Cairo University

Faculty of Economics and Political Science

English Section

Statistics Department

3rd Year

Spatial Analysis of Temperature Patterns in California

Submitted by:

Saif Essam Abdelmaboud (5200356)

Youssef Hatem Mohamed (5200836)

Yehia Mohamed Quolquela (5200901)

Under The Supervision of:

Dr. Sara Osama

Table of Contents

➢ Introduction:

- Climate change has become a critical issue in the modern world, with temperatures rising at an unprecedented rate due to human activities. In this context, it is essential to understand the spatial patterns and variability of temperature in different regions, as well as the factors that drive temperature changes.
- California is a state that has seen a significant increase in temperature over the last few decades, making it an ideal location for studying the spatial patterns of temperature. In this report, we conduct a complete spatial data analysis of temperature values in California using real temperature data from California state. Our analysis aims to identify the spatial patterns and variability of temperature in California, as well as the factors that drive these patterns.

➢ Objectives:

1. To conduct an exploratory data analysis of the temperature values in California using the real temperature data, including examining the distribution of the data, outliers and spatial autocorrelation.
2. To assess the spatial autocorrelation in the temperature values, which means, the degree to which nearby locations have similar temperature values, using spatial autocorrelation statistics such as Moran's I or Geary's C.
3. To cluster the temperature values based on their spatial dependency and identify groups of locations with similar temperature patterns.
4. To interpolate the temperature values at unsampled locations using spatial interpolation techniques such as inverse distance weighting and kriging.
5. To fit a statistical model to the temperature values to identify the factors that influence temperature variation, using spatial regression techniques such as geographically weighted regression.

We aim to gain insights into the spatial patterns and variability of temperature in California, as well as the factors that drive these patterns. This knowledge can be valuable for policymakers and researchers working to mitigate the impacts of climate change in California and other regions.

➢ Data Analysis:

1. Data Description:

i. Checking the Normality Assumption:

- Figure (1): From the histogram it appears that it is asymmetric distribution and the data are skewed to the right indicating that the normality assumption does not hold and it is indication of a potential outliers.

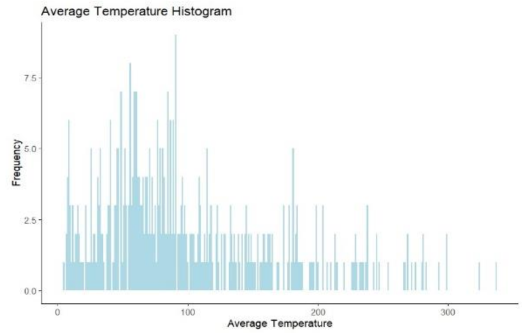


Figure (1): Histogram of the data

ii. Identifying Outliers:

- Figure (2): The box plot reveals the presence of outliers in the temperature data. Outliers are observations that lie significantly outside the whiskers of the box plot, indicating extreme values that are distinct from the majority of the data.

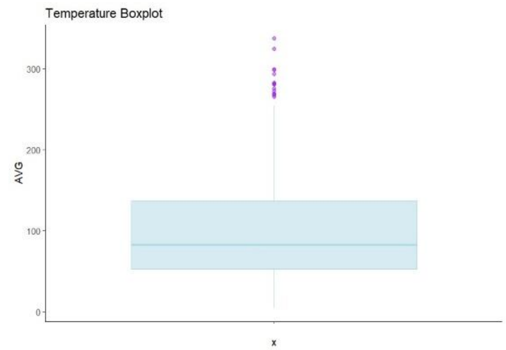


Figure (2): Temperature Box Plot

iii. Exploring Relationships:

- Figure (3): Shows that the difference in values between the top and bottom of the bubble plot suggests a vertical pattern or gradient in the temperature data. This means that as you move from the bottom to the top of the plot, there is a systematic increase in temperature values.

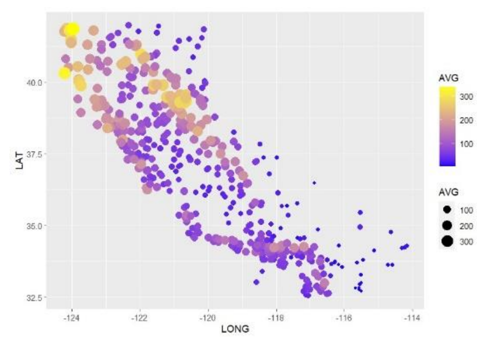


Figure (3): Bubble Plot

2. Measuring the Spatial Dependency and Autocorrelation:

- First of all, we have to check if there is an autocorrelation or spatial dependency and to understand if there is a spatial pattern of similarity or dissimilarity in the temperature across California.
- One of the methods to check the autocorrelation is Moran's I statistic.

- **<u>Moran's I:</u>**
  - Moran's I statistic is a commonly used measure in spatial statistics to assess the presence and strength of spatial autocorrelation in a dataset. The goal is to uncover any significant spatial patterns and provide insights into the spatial dependence of temperature measurements.

Upon calculating Moran's, I for the temperature dataset, we obtained the following results:

- Observed Moran's I: 0.2548929
- Expected Moran's I under the assumption of spatial randomness: -0.002197802
- Standard deviation (SD) of Moran's I: 0.005153807
- p-value is equal 0.

Since the (observed Moran's I = 0.2548929) is greater than the (Expected Moran's I = -0.002197802) and the p-value = 0 is less than 0.05, therefore this indicates **statistical significance weak positive autocorrelation** in the temperature data.

This suggests that locations with similar temperature values tend to be spatially clustered, while locations with dissimilar values are also clustered together.

- **<u>Moran's I plot:</u>**

In the plot, positive values of Moran's I are shown in the upper-right (High-High) quadrant, indicating clusters of locations with similar temperature values (Hot Spot). While, in the lower-left quadrant (Low-Low), representing clusters of locations with similar temperature values (Cold Spot). The horizontal line at Moran's I = 0 indicates the threshold for spatial randomness.

<u>Figure (4):</u> Shows that there are clusters in the high-high quadrant and in the low-low quadrant which indicates that there is hot spot and cold spot. Moreover, the values in the high-low and low-high quadrant are considered outliers.

