**Abstract**

This article studies the relationship between spatial attributes (altitude, latitude, and longitude) and mean annual temperature using climate data from 34 weather stations across France. The dataset contains key climate variables, including altitude, latitude, temperature, humidity, and precipitation. We used multiple linear regression to identify significant predictors of mean temperature, excluding non-significant variables. A 3D scatterplot visualization was created to represent the relationship between altitude, latitude, and mean temperature, while plotting a regression plane to further emphasize the model’s fit. Our results indicate that altitude and latitude are strong predictors of mean temperature, with an inverse relationship between both altitude and latitude with temperature. This analysis shows the potential of these spatial data particularly in predicting local climate conditions.

**Introduction**

Overview of datasets

The dataset contains climate data from 36 weather stations located across France. The data spans 12 variables collected for each 36 weather stations, providing insights into regional temperature, humidity, precipitation, and other factors.

The variables include ‘*station name’, ‘altitude’, ‘latitude’, ‘longitude’, ‘mean temperature’, ‘maximum temperature’, ‘minimum temperature’, ‘relative humidity’, ‘mean precipitation’, ‘maximum precipitation within 24 hours’, ‘number of rainy days’* and ‘*number of sunshine hours per year’.*

**Data Preparation and Cleaning**

Upon exploration of our dataset, we found that the ‘*station name*’ was reported as a character, every other variable was appropriately represented as either numbers or integers except for the *‘altitude’* and *‘mean precipitation’*. We converted these variables to integers to allow for correct analysis.

**Descriptive Analysis**

A summary of our dataset reveals that:

* Altitude ranges from 3m to 2860m with most stations located at lower altitudes, with the median altitude being 96m.
* Latitude ranges from 42.55deg N to 50.73deg N, with the average of 46.19deg N. Most stations are situated between 44deg N and 48deg N, representing southern and central France.
* The mean temperature ranges between -1.20deg C and 16.40deg C, with a median temperature of 11.15deg C.
* The relative humidity ranges from 66% to 89%, with a median of 78%. This indicates a generally high level of moisture across the stations.
* The mean precipitation shows a disparity ranging from 42mm to 201mm. The average being 90.03mm. This indicates averagely moderate precipitation across the stations.
* The annual sunshine hours range from 1574 to 2805 hours, indicating a significant variation in sunshine exposure across different regions.

Figure 1: Variation of Mean Temperature with Relative Humidity and Altitude Across Climate Stations in France

Figure 2: Shows the mean temperature of all stations plotted against the altitude.

Figure 3: Shows the mean temperature of all stations by region

Figure 1 presents the map of France with all 34 climate stations plotted based on their longitude and latitude. The mean temperature at each station is represented by a color scale showing how the temperature varies across different locations across the country. Countries in the South generally experiences higher temperatures, indicated with higher mean temperatures on the map, while stations in the Northern region of France experience cooler mean temperatures.

Figure 2 uses scatter plot to visualize individual data points for the 34 climate stations, with a linear regression line used to reveal a negative correlation trend. We demonstrate through the plot a clear negative relationship between altitude and mean temperature. As altitude increases, the mean temperature decreases. Figure 3 displays a scatter plot of relative humidity and altitude with the color of the points representing the intensity of the mean temperature. The mean temperature provides further insight into how temperature varies across humidity and altitude. Warmer stations (indicated by yellow) are often located altitudes with moderate to high relative humidity, while cooler stations (indicated by blue) tend to be more situated at higher altitudes with lower humidity.

Modelling

**Model 1**

We excluded Mont-Ventoux and Pic-du-midi observations from the data sets before our analysis began because they are extremely mountainous observations able to impact the fit of our model.

To explain the mean temperature by spatial attributes namely the latitude, longitude and altitude, we ran a multiple linear regression model as follow:

model <- lm(mean\_temp ~ altitude + latitude + longitude, data = climfrar)

(we set our dependent variable to the mean temperature, and our predictors as latitude, longitude and altitude).

We have the result presented as follows:

Coefficients:

#               Estimate Std. Error t value Pr(>|t|)

# (Intercept) 37.2650364  2.6220099  14.212 7.29e-15 \*\*\*

# altitude    -0.0064139  0.0008688  -7.383 3.17e-08 \*\*\*

# latitude    -0.5339603  0.0557546  -9.577 1.24e-10 \*\*\*

# longitude    0.0321010  0.0395728   0.811    0.424

# ---

# Signif. codes:  0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

# Residual standard error: 0.7308 on 30 degrees of freedom

# Multiple R-squared:  0.8329,    Adjusted R-squared:  0.8162

# F-statistic: 49.84 on 3 and 30 DF,  p-value: 9.112e-12

**Model 2**

Finding that the predictor longitude was not statistically significant, we excluded it from our multiple linear regression model, and ran another model ‘refined model’.

refined\_model <- lm(mean\_temp ~ altitude + latitude, data = climfrar

# Coefficients:

#               Estimate Std. Error t value Pr(>|t|)

# (Intercept) 37.9147567  2.4828724   15.27 5.68e-16 \*\*\*

# altitude    -0.0062643  0.0008443   -7.42 2.34e-08 \*\*\*

# latitude    -0.5465325  0.0532610  -10.26 1.72e-11 \*\*\*

# ---

# Signif. codes:  0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

# Residual standard error: 0.7268 on 31 degrees of freedom

# Multiple R-squared:  0.8292,    Adjusted R-squared:  0.8182

# F-statistic: 75.26 on 2 and 31 DF,  p-value: 1.268e-12

What about ‘Mont-Ventoux’ and ‘Pic-du-midi’?

