Long-term rainfall averages for Ireland (1990-2020)

"What is the percentage change in monthly average rainfall in Ireland?"

The thesis is submitted to University College Dublin in part fulfillment of the requirements for the degree of M.Sc. Statistics. The thesis was conducted under the School of Mathematics and Statistics at UCD.

Author:

Anisha Mittal

Supervisor:

Associate Professor Gabrielle Kelly

Submission Date: 8th August 2021

Acknowledgments

I am thankful to my supervisor, Gabrielle Kelly for her valuable time, guidance, remarks, and engagement through the learning process of this dissertation.

I would like to send my immense gratitude to all the employees at Met Éireann, who work tirelessly to provide valuable data about the climate and weather.

I would also like to thank my family and friends for their love and support.

Table of Contents

1. List of Figures

2. List of Tables

3. Abstract

Effects of global warming are visible through out the world. Amount of rainfall is highly affected by global warming. Ireland is an Island, which is particularly vulnerable to changes in sea levels. This thesis aims to study the change in amount of rainfall for Ireland over the past 61 years. Long-term rainfall averages (LTAs) or climate normals are 30 years averages of rainfall which are used here to place the current weather of Ireland in perspective. Monthly LTAs from year 1960 – 1990 and 1991 - 2020 are calculated and compared to see the change in amount of rainfall in Ireland.

Rainfall is a stochastic process and is difficult to study due to rapidly changing weather patterns. Three different models are developed to see the effects of geography such as elevation of stations, cartesian distance from the sea of station, location of stations, and time has on the amount of rainfall. The models are linear model, ARIMA model, and spatial model.

Spatial model was found to be the best model which studies amount of rainfall in Ireland. It indicates that stations with higher elevation receive more rainfall, and stations which are far from the sea receive less rainfall. It also indicates that the geographical effects are spatially correlated up to the radius 20kms. In the end it was found that there was a 6.15% increase in the amount of rainfall in Ireland in 1991-2020 as compared to 1960-1990.

4. Introduction

The average global temperature on Earth has increased by a little more than 1^0 Celsius since 1880[1]. A change of as little as 1^0 Celsius can create a huge impact. Effects of Global Warming are evident as in the past, a drop of $1\text{-}2^0$ Celsius was enough to plunge the Earth into the Little Ice Age[1]. A warming planet leads to climate change which negatively affects the world in many ways. The effects of climate change can be seen in many areas, such as[2]:

1. Deltas, low-lying islands, and coasts being flooded due to rising sea levels.
2. Productive land being turned into deserts.
3. Weather conditions becoming increasingly extreme, and many more.

The increase in temperature due to global warming has also influenced Ireland's natural environment. It has changed the growing season which affects farming. An irregular weather condition could lead to change in rainfall and storms. There has also been an increase in the frequency and impact of storms in the last few decades. Ireland, being an island nation is particularly vulnerable to changing sea levels with coastal regions facing issues of flooding.

Changes in rainfall and other forms of precipitation are one of the most critical factors determining the overall impact of climate change. It is much more difficult to predict rainfall than temperature. The changes in weather patterns make predicting rainfall particularly difficult. Future warming on a global scale is in broad agreement with different climate models, however, there is low agreement at a detailed level when it comes to predicting how these changes will impact whether – and consequently rainfall. Models and observations based on climate are improving each day and the reliability of predictions is improving significantly too.

People say that the climate and the weather have changed as compared to the previous year. We cannot say the climate changed in a year. It is not the correct method to measure this change. The change in climate and weather is studied over a longer period, approximately 30 years. We study the long-term averages (LTAs) or Climate Normals to place the current climate and weather in perspective[3]. Therefore, we will use the long-term rainfall monthly averages from years 1960 to 2020 of Ireland to answer our research question:

"What is the percentage change in monthly average rainfall in Ireland?"

Several models have been proposed to bring in the best study for amount of rainfall in the past. However, due to the stochastic and non-linear nature of rainfall, it is hard to study and/or model rainfall using the statistical models. This paper will fit a linear regression model, an ARIMA model, and geo-statistical model to the monthly LTAs of Ireland from years 1960 to 2020. All these models will aim to study the geographical effects stations has on the amount of rainfall. LTAs for the period 1960-1990 and 1991-2020 will compared to answer the research question.