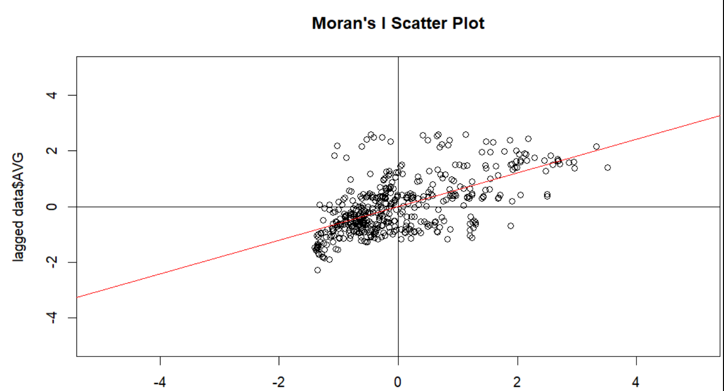


Figure (4): Moran's I Plot

- **LISA plot:**

- The LISA plot provides insights into the spatial clustering patterns and identifies local areas of significance in terms of similarity or dissimilarity in the variable of interest.
- Figure (5): Shows a bubble-plot of observations against spatial coordinates. Above mean values are signified by red circles. While below mean values are signified by black squares. Thus, spatial hot-spots are represented by red filled circles and cold-spots by black filled squares.

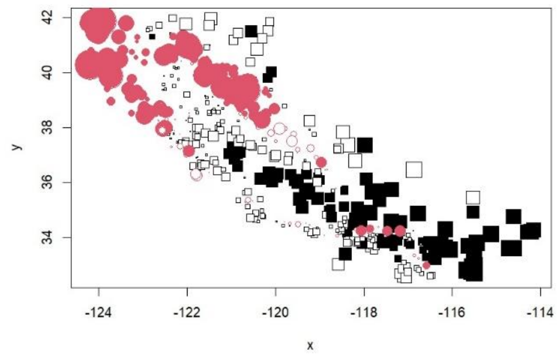


Figure (5): LISA Plot

- Therefore, as we go from downward to upward, the values of the temperature become higher. And in the upper-left quadrant there is a hot spot, while in the lower-right quadrant there is a cold spot. Additionally, we have a high low outlier in the upper right quadrant, also we have low-high outliers in the lower-right quadrant.

3. Trend Surface Model:

- The trend surface model is a widely used technique in spatial analysis and geo-statistics that aims to capture and describe the underlying spatial trends or patterns in a dataset. It is particularly useful when dealing with continuous variables, such as temperature, where there is an expectation of smooth and gradual changes across the study area. The trend surface model allows us to estimate a surface or function that best represents the overall trend in the data, providing valuable insights into spatial patterns and variations.
- Moreover, it is a regression model where the explanatory variables are the spatial coordinates (longitude, latitude), and the response variable is Z and It follows the same assumptions of regression models and it can be used as interpolation tool but in certain condition such that there is very low value of autocorrelation and it depends on the relationship between x and y.

First, we need to make 3D plot of the data as the 3D plots can figure out the type of the relation.

- Figure (6): We can see from this figure that there are many curves, so we have to check the model that best fits our data. However, we have to check all the probable models.

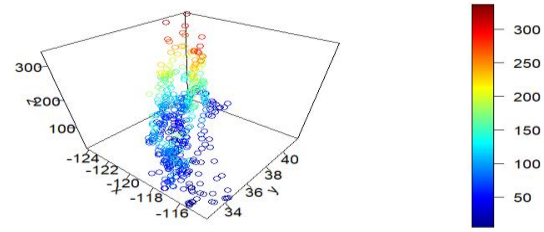


Figure (6): 3D Scatter Plot

We will choose according to Akaike Information Criteria (AIC) rather than the R-squared as the AIC indicates the trade-off between model complexity and goodness of fit, with lower values indicating a better balance between the two. The AIC increases by increasing number of parameters and considers the problem of losing the degrees of freedom and it tells us the difference between the two models is not worth adding parameters so, we will go forward to the fourth order trend surface model as it has lowest AIC.

- **The Fourth Order Trend Surface Model:**

- Figure (7): Shows the fourth order trend surface model which indicates continuous surface that approximates the temperature distribution across the study area. It allows us to visualize the overall trend and identify regions of higher or lower temperatures. Additionally, the model enables the estimation of temperature values at unobserved locations based on the spatial relationships captured in the data.

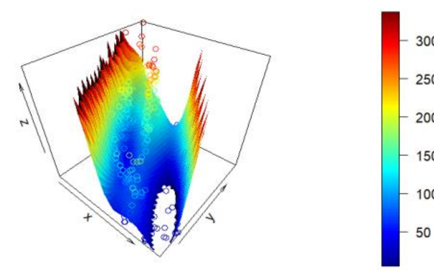


Figure (7): 4th Order Trend Surface Model Plot

Other Model fits are in the Appendix.

- **Contour Lines:**

- In the context of the temperature analysis in California, contour lines provide a visual representation of the temperature values at different locations, highlighting areas of higher and lower temperatures. By plotting contour lines, we can observe spatial patterns, identify temperature gradients, and gain insights into the overall temperature landscape.

- Figure (8): Shows the contour lines generated from the trend surface model provide a clear depiction of the spatial distribution of temperature in California. Each contour line represents a specific temperature value, connecting locations with similar temperature readings. The contour lines can be interpreted as isolines, separating areas with different temperature ranges.

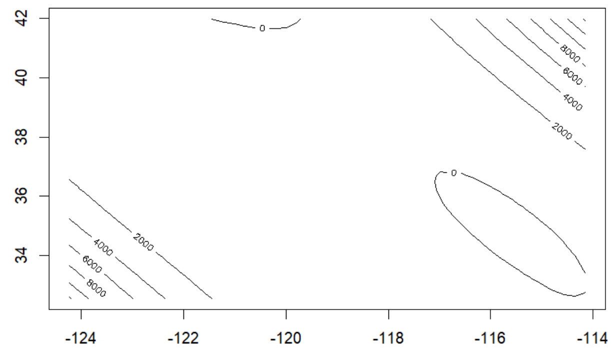


Figure (8): Contour Lines

- We can exactly specify the values lying on the contour lines such as the points with values of 2000/4000/6000/8000/10000, but we cannot do the same for the points lying between two contours. The range between the contour lines are constant = 2000.
- Since the area between contour is large, this indicates having more variables.

