1000
500
0
25
50
75
100
125
Year


1−Near Location
2−Near Location
0.00
0.25
0.50
0.75
1.00
Correlation value


Location: (76.5°E,15.5°N)
Location: (74.5°E,19.5°N)
Location: (80.5°E,13.5°N)
Location: (83.5°E,19.5°N)
Sample Quantiles
Theoretical Quantiles


Location : (76.5°E,15.5°N)
Location : (74.5°E,19.5°N)
Location : (80.5°E,13.5°N)
Location : (83.5°E,19.5°N)
ACF
Lag

Model Description

- **Model:**

For $i = 1, \ldots, T$, and $j = 1, \ldots, S$,

$$Y_{ij} = \beta_{j1}\,\bar{Y}_{ij,1} + \beta_{j2}\,\bar{Y}_{ij,2} + \beta_{j3}\,\bar{Y}_{i,1} + \beta_{j4}\,\bar{Y}_{i,2} + \epsilon_{ij}.$$

$$\Rightarrow Y_{ij} = \boldsymbol{x}_j^T \boldsymbol{\beta}_j + \epsilon_{ij}, \text{ where } \epsilon_{ij} \overset{\text{iid}}{\sim} N(0, \sigma^2).$$

Here $\boldsymbol{x}_j^T = [\bar{Y}_{ij,1}, \bar{Y}_{ij,2}, \bar{Y}_{i,1}, \bar{Y}_{i,2}]'$ and $\boldsymbol{\beta}_j = [\beta_{j1}, \ldots, \beta_{j4}]'$.

$$\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_S \overset{\text{iid}}{\sim} N_4(\boldsymbol{\theta}, \Sigma)$$

Model Description

- **Model:**

  For $i = 1, \ldots, T$, and $j = 1, \ldots, S$,

$$Y_{ij} = \beta_{j1}\bar{Y}_{ij,1} + \beta_{j2}\bar{Y}_{ij,2} + \beta_{j3}\bar{Y}_{i,1} + \beta_{j4}\bar{Y}_{i,2} + \epsilon_{ij}.$$

$$\Rightarrow Y_{ij} = \boldsymbol{x}_j^T\boldsymbol{\beta}_j + \epsilon_{ij}, \text{ where } \epsilon_{ij} \overset{\text{iid}}{\sim} N(0, \sigma^2).$$

  Here $\boldsymbol{x}_j^T = [\bar{Y}_{ij,1}, \bar{Y}_{ij,2}, \bar{Y}_{i,1}, \bar{Y}_{i,2}]'$ and $\boldsymbol{\beta}_j = [\beta_{j1}, \ldots, \beta_{j4}]'$.

$$\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_S \overset{\text{iid}}{\sim} N_4(\boldsymbol{\theta}, \Sigma)$$

- **Description:**

  $\bar{Y}_{ij,1}$: Mean of Near-1-Neighbor

  $\bar{Y}_{ij,2}$: Mean of Near-2-Neighbor

  $\bar{Y}_{i,1} = \sum_{j=1}^{S} \bar{Y}_{ij,1}$, and $\bar{Y}_{i,2} = \sum_{j=1}^{S} \bar{Y}_{ij,2}$

Model Description

- **Model:**

For $i = 1, \ldots, T$, and $j = 1, \ldots, S$,

$$Y_{ij} = \beta_{j1}\,\bar{Y}_{ij,1} + \beta_{j2}\,\bar{Y}_{ij,2} + \beta_{j3}\,\bar{Y}_{i,1} + \beta_{j4}\,\bar{Y}_{i,2} + \epsilon_{ij}.$$

$$\Rightarrow Y_{ij} = \boldsymbol{x}_j^T \boldsymbol{\beta}_j + \epsilon_{ij}, \text{ where } \epsilon_{ij} \overset{\text{iid}}{\sim} N(0, \sigma^2).$$

Here $\boldsymbol{x}_j^T = [\bar{Y}_{ij,1}, \bar{Y}_{ij,2}, \bar{Y}_{i,1}, \bar{Y}_{i,2}]'$ and $\boldsymbol{\beta}_j = [\beta_{j1}, \ldots, \beta_{j4}]'$.

$$\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_S \overset{\text{iid}}{\sim} N_4(\boldsymbol{\theta}, \Sigma)$$

- **Description:**

$\bar{Y}_{ij,1}$: Mean of Near-1-Neighbor

$\bar{Y}_{ij,2}$: Mean of Near-2-Neighbor

$$\bar{Y}_{i,1} = \sum_{j=1}^{S} \bar{Y}_{ij,1}, \text{ and } \bar{Y}_{i,2} = \sum_{j=1}^{S} \bar{Y}_{ij,2}$$