5. Data Collection and Combination

5.1 About the Data

Met Éireann is the national Meteorological service of Ireland and is the leading provider of weather information and climate related services for Ireland. Data available on its website is easily accessible to all. All the data was very easily collected from the website of Met Éireann. The website has 750+ stations across Ireland which records the precipitation amount (mm), greatest daily fall (mm), number of rain days (0.2mm or more), number of wet days (1.0 mm or more). The values are stored on an hourly, daily, and monthly basis. Monthly averages data from the year 1960 to 2020 was collected for this research. Not all stations are operational for the required time frame. Also, the data is recorded manually at the stations and might be taken on an irregular basis, leading to missing data. Therefore, to avoid missing values 28 stations were selected which were spread across Ireland. They record the data regularly and has the least number of missing values. Only separate CSV files for each station was available to download. A code on R was written to merge all separate files. (List of station in appendix A)

Figure 1 Map of Ireland with plots of stations

To understand the effects geography of stations has on the amount of rainfall, the observed variables are:

- **Rain (mm):** Monthly averages of amount of precipitation in millimeters
- **Height (m) of the station:** This will capture how the elevation at which the rain is recorded affects the amount of rainfall.
- **Easting (m):** The latitude location of the station will capture the spatial trends.
- **Northing (m):** The longitude location of the station will capture the spatial trends.
- **Distance between the station and the sea (m)** will model the coastal effects.
- **Year** will capture the change over time
- **Month** will capture the change over time

In total there were 732*28 = 20,496 observed observations. There were only 1.3% missing values. Data from nearby stations were used for the missing values. As height of station is an explanatory variable used for model building, the average difference in height of the main stations and nearby stations is 11.76m. Also, the nearby stations were on an average approximately 5-8km away from the main stations. Met Éireann does not record the station's distance from the sea, this was done manually. Location of the nearest coastline was selected by eyeballing and the cartesian distance was calculated using R language. The following Figure shows the first eighteen observations of the final data.

	station.id	station.name	county	heightm	easting	northing	distance	year	month	rainmm
1	108	Foulkesmill (Longraigue)	Wexford	71	284185	118335	4480	1960	1	91.6
2	108	Foulkesmill (Longraigue)	Wexford	71	284185	118335	4480	1960	2	89.6
3	108	Foulkesmill (Longraigue)	Wexford	71	284185	118335	4480	1960	3	93.2
4	108	Foulkesmill (Longraigue)	Wexford	71	284185	118335	4480	1960	4	77.7
5	108	Foulkesmill (Longraigue)	Wexford	71	284185	118335	4480	1960	5	49.8
6	108	Foulkesmill (Longraigue)	Wexford	71	284185	118335	4480	1960	6	48.8
7	108	Foulkesmill (Longraigue)	Wexford	71	284185	118335	4480	1960	7	138.3
8	108	Foulkesmill (Longraigue)	Wexford	71	284185	118335	4480	1960	8	110.4
9	108	Foulkesmill (Longraigue)	Wexford	71	284185	118335	4480	1960	9	120.6
10	108	Foulkesmill (Longraigue)	Wexford	71	284185	118335	4480	1960	10	207.3
11	108	Foulkesmill (Longraigue)	Wexford	71	284185	118335	4480	1960	11	190.7
12	108	Foulkesmill (Longraigue)	Wexford	71	284185	118335	4480	1960	12	102.8
13	518	Shannon airport	Clare	15	137900	160300	10	1960	1	83.1
14	518	Shannon airport	Clare	15	137900	160300	10	1960	2	47.2
15	518	Shannon airport	Clare	15	137900	160300	10	1960	3	44.7
16	518	Shannon airport	Clare	15	137900	160300	10	1960	4	70.6
17	518	Shannon airport	Clare	15	137900	160300	10	1960	5	109.4
18	518	Shannon airport	Clare	15	137900	160300	10	1960	6	80.2

Figure 2 First 18 observations of the final combined data

5.2 Exploratory Data Analysis

To evaluate the geographical factors affecting the amount of rainfall in Ireland, a correlation analysis was performed. As the explanatory variables is a count data, it does not meet the assumptions of Pearson correlation tests. Therefore, Spearman correlation is used to check the correlation between the explanatory and response variable. The test generates value between -1 and 1. Where values closer to 1 represent high positive correlation between two variables, values closer to -1 represent high negative correlation between two variables, and 0 represent no correlation between two variables. A positive relation implies as one variable increases the other variable increases as well. A negative relation implies as one variable increases the other variable decreases and vis-a-versa. The Figureure below shows the correlation chart plot of all the variables. The upper triangle of the plot gives the Spearman correlation value, and the stars show that they are statistically significant. The diagonal shows the distribution of and density plot of the variables. The lower triangle shows the scatter plot with lines of regression.