4. Spatial Interpolation:

1) Inverse Distance Weighting:

- It is one of the most common methods in spatial interpolation, it is based on the idea that the interpolating area should be influenced by the closer stations than the distant ones.
- We will use IDW to give us an indicator of the dependency structure of the stations recording air quality in California.
- We will use the most common form of IDW:

$$\hat{z}(s_0) = \frac{\sum_{i=1}^{n} z(s_i) d_{i0}^{-p}}{\sum_{i=1}^{n} d_{i0}^{-p}}$$

  - Where $\hat{Z}(S_0)$ is the unsampled locations which we want to interpolate. We will use a grid as an interpolating surface.
  - $Z(S_i)$: is the sampled air quality at location "i" in California.
  - d_{i0}: is the distance between the interpolating surface (unsampled point on the grid) and the sampled locations.
  - P: the power of the distance, it will be determined using the MSE.

We will get the interpolated values of the last 56 observations and compare to the to their true sampled value using Mean Square Errors (MSE). We will compute MSE for five IDW models whose power of the distance 'P' ranges from 1 to 5.

- <u>Figure (9):</u> Show the MSE for IDW models from p=1 to p=5. The figure shows that the IDW model with p =5 has the lowest MSE, so it is the best IDW model for interpolation.

MSE_IDW1	656.477289389667
MSE_IDW2	531.099127955837
MSE_IDW3	442.730255437393
MSE_IDW4	418.492396008508
MSE_IDW5	414.951592302547

Figure (9): R Output for the MSE for all p=1, ..., 5

<u>Note that:</u> we tried IDW models with P > 5 & their MSE was greater than that of IDW model with p =5.

- **<u>Predicted temperature of California using IDW:</u>**

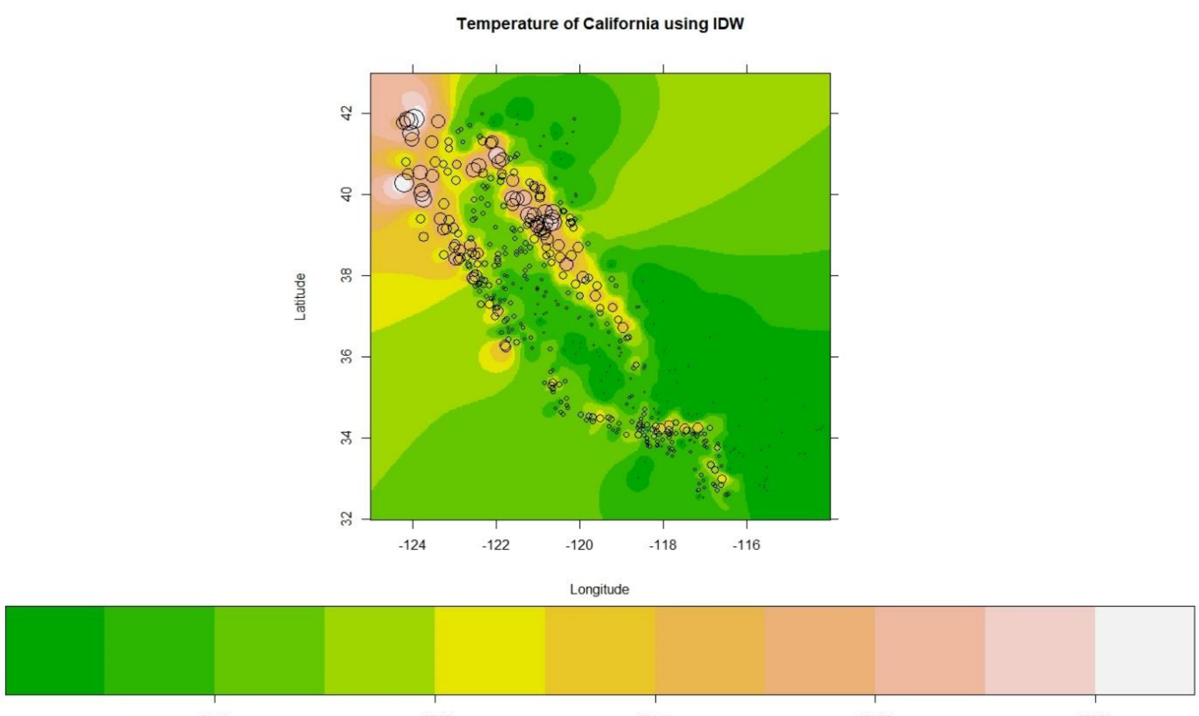


Figure (10): Temperature of California Using IDW

- <u>Figure (10):</u> Shows that the temperature is lower in the south of California compared to the north especially north west of California. The temperature in the southern parts of California is generally around 70. There is a small cluster in south with higher temperature than its surrounding points (around140) at (-120, -117) longitude& 34 latitude.

- It is the hottest in the north west of California where it reaches temperatures around 337. The points on the graph are the values of the temperature of the sampled locations (the bigger the points the higher the value of the temperature).
- We can notice positive correlation between the values of the sampled locations and the predicted values, this indicates low prediction error at the sampled points.

2) Kriging:

- Kriging is a geostatistical interpolation technique commonly used in spatial analysis to estimate values at unobserved locations based on the values observed at nearby locations. It is particularly useful when dealing with spatially correlated data, such as temperature measurements, where the values at neighboring locations tend to be more similar than those farther apart. Kriging takes into account the spatial autocorrelation of the data to provide a spatially continuous estimate that captures the underlying trends and patterns in the variable of interest.0
- It is an optimal linear estimator given by the following formula:

$$\hat{Z}(s_0) = \sum_{i=1}^{N} \lambda_i Z(s_i)$$

  - Where S_0: the prediction location.
  - λ_i: weights for the sampled values at the i^{th} location.
  - S_i: the sampled value at the i^{th} location.
  - N: number of the sampled location.
- Kriging is a better spatial interpolator than IDW and trend surface model, as kriging takes into consideration the dependency or the spatial autocorrelation between the sampled location while IDW is a mathematical formula and trend surface model is a regression model (regression assumptions states no autocorrelation).
- Kriging is a 2-step process: first, we want to determine the covariance structure of the sampled location by fitting a suitable variogram; and second, we estimate the weights λ_i using the fitted variogram which are used to spatial interpolation.