- **Prior:**

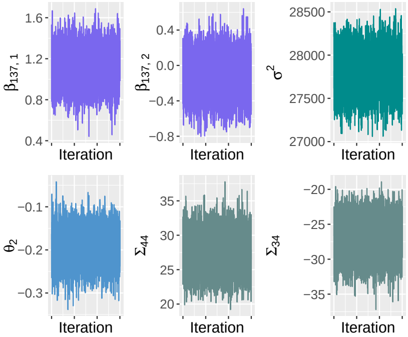
$$\boldsymbol{\theta} \sim N(\boldsymbol{\mu}_0, \boldsymbol{\Lambda}_0)$$

$$\Sigma \sim \text{Inv-Wish}(\eta_0, \boldsymbol{S}_0)$$

$$\sigma^2 \sim \text{Inv-Gam}\left(\frac{\nu_0}{2}, \frac{\nu_0 \sigma_0^2}{2}\right)$$

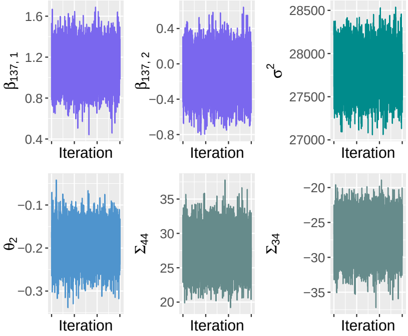
MCMC Diagnostistics

- Traceplot:

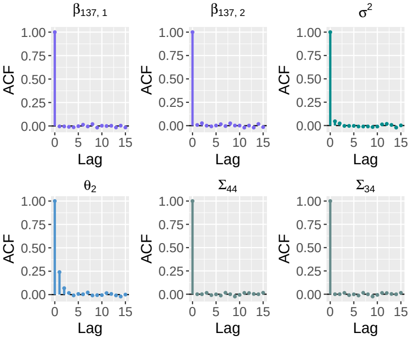


MCMC Diagnostistics

- Traceplot:

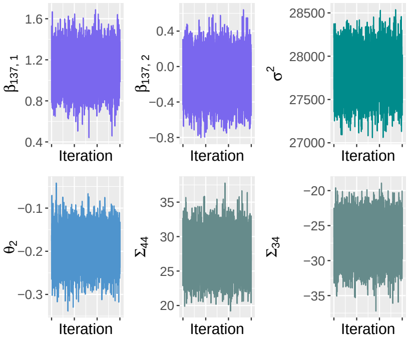


- ACF Plot:

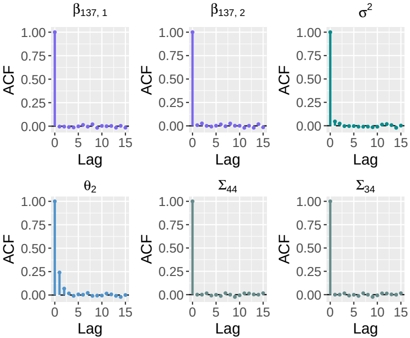


MCMC Diagnostistics

- Traceplot:



- ACF Plot:



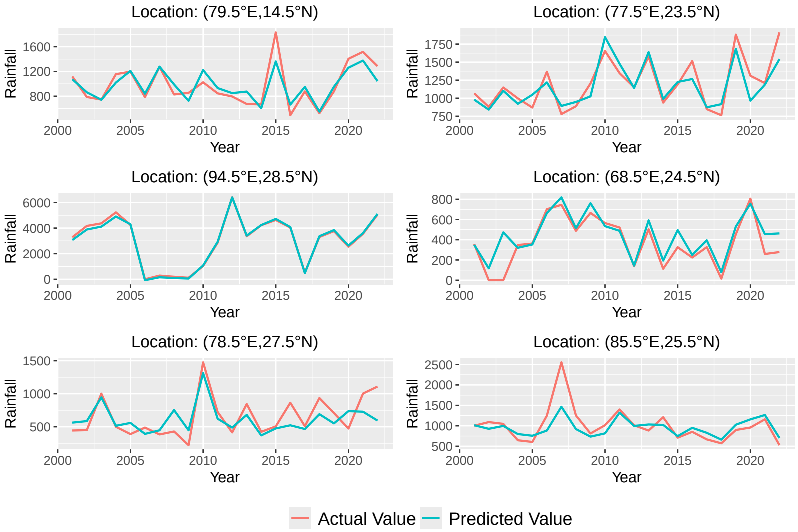
- ESS:

$\text{ESS}(\sigma^2) = 5249$

$\text{ESS}(\theta) = 3255$

	Min	Q1	Mean	Q3	Max
β	5010	5775	5938	6000	6000
Σ	5715	5719	5912	6000	6000

Prediction


Location: (79.5°E,14.5°N)
Rainfall
1600
1200
800
2000 2005 2010 2015 2020
Year
Location: (77.5°E,23.5°N)
Rainfall
1750
1500
1250
1000
750
2000 2005 2010 2015 2020
Year
Location: (94.5°E,28.5°N)
Rainfall
6000
4000
2000
0
2000 2005 2010 2015 2020
Year
Location: (68.5°E,24.5°N)
Rainfall
800
600
400
200
0
2000 2005 2010 2015 2020
Year
Location: (78.5°E,27.5°N)
Rainfall
1500
1000
500
2000 2005 2010 2015 2020
Year
Location: (85.5°E,25.5°N)
Rainfall
2500
2000
1500
1000
500
2000 2005 2010 2015 2020
Year
Actual Value
Predicted Value