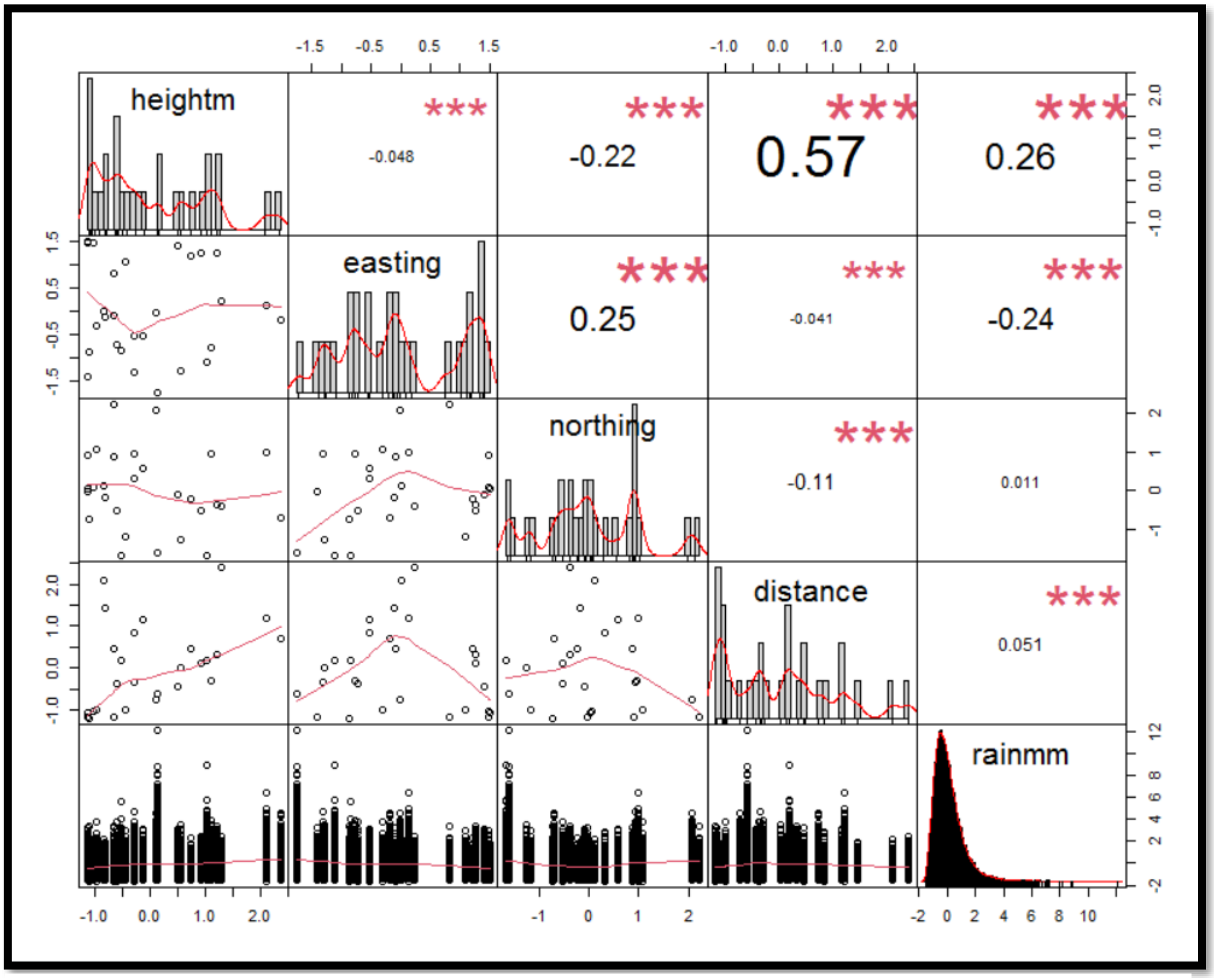

Figure 3Correlation chart of response and explanatory variables

From the correlation chart we see that rain is statistically correlated to all the geographical effects. But the value of correlation is small. The values are closer to zero and therefore represents that they might have less effect on the amount of rainfall. The correlation values are statistically significant because there are ties present in the data. Even though correlation does not mean causation, we get some inference that we would need some new modeling methods to study the geographical effects of stations on rainfall.

It was noticed that due to storm Eva, the maximum amount of rainfall recorded in the past 61 years was 910.7mm at Cloone Lake, Kerry in December 2015. The minimum amount of recorded rainfall was 0.2mm at Drumshanbo, Leitrim in February 1986. The average rainfall from the year 1960 to 2020 was 107.1mm. The following plot shows the change in the average amount of rainfall over the years. We see the average rainfall has an increasing trend but at a slow rate as suggested by the intercept. This means that the amount of rainfall has increased in Ireland, but the change is increasing at a very slow speed.