- **Step 1: Fitting a Suitable Variogram:**

First, we need to fit and plot the **empirical semi-variogram:**

- Figure (11): Shows the fitted empirical variogram.
- However, we cannot use empirical variogram as estimate for variogram as it does not satisfy the condition of positive definite covariance and negative definite variogram so the variance will be with negative value thus we have to fit it using one of the parametric variograms.

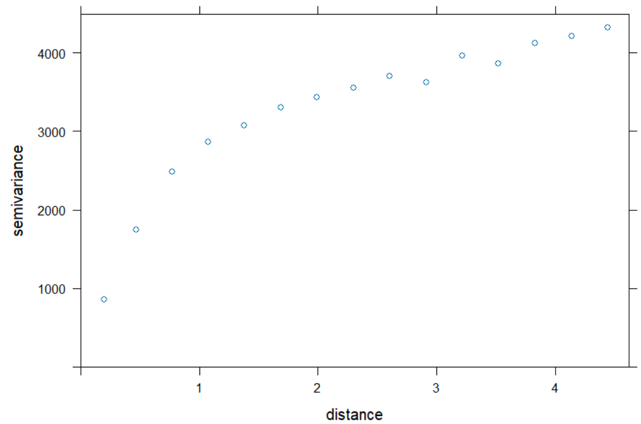


Figure (11): The Empirical Semi-Variogram

Second, we have to **choose one of the parametric semi-variograms:**

- According to the plots we found that **exponential variogram is best fit to our data**.
- And according to **Error Sum of Squares the exponential semi-variogram is the lowest so it is best.**
- Figure (12): Shows the exponential semi-variogram model which is represented by a smooth curve that best fits these empirical values indicating that the model captures the spatial dependence and variability of the temperature data accurately. And with a value of ESS = 237956381 which is the lowest among the other parametric semi-variogram models.

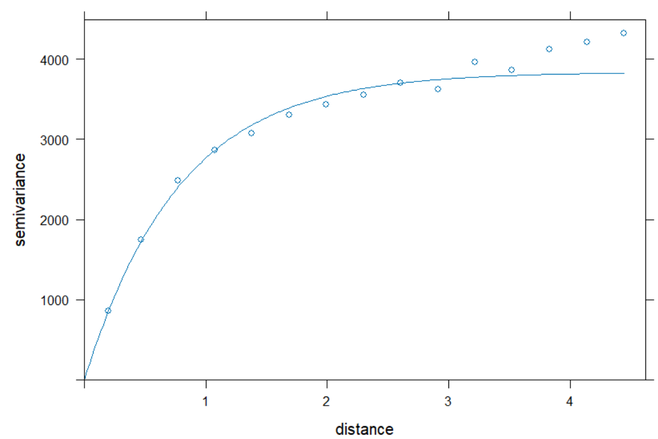


Figure (12): Exponential Semi-Variogram Plot

Then, we have to **estimate the semi-variogram:**

- **The Exponential Model:**

$$\gamma o = c_o + c_1\,[1 - \exp(-3h/r)]$$

- **<u>The Estimated one by using weighted least squares:</u>**

$$\widehat{\delta o} = 0 + 3838.989[\ 1\text{- exp} (\text{-}3h / 0.7846972)$$

- **<u>The Interpretations:</u>**
  - $\hat{C}$ (0): The variation between two points tends to be 0 as the distance tends to zero.
  - $\hat{C}$ (1): Partial sill =3838.989
  - Sill: The maximum variation is 3838.989+0 = 3838.989
  - $\hat{R}$: The distance at which the variogram is maximized or the distance at which the maximum amount of variation reached is equal to 0.7846972.

<u>Step 2: Estimating the weights:</u>

- *This step is done automatically using R program.*
- With the weights calculated, we can generate a surface of predicted temperature values across California. Additionally, we can derive a variance map that represents the uncertainty associated with these predicted values.
- The predicted values of temperature obtained through ordinary kriging represent our best estimates of temperature at each location in California. These estimates incorporate both the observed values and the spatial relationships between the data points, resulting in a continuous surface that captures the underlying temperature patterns.
- Also, by analyzing the predicted values and variance map of ordinary kriging, we gain insights into the spatial distribution of temperature across California. We can identify areas with high or low predicted temperatures and assess the level of uncertainty associated with these predictions. This information is crucial for decision-making processes that rely on accurate temperature estimates and the associated uncertainty.

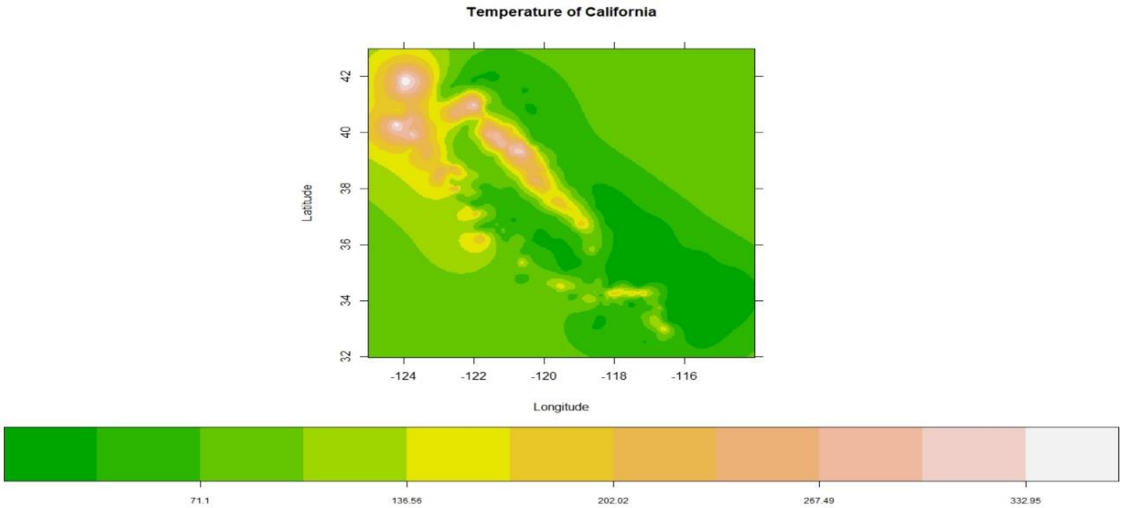


Figure (13): Predicted Values of Temperature in California.