Since we had excluded ‘Mont-Ventoux’ and ‘Pic-du-midi’ from our model. We can now predict their respective average temperature using our robust multiple linear regression model.

Our model predicts the average

Mont-Ventoux:

* Measured mean temperature: 3.6°C
* Predicted mean temperature: 6.19°C
* Prediction confidence interval: [4.33°C, 8.04°C]

Pic-du-Midi:

* Measured mean temperature: -1.2°C
* Predicted mean temperature: -3.46°C
* Prediction confidence interval: [-8.13°C, 1.21°C]

The measured mean temperature of 3.6°C for Mont-Ventoux falls within the confidence interval of the prediction (4.33°C to 8.04°C).

Likewise, the measured mean temperature of -1.2°C for Pic-du-Midi also falls within the prediction's confidence interval of -8.13°C to 1.21°C.

From the table above, we see that our prediction aligns with the observed values of mean temperature for both   
‘Mont-Ventoux’ and ‘Pic-du-midi’. Although it does not precisely predict the exact figure, it falls within our confidence interval (bounces).

**Model Interpretation**

**Coefficients Analysis**

1. **Intercept (Estimate = 37.265)**:

The intercept represents the predicted mean temperature when all the **altitude**, **latitude**, and **longitude** are zero.

1. **Altitude (Estimate = -0.0064)**:

The altitude indicates that for every 1-meter increase in altitude, the mean temperature decreases by approximately **0.0064°C**.

The standard error for altitude is **0.0009**, leading to a very large **t-value (-7.383)** and an extremely small p-value (**p < 0.001**). Therefore, altitude is statistically significant.

1. **Latitude (Estimate = -0.534)**:

For every 1-degree increase in latitude, the mean temperature decreases by approximately **0.534°C**.

With a standard error of **0.056**, the corresponding **t-value (-9.577)** is also large, and the p-value (**p < 0.001**). Therefore, latitude is statistically significant.

1. **Longitude (Estimate = 0.0321)**:

For every 1-degree increase in longitude, the mean temperature increases by approximately **0.0321°C**.

However, the standard error for longitude is **0.0396**, yielding a relatively small **t-value (0.811)** and a high p-value (**p = 0.424**). Therefore, longitude is not statistically significant.

**Residual Analysis**

* **Residual Standard Error (RSE = 0.7308)**:

The average deviation of the observed mean temperatures from the model fit. 0.7308 suggests a moderately good fit.

**Multiple R-squared (0.8329)**:

* + Approximately **83.29%** of the variance in mean temperature is explained by altitude, latitude, and longitude.

**Adjusted R-squared (0.8162)**:

This indicates the model has strong explanatory power, even after accounting for the number of predictors.

**Overall Model Significance**

* **F-statistic (49.84, p < 0.001)**:
  + Our model, with a very high F-statistic and a p-value well below 0.001, is highly significant overall.

After removing the Longitude

**3D Scatterplot**

We used our plot here to visualize the relationship between altitude, latitude and our response variable mean temperature, and we overlayed the regression plane to illustrate the model’s fit.

The observed points represent the actual mean temperature values for different altitudes and latitudes while the regression plane is our model fit based on our predictors. Figure 4 shows a

**Model Interpretation**

**Coefficients Analysis**

**Intercept (Estimate = 37.91)**:

The predicted mean temperature when all the **altitude**, **latitude**, and **longitude** are zero is **37.91.**

**Altitude (Estimate = -0.0063)**:

The altitude indicates that for every 1-meter increase in altitude, the mean temperature decreases by approximately **0.0063°C**.

The standard error for altitude is **0.0008**, leading to a very large **t-value (-7.42)** and an extremely small p-value (**p < 0.001**). Therefore, altitude is still statistically significant.

**Latitude (Estimate = -0.546)**:

For every 1-degree increase in latitude, the mean temperature decreases by approximately **0.546°C**.

With a standard error of **0.053**, the corresponding **t-value (-10.26)** is also large, and the p-value (**p < 0.001**). Therefore, latitude is statistically significant.

**Residual Analysis**

**Residual Standard Error (RSE = 0.7268)**:

Upon removal of the longitude as a predictor, the average deviation of the observed mean temperatures from the model fit further reduced to 0.7268 suggesting a good fit.

**Multiple R-squared (0.8292)**:

* + Approximately **82.92%** of the variance in mean temperature is explained by altitude, latitude, and longitude.

**Adjusted R-squared (0.8182)**:

This indicates the model has strong explanatory power, after removing longitude as a predictor.

**Overall Model Significance**

* **F-statistic (75.26, p < 0.001)**:
  + Our refined model has a higher F-statistic and a p-value well below 0.001, suggesting an overall better model.

Conclusion  
**Altitude** and **latitude** are the primary drivers of mean temperature variation in this model, both having highly significant negative relationships with temperature. However, **Longitude** does not contribute significantly to explaining temperature variability and was excluded in our refined model.

The model is robust, explaining over approximately 83% of the variance in mean temperature, with a low residual error and strong overall significance