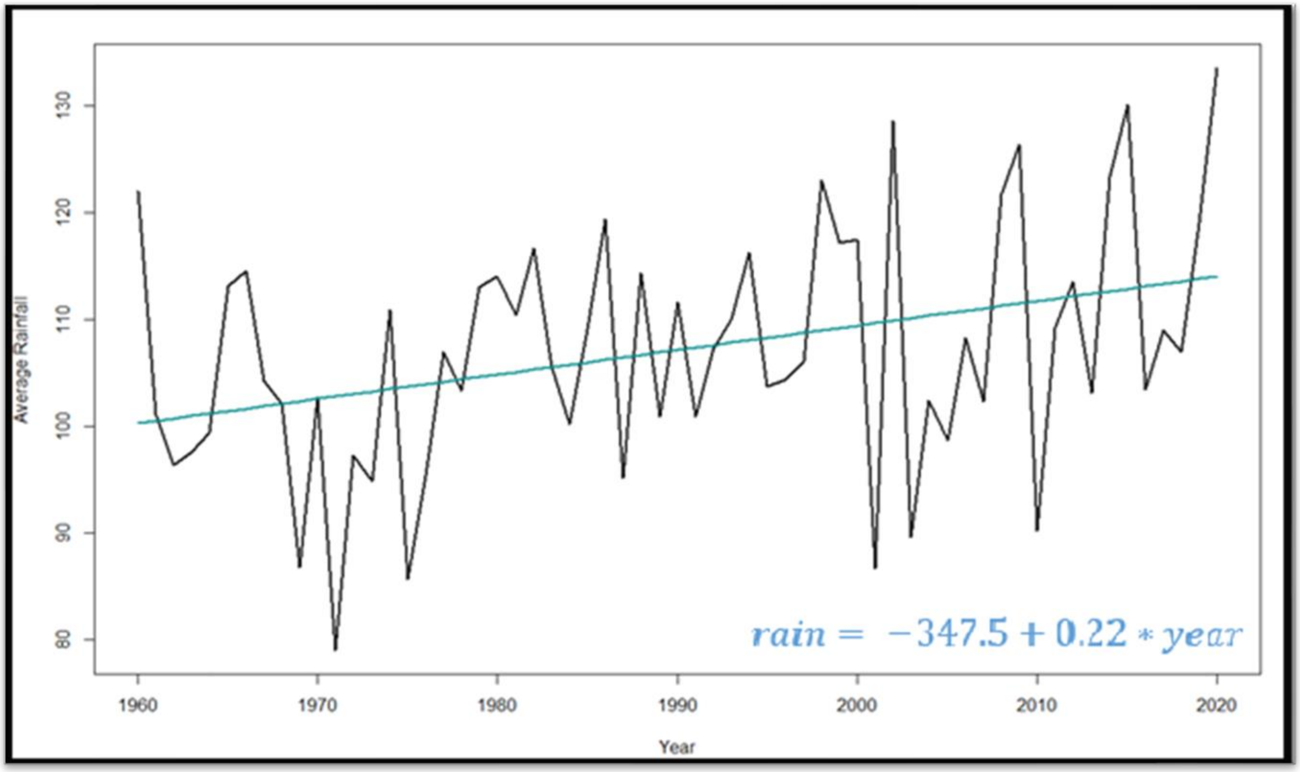

Figure 4Plot of amount of average rainfall over the years with trend line

To see the effects of other variables, an ordinary linear model was developed.

6. Linear Model

A linear model is a very basic and most used predictive model. It models the relationship, if it exists, between a response variable and one or more explanatory variables. The model is developed by estimating unknown model parameters statistically using the data. Linear regression is widely used

A linear model is assumed to fit well to a data set if each data point provides equally precise information about the explanatory variables. In other words, the standard deviation of the error terms should be constant over all values of the explanatory variables. If a model does not meet this assumption, weighted least square method is used to each data points a proper amount of influence over the estimating parameters.

Generalization of an ordinary least squares and linear regression is called Weighted least square linear model. Here the results from an ordinary linear model are used. The reciprocal of the residual variances is used as weights to each data point.

A weighted least square linear regression model is used to model the amount of rainfall. A model with interaction terms using a stepwise regression for the main effects was conducted, namely

$$Rain = (\beta_0 + \beta_1 height + \beta_2 easting + \beta_3 northing + \beta_4 distance + \beta_5 year + \beta_6 month)^2$$

Here all the explanatory variables have the same unit except year and month. Months are treated as a factor variable instead of a numerical variable to capture the monthly change. January is taken as the base year to compare the change in rainfall with respect to other months.