- <u>Figure (13):</u> We can see from this figure that the temperature in the south east of California is less than 71 except for some parts which are around 140. The temperature keeps increasing as we go from southeast to northwest till it reaches values around 267, except for some locations where it reaches a temperature of more than 333. In the northeast & southwest, the temperature is quite similar (between 71 & 136).

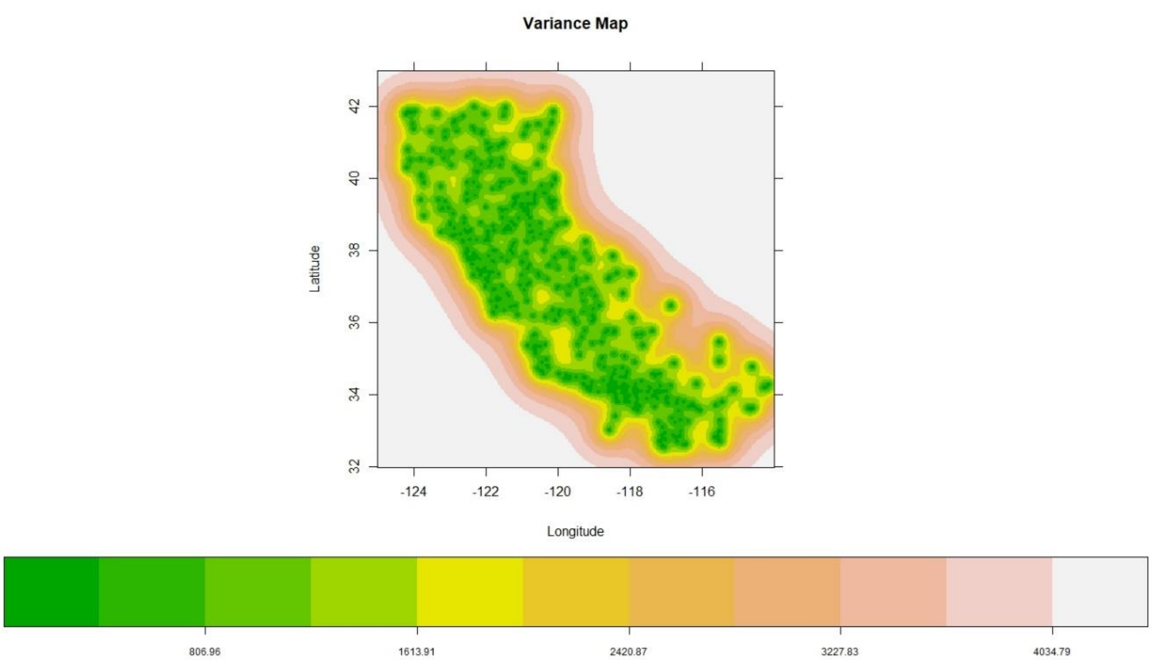


Figure (14): Variance Map of the Predicted Values.

- Figure (14): We can see from this figure that the variance of the predicted temperature is very high in the northeast and the southwest. This indicates the prediction in both regions is highly inaccurate. We can mitigate this issue in the future by putting weather stations in the northeast and the southwest of California thus improving the prediction. Along the diagonal from the northwest to the southeast the prediction error is relatively low, hence the prediction along this diagonal is accurate and reliable.

5. Geographical Weighted Regression Model (GWRM):

- GWR is a non-stationary technique that models spatially varying relationships. Compared with a basic (global) regression, the coefficients in GWR are functions of spatial location.

- **Step 1: Fitting Global Regression Model:**

- Global Model is:

$$Y_i = \beta_0 + \beta_1 \text{ long} + \beta_2 \text{lat}$$

$$Y_i = -1521.238 - 11.735 \text{ long} + 5.746 \text{lat}$$

β_1: By increasing longitude by 1 so on average temperature decreases by 11.75 holding latitude constant.

β_2: By increasing latitude by 1 so on average temperature increases by 5.746 holding longitude constant.

β_0: By holding longitude and latitude constant the temperature on average will be -1521.238.

- **Step 2: Obtain the Residuals and Plot of it:**

- Figure (15): Shows that at south we found that there is a group of residuals with high values and at north in the center we found also points of residuals with high values and at north west also so we concluded that the global model predicts it badly, so we need to go Forword geographical to get models to this are and predict it.

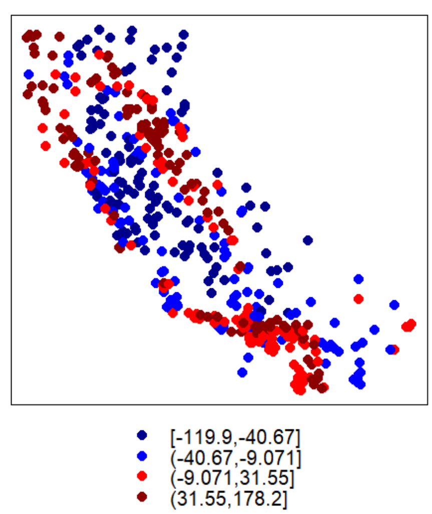


Figure (15): Geographical Weighted Regression Plot

- **<u>Step 3: Obtain Geographical Weighted Regression Model:</u>**

$$y_i = \beta_{i0} + \sum_{k=1}^{m} \beta_{ik} x_{ik} + \varepsilon_i$$

- We need to estimate parameter at location I and kth independent variable. So, the parameter of set of regression coefficients is estimated by weighted least squares.

$$\hat{\beta}_i = \left(X^T W_i X\right)^{-1} X^T W_i y$$

- The weighting scheme Wi is calculated with a gaussian kernel function based on the proximities between regression point i and the N data points around it.
- where d_{ij} is the distance between observation point j and regression point i, and b is the kernel bandwidth.
- So, we get kernel bandwidth using adaptive bandwidth according to cross-validation.
- We found that b = 0.004464932.

$\beta_{0i:}$ Holding longitude and latitude constant the temperature on average will be ranges between -95675.044 to 50391.863 for all locations.

β_{1i}: By increasing longitude by 1 so on average temperature will ranges from -378.877 to 552.336 for all locations holding latitude is constant.

$\beta_{2i:}$ By increasing latitude by 1 so on average temperature will ranges from -607.398 to -11.7354 for all locations holding longitude constant.