6.1 Model Summary

The model summary is as follows:

β_i	Coefficient	Estimate	P-Value
β_0	Intercept	$-7.005*10^{02}$	$2.28*10^{-09}$
β_1	Height	$4.639*10^{-01}$	$<2*10^{-16}$
β_2	Easting	$2.189*10^{-03}$	$9.46*10^{-06}$
β_3	Northing	$-2.432*10^{-04}$	$< 2*10^{-16}$
β_4	Distance	$-1.748*10^{-03}$	$< 2*10^{-16}$
β_5	February	-2.974*10	$< 2*10^{-16}$
β_6	March	-3.118*10	$< 2*10_{-16}$
β_7	April	-4.600*10	$< 2*10^{-16}$
β_8	May	-4.294*10	$< 2*10^{-16}$
β_9	June	-4.250*10	$< 2*10^{-16}$
β_{10}	July	-4.084*10	$< 2*10^{-16}$
β_{11}	August	-2.400*10	$< 2*10^{-16}$

β_i	Coefficient	Estimate	P-Value
β_{12}	September	2.505*10	$< 2*10^{-16}$
β_{13}	October	-2.698*10	0.1894
β_{14}	November	$6.386*10^{-01}$	0.7601
β_{15}	December	4.877	0.0224
β_{16}	Year	$4.504*10^{-01}$	$2.07*10^{-14}$
β_{17}	Height*Easting	$-8.209*10^{-07}$	$8.45*10^{-11}$
β_{18}	Height*Distance	$-1.635*10^{-06}$	$2.97*a10^{-15}$
β_{19}	Easting*Northing	$1.229*10^{-09}$	$< 2*10^{-16}$
β_{20}	Easting*Distance	$3.603*10^{-09}$	$1.14*10^{-07}$
β_{21}	Northing*Distance	$3.439*10^{-09}$	$< 2*10^{-16}$
β_{22}	Easting*Year	$-1.321*10^{-06}$	$1.05*10^{-07}$

Table 1Estimated parameters of Linear Model

The residual standard error is 1.284 on 20473 degrees of freedom and the adjusted R-squared associated with the model is 0.2656. The related F-statistic with the model is 337.9 on 22 and 20473 degrees of freedom with a p-value less than $2.2*10^{-16}$. Notice that the model does not give an estimate for the month of January. The model takes January as the base month to get the estimates for all the other months. We can see that p-value associated with the month of October and November are not statistically significant, while the others are. The adjusted R-squared associated with the model is less because the estimates for coefficients are very close to zero. This suggests the geographical variable have a very small effect on the amount of rainfall statistically.

6.2 Model Interpretation

The interpretation of the model is as follows:

- A 1% increase in **height of station**, holding other variables as constant, is associated with a **46.39% increase** in the **amount of rainfall**.

- A 1% increase in **easting of station**, holding other variables as constant, is associated with a **0.21% increase** in the **amount of rainfall.**

- A 1% increase in **northing of station**, holding other variables as constant, is associated with a **0.024% decrease** in the **amount of rainfall**.

- A 1% increase in **distance of station from sea**, holding other variables as constant, is associated with a **0.17% decrease** in the **amount of rainfall**.

- A 1% increase in **no. of years**, holding other variables as constant, is associated with a **45.04% increase** in the **amount of rainfall.**

- A 1% increase in the **amount of rainfall in January**, holding other variables as constant, is associated with a **2974% decrease** in the **amount of rainfall in February.**

- A 1% increase in the **amount of rainfall in January**, holding other variables as constant, is associated with a **3118% decrease** in the **amount of rainfall in March**.

- A 1% increase in the **amount of rainfall in January**, holding other variables as constant, is associated with a **4600% decrease** in the **amount of rainfall in April**.

- A 1% increase in the **amount of rainfall in January**, holding other variables as constant, is associated with a **4294% decrease** in the **amount of rainfall in May**.

- A 1% increase in the **amount of rainfall in January**, holding other variables as constant, is associated with a **4250% decrease** in the **amount of rainfall in June**.

- A 1% increase in the **amount of rainfall in January**, holding other variables as constant, is associated with a **4084% decrease** in the **amount of rainfall in July**.

- A 1% increase in the **amount of rainfall in January**, holding other variables as constant, is associated with a **2400% decrease** in the **amount of rainfall in August**.

- A 1% increase in the **amount of rainfall in January**, holding other variables as constant, is associated with a **2505% decrease** in the **amount of rainfall in September**.

- A 1% increase in the **amount of rainfall in January**, holding other variables as constant, is associated with a **487.7% increase** in the **amount of rainfall in December**.

- For any fixed value of **easting,** $easting_0$, the effect for **height** is given by $0.4 - 0.0000008*easting_0$. This suggests that there is no unique effect of height, it is different for each possible easting value.

- For any fixed value of **distance,** $disctance_0$, the effect for **height** is given by $0.4 - 0.000001*distance_0$. This suggests that there is no unique effect of height, it is different for each possible distance value.

- For any fixed value of **northing,** $northing_0$, the effect for **easting** is given by $0.002 + 0.000000001*northing_0$. This suggests that there is no unique effect of easting, it is different for each possible northings value.

- For any fixed value of **distance, distance$_0$**, the effect for **easting** is given by $0.002 + 0.00000003*distance_0$ **.** This suggests that there is no unique effect of easting, it is different for each possible distance value.

- For any fixed value of **distance,** $distance_0$, the effect for **northing** is given by $-0.0002 + 0.00000003*distance_0$. This suggests that there is no unique effect of northing, it is different for each possible northing value.

- For any fixed value of **year,** $year_0$, the effect for **easting** is given by $0.002 - 0.000001*easting_0$. This suggests that there is no unique effect of easting, it is different for each possible year value.

This is a very complex model and gives a log-likelihood ratio of -109473 with 24 degrees of freedom. All these results when compared to the real data did not give many correct values, to see if this model is good fit, the residual plots were seen.

6.3 Model Diagnostics

From the diagnostics plots, the residuals vs leverage plot suggests all the points are under Cook's Distance and hence does not indicate that there are any influential or leverage points in the data. The scale-Location plot suggests a constant band around one and therefore indicate that residuals are spread equally along the range of predictors. But the QQ-plot does not suggest the residuals are normally distributed and performing a Lilliefors test for normality on the residuals gives a p-value very close to zero which means non-normality, but because of the large data set the central limit theorem suggests normality of the residuals . The residuals versus fitted values plot shows the residuals are around zero but the variability very slowly increases over time, so the residuals experience some variance. This means that the amount of rainfall has changed over the years, which makes the predictive model so uncertain.

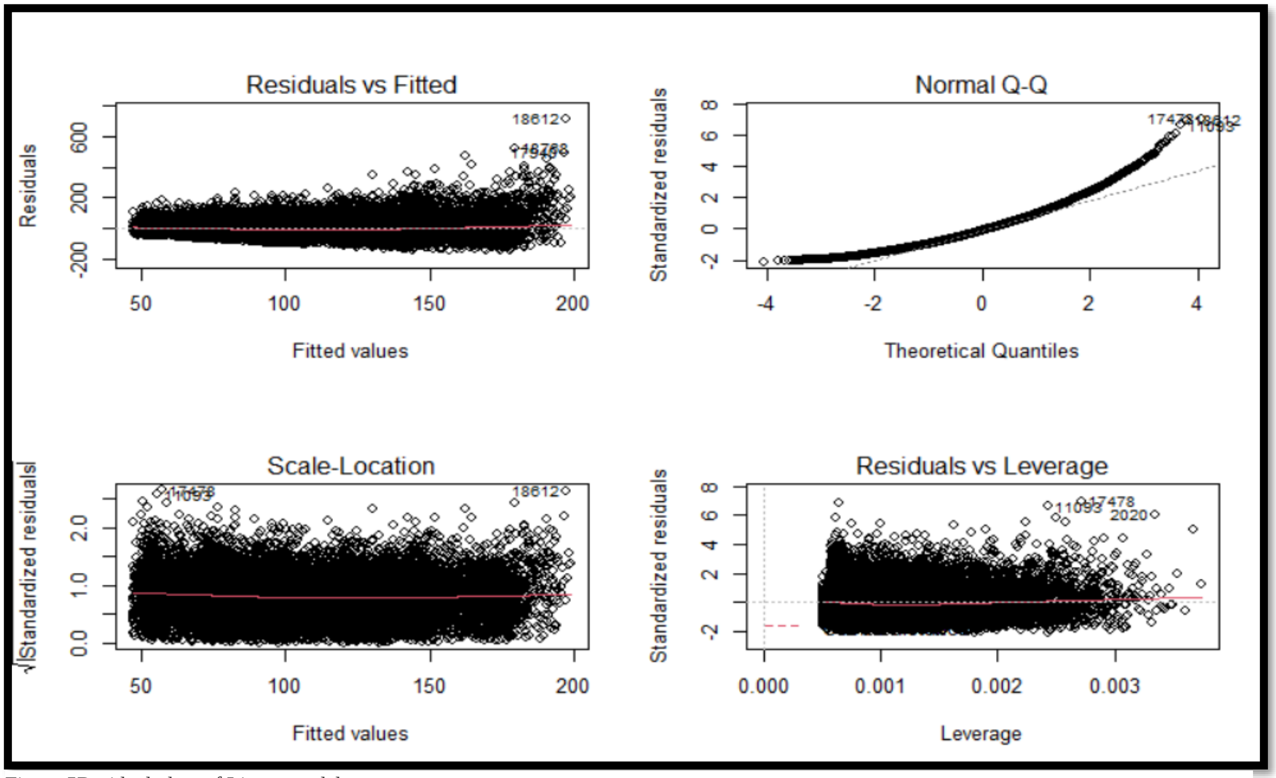

Figure 5Residual plots of Linear model

Therefore, there is a need for a better model which can account the geographical effects and make better predictions.