➢ Conclusion:

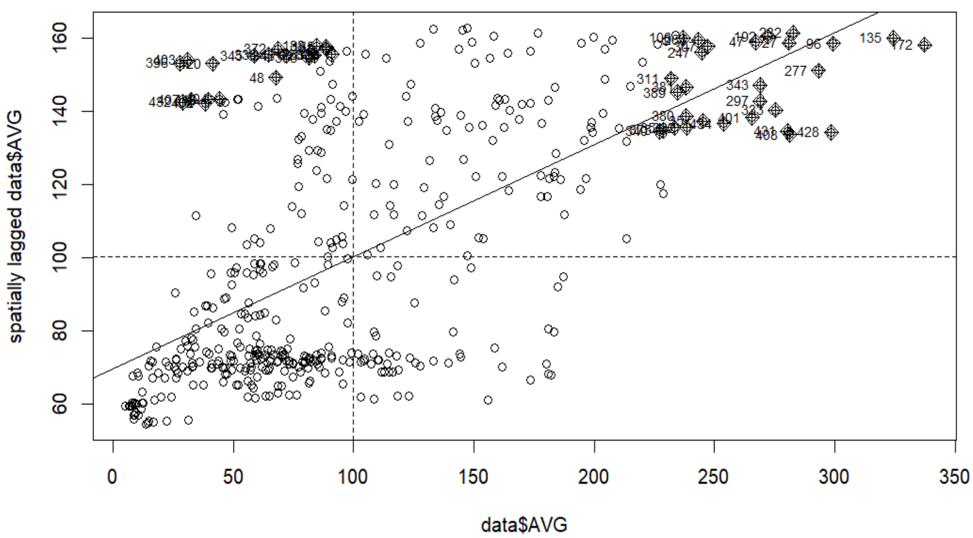
- In conclusion, this report conducted a comprehensive spatial data analysis of temperature values in California using various statistical techniques and methods. The main objectives of the analysis were to understand the spatial patterns, explore trends, and estimate temperature values across the region.
- We have begun with some data description by checking the normality assumption and outliers, with visual representation of the spatial distribution of temperature values.
- Then, checking the spatial autocorrelation using Moran's I indicating positive autocorrelation and visualizing the clustering pattern, showing significant spatial clusters of similar temperature values.
- After that, Trend surface modeling was employed to identify trends and spatial patterns in the temperature data. The fourth-degree trend surface model, with the lowest AIC, was selected as the best-fit model, indicating a smooth variation in temperature across the study area.
- Also, spatial interpolation techniques, such as kriging, were utilized to estimate temperature values at unmeasured locations. The variogram analysis aided in selecting an appropriate variogram model, with the exponential variogram demonstrating the best fit to the data.
- And then, the geographical weighted regression modeling captured spatial heterogeneity by considering the local effects of predictor variables on temperature. The global regression model was initially fitted, followed by the exploration of residuals and the identification of spatially varying relationships using the geographically weighted regression model.
- Overall, this report demonstrates the value of spatial data analysis in understanding the spatial patterns and variations of temperature in California. The findings contribute to the broader knowledge of environmental statistics and provide insights for informed decision-making in spatially dependent phenomena.

Recommendations:

1. Spatial Monitoring: Implement a spatial monitoring system to collect temperature data across California especially in the areas with high variance. This will enhance the understanding of spatial patterns, identify temporal trends, and provide valuable information for climate studies.
2. Localized Interventions: Given the spatial heterogeneity observed in temperature patterns, we may consider implementing localized interventions or adaptation strategies that take into account the specific characteristics of different regions within California especially in the north as it has the highest temperature values. For example, green infrastructure planning to mitigate extreme temperature conditions in the north.
3. Climate Change Resilience: By understanding the spatial variations in temperature patterns can help identify areas that are exposed to increasing temperatures, especially when we go from south to north, and guide the development of adaptation strategies to mitigate potential impacts on ecosystems, agriculture, and human health.
4. Further Research: We have to encourage further research to investigate the underlying factors contributing to the observed spatial patterns of temperature. This could include exploring additional variables such as elevation, proximity to water bodies, and atmospheric conditions.

➢ Appendix:

○ Zoomed Moran's I plot:



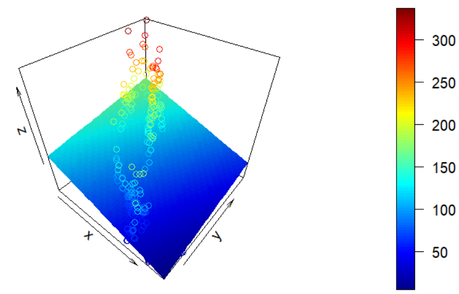
▪ Trend Surface Models:

Summary of the fourth model:

```
> summary(fit.sfc4)
Analysis of Variance Table
 Model: surf.ls(np = 4, x = x, y = y, z = z)
              Sum Sq  Df  Mean Sq F value     Pr(>F)
Regression 1183652.5  14 84546.61 41.9297 < 2.22e-16
Deviation   889227.8 441  2016.39
Total      2072880.3 455
Multiple R-Squared: 0.571,      Adjusted R-squared: 0.5574
AIC: (df = 15) 3484.481
Fitted:
   Min     1Q Median     3Q    Max
-43.66  68.46  83.80 138.12 297.42
Residuals:
     Min       1Q    Median       3Q      Max
-125.092  -28.902   -5.492   24.449  143.094
```

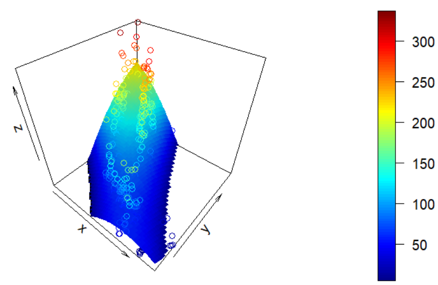
Linear trend surface

```
> summary(fit.sfc1)
Analysis of Variance Table
 Model: surf.ls(np = 1, x = x, y = y, z = z)
              Sum Sq  Df     Mean Sq  F value     Pr(>F)
Regression  701796.9   2 350898.435 115.9353 < 2.22e-16
Deviation  1371083.4 453   3026.674
Total      2072880.3 455
Multiple R-Squared: 0.3386,     Adjusted R-squared: 0.335
AIC: (df = 3) 3657.93
Fitted:
   Min     1Q Median     3Q    Max
 15.44  64.72 109.64 130.39 176.71
Residuals:
     Min       1Q   Median       3Q      Max
-119.890  -40.672   -9.071   31.555  178.224
```