Next, we look if a time series model can help answer the research topic.

7. ARIMA Model

ARIMA has been a widely recognized choice for making forecasting based on statistical inference. The auto- regressive (AR),Integrated (I), moving average(MA), popularly known as the ARIMA model has been a widely recognized model for making future predictions of a time series-based data based on some linear function of previous values that are known to have direct effect on the forecasts being made. In theory, ARIMA models are used for forecasting a time series that can be made stationary by differentiating. A random variable is stationary if it is found to have a constant statistical feature that is if its autocorrelations that is the correlation with its own prior derivations around mean remains constant over time. This can be tested using the Augmented Dickey Fuller (ADF) test. ADF tests the null hypothesis that a series is not stationary and alternatively, the series is stationary. Where the null hypothesis is rejected if the p-value is less than the significance level of 0.05.

A non-seasonal ARIMA model defined by an autoregressive term, a moving average and differentiating term or ARIMA(p,d,q) where p is the number of autoregressive terms, d is the number of non- seasonal differences needed for stationarity and q is the number of lagged forecast errors in the prediction equation.

Since we must see how the amount of rainfall is affected by the geographical features of the stations, in this study along with time; height, distance from the sea, easting, and northing of the stations are also used as regressors. The auto.arima function from the forecast package in R, takes the argument of choosing different numerical regressors.

For the rainfall data, the Augmented Dickey-Fuller provided a p-value of 0.01, which suggests there is enough evidence to reject the null hypothesis and assume there does not exist any unit root. Therefore, there will be no Integrated (I) component in the ARIMA model.

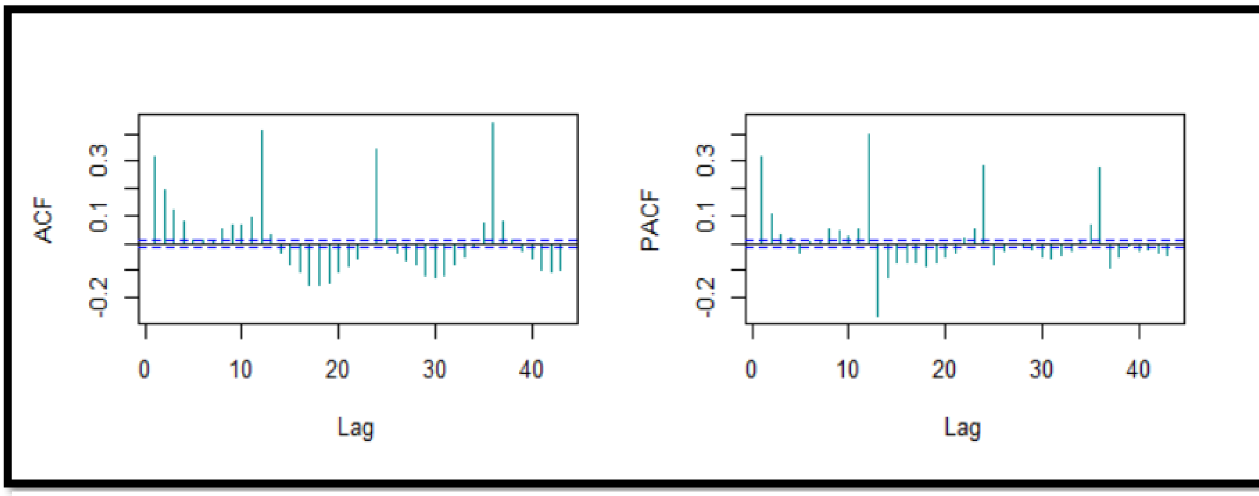

Figure 6ACF and PACF plots of amount of rainfall as a time series

The ACF and the PACF plots suggests there exists seasonality in the data as we can see their graph looks like a plot of sine function. So, we cannot decide the p, d, and q values to estimate the parameters for the model. Hence the auto.arima function, as the name suggests, fits an ARIMA model to the data on its own.

The data was transformed by adding dummy variables instead of stating months as factors. Month of January was taken as the base with all the value as 1. The following Figure shows the transformed data

	station.id	station.name	county	heightm	easting	northing	distance	year	month_1	month_2	month_3	month_4	month_5	month_6	month_7	month_8	month_9	month_10	month_11	month_12	rainmm
1	108	Foulkesmill (Longraigue)	Wexford	71	284185	118335	4480	1960	1	0	0	0	0	0	0	0	0	0	0	0	91.6
2	108	Foulkesmill (Longraigue)	Wexford	71	284185	118335	4480	1960	1	1	0	0	0	0	0	0	0	0	0	0	89.6
3	108	Foulkesmill (Longraigue)	Wexford	71	284185	118335	4480	1960	1	0	1	0	0	0	0	0	0	0	0	0	93.2
4	108	Foulkesmill (Longraigue)	Wexford	71	284185	118335	4480	1960	1	0	0	1	0	0	0	0	0	0	0	0	77.7
5	108	Foulkesmill (Longraigue)	Wexford	71	284185	118335	4480	1960	1	0	0	0	1	0	0	0	0	0	0	0	49.8
6	108	Foulkesmill (Longraigue)	Wexford	71	284185	118335	4480	1960	1	0	0	0	0	1	0	0	0	0	0	0	48.8
7	108	Foulkesmill (Longraigue)	Wexford	71	284185	118335	4480	1960	1	0	0	0	0	0	1	0	0	0	0	0	138.3
8	108	Foulkesmill (Longraigue)	Wexford	71	284185	118335	4480	1960	1	0	0	0	0	0	0	1	0	0	0	0	110.4
9	108	Foulkesmill (Longraigue)	Wexford	71	284185	118335	4480	1960	1	0	0	0	0	0	0	0	1	0	0	0	120.6
10	108	Foulkesmill (Longraigue)	Wexford	71	284185	118335	4480	1960	1	0	0	0	0	0	0	0	0	1	0	0	207.3
11	108	Foulkesmill (Longraigue)	Wexford	71	284185	118335	4480	1960	1	0	0	0	0	0	0	0	0	0	1	0	190.7
12	108	Foulkesmill (Longraigue)	Wexford	71	284185	118335	4480	1960	1	0	0	0	0	0	0	0	0	0	0	1	102.8
13	518	Shannon airport	Clare	15	137900	160300	10	1960	1	0	0	0	0	0	0	0	0	0	0	0	83.1
14	518	Shannon airport	Clare	15	137900	160300	10	1960	1	1	0	0	0	0	0	0	0	0	0	0	47.2
15	518	Shannon airport	Clare	15	137900	160300	10	1960	1	0	1	0	0	0	0	0	0	0	0	0	44.7
16	518	Shannon airport	Clare	15	137900	160300	10	1960	1	0	0	1	0	0	0	0	0	0	0	0	70.6
17	518	Shannon airport	Clare	15	137900	160300	10	1960	1	0	0	0	1	0	0	0	0	0	0	0	109.4
18	518	Shannon airport	Clare	15	137900	160300	10	1960	1	0	0	0	0	1	0	0	0	0	0	0	80.2

Figure 7First 18 observations of the transformed data for ARIMA model

Using the auto.arima function on R and using columns 4 to 8 and 10 to 20 as regressors, an ARIMA(4,0,1) model was fitted by to the data. Maximum Likelihood method is used to estimate the coefficients.

The Model summary is as follows:

$\beta_i/\varphi_i/\theta_i$	Coefficient	Estimate	P-Value
φ_1	Ar1	1.0356	$< 2.2*10^{-16}$
φ_2	Ar2	-0.0859	$< 2.2*10^{-16}$
φ_3	Ar3	0.0047	0.637164
φ_4	Ar4	0.0240	0.001411
θ_1	Ma1	-0.9308	$< 2.2*10^{-16}$
β_0	Intercept	-292.7417	0.034856
β_1	Height	0.3225	$< 2.2*10^{-16}$
β_2	Easting	$-2.5896*10^{-04}$	$< 2.2*10^{-16}$
β_3	Northing	$3.4350*10^{-05}$	$1.805*10^{-07}$
β_4	Distance	$-6.3928*10^{-04}$	$< 2.2*10^{-16}$
β_5	February	-3.1596*10	$< 2.2*10^{-16}$

$\beta_i/\varphi_i/\theta_i$	Coefficient	Estimate	P-Value
β_6	March	-3.5151*10	$< 2.2*10^{-16}$
β_7	April	-5.5467*10	$< 2.2*10^{-16}$
β_8	May	-5.2461*10	$< 2.2*10^{-16}$
β_9	June	-5.2581*10	$< 2.2*10^{-16}$
β_{10}	July	-4.9723*10	$< 2.2*10^{-16}$
β_{11}	August	-3.1104*10	$< 2.2*10^{-16}$
β_{12}	September	-3.0993*10	$< 2.2*10^{-16}$
β_{13}	October	-5.5534	0.002975
β_{14}	November	$-6.6763*10^{-01}$	0.720857
β_{15}	December	5.2098	0.003583
β_{16}	Year	$2.2816*10^{-01}$	0.001062

Table 2Estimated parameters of ARIMA Model

We can see that p-value associated with the month of November and AR3 component are not statistically significant.