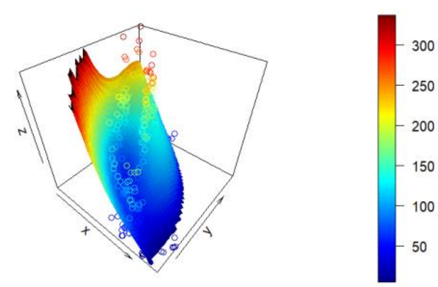
Quadratic trend Surface Model:

```
> summary(fit.sfc2)
Analysis of Variance Table
 Model: surf.ls(np = 2, x = x, y = y, z = z)
              Sum Sq  Df     Mean Sq  F value     Pr(>F)
Regression  813554.2   5 162710.839 58.14211 < 2.22e-16
Deviation  1259326.1 450   2798.502
Total      2072880.3 455
Multiple R-Squared: 0.3925,     Adjusted R-squared: 0.3857
AIC: (df = 6) 3625.159
Fitted:
   Min     1Q Median     3Q    Max
-14.46  64.29  94.42 128.06 223.08
Residuals:
     Min       1Q   Median       3Q      Max
-122.654  -41.077   -4.256   32.716  183.976
```



Cubic trend:

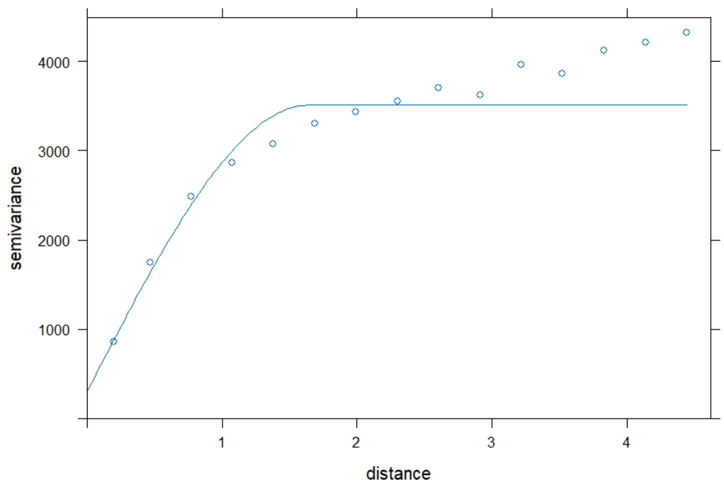
```
> summary(fit.sfc3)
Analysis of Variance Table
 Model: surf.ls(np = 3, x = x, y = y, z = z)
              Sum Sq  Df    Mean Sq  F value     Pr(>F)
Regression  997106.1   9 110789.57 45.93171 < 2.22e-16
Deviation  1075774.1 446   2412.05
Total      2072880.3 455
Multiple R-Squared: 0.481,     Adjusted R-squared: 0.4706
AIC: (df = 10) 3561.323
Fitted:
   Min     1Q Median     3Q    Max
-53.95  63.80  93.52 134.68 242.32
Residuals:
     Min       1Q   Median       3Q      Max
-103.240  -33.080   -7.695   28.756  160.634
```



- **Variogram:**

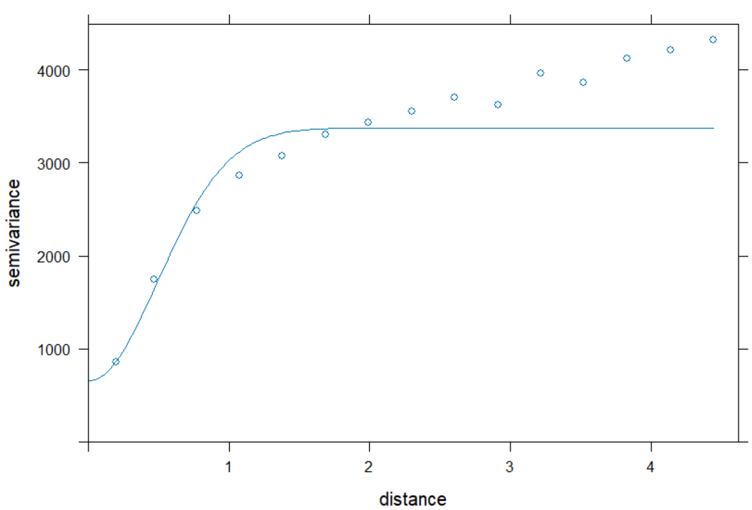
In Choosing the Parametric Semi-Variogram:

- **By using spherical semi-variogram**



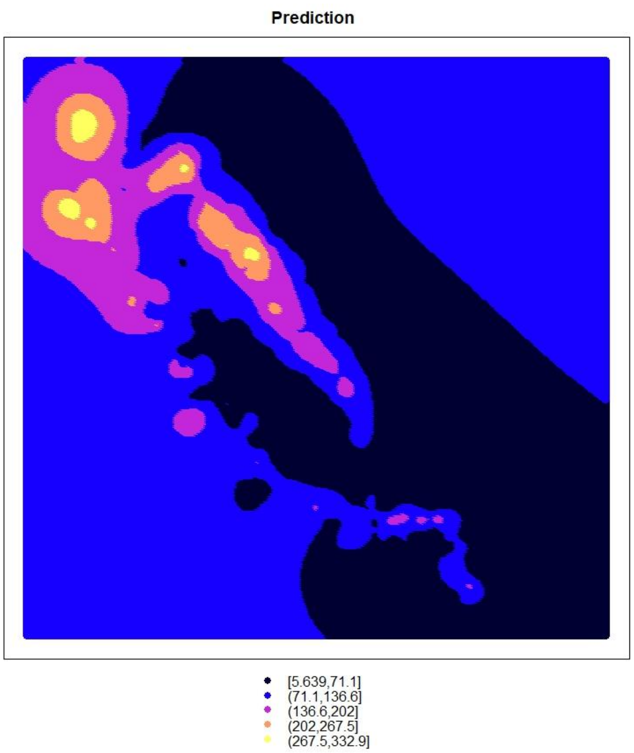
The value of ESS = 1117637743

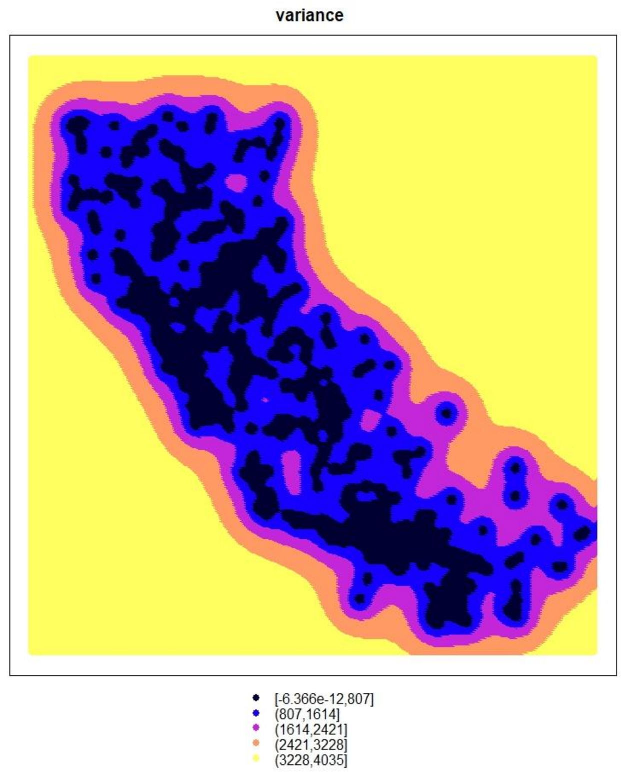
- **By using gaussian semi-variogram**



The value of ESS = 1430944346

▪ Another way to present the kriging and its variance map:





▪ Geographical Weight Regression Model:

○ Fitting Global Regression Model:

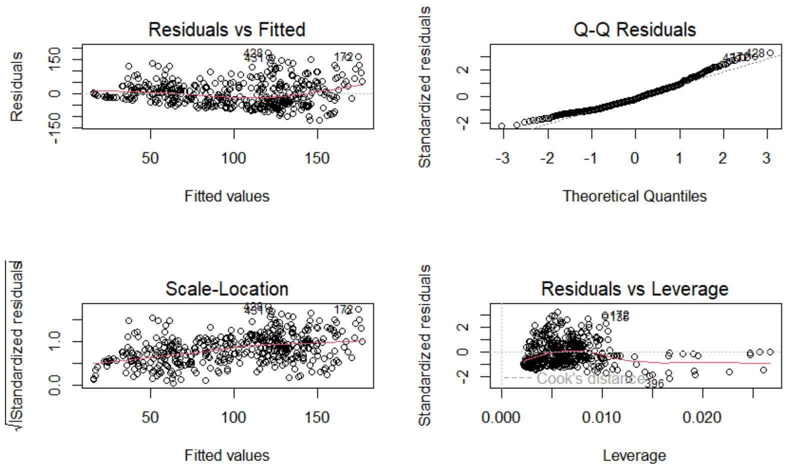
```
Call:
lm(formula = data$AVG ~ data$LONG + data$LAT)

Residuals:
     Min       1Q   Median       3Q      Max 
-119.890  -40.672   -9.071   31.555  178.224 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) -1521.238    196.769  -7.731 6.95e-14 ***
data$LONG     -11.735      2.061  -5.694 2.23e-08 ***
data$LAT        5.746      1.771   3.244  0.00127 ** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 55.02 on 453 degrees of freedom
Multiple R-squared:  0.3386,	Adjusted R-squared:  0.3356 
F-statistic: 115.9 on 2 and 453 DF,  p-value: < 2.2e-16
```

- ## Plots of Model Global:



- ## Getting kernel bandwidth using adaptive bandwidth:

```
> GWRbandwidth <- gwr.sel(data$AVG ~ data$LAT+data$L(
Adaptive q: 0.381966 CV score: 1278201
Adaptive q: 0.618034 CV score: 1343044
Adaptive q: 0.236068 CV score: 1211903
Adaptive q: 0.145898 CV score: 1117949
Adaptive q: 0.09016994 CV score: 1011332
Adaptive q: 0.05572809 CV score: 900683
Adaptive q: 0.03444185 CV score: 763381.2
Adaptive q: 0.02128624 CV score: 635383.6
Adaptive q: 0.01315562 CV score: 523539.7
Adaptive q: 0.008130619 CV score: 413113.4
Adaptive q: 0.005024999 CV score: 319371.7
Adaptive q: 0.00310562 CV score: NA
Adaptive q: 0.00621124 CV score: 357598.4
Adaptive q: 0.004291861 CV score: NA
Adaptive q: 0.005478103 CV score: 333078.3
Adaptive q: 0.004744965 CV score: 311981.8
Adaptive q: 0.004571895 CV score: 308056
Adaptive q: 0.004464932 CV score: 305956
Adaptive q: 0.004398825 CV score: NA
Adaptive q: 0.004505788 CV score: 306724.3
Adaptive q: 0.004464932 CV score: 305956
```

```
> gwr.model
Call:
gwr(formula = data$AVG ~ data$LAT + data$LONG, data = data, coords = cbind(x,
    y), adapt = GWRbandwidth, hatmatrix = TRUE, se.fit = TRUE)
Kernel function: gwr.Gauss
Adaptive quantile: 0.004464932 (about 2 of 456 data points)
Summary of GWR coefficient estimates at data points:
                   Min.    1st Qu.     Median    3rd Qu.       Max.     Global
X.Intercept. -95675.044  -8226.226  -2015.646   3551.770  50391.863 -1521.2384
data.LAT       -378.877    -34.332     19.619     93.630    552.336     5.7456
data.LONG      -607.398    -73.506    -14.443     35.737    409.208   -11.7354
Number of data points: 456
Effective number of parameters (residual: 2traceS - traceS'S): 197.5304
Effective degrees of freedom (residual: 2traceS - traceS'S): 258.4696
Sigma (residual: 2traceS - traceS'S): 23.86579
Effective number of parameters (model: traceS): 149.6486
Effective degrees of freedom (model: traceS): 306.3514
Sigma (model: traceS): 21.92152
Sigma (ML): 17.96794
AICc (GWR p. 61, eq 2.33; p. 96, eq. 4.21): 4379.889
AIC (GWR p. 96, eq. 4.22): 4078.114
Residual sum of squares: 147218.1
Quasi-global R2: 0.928979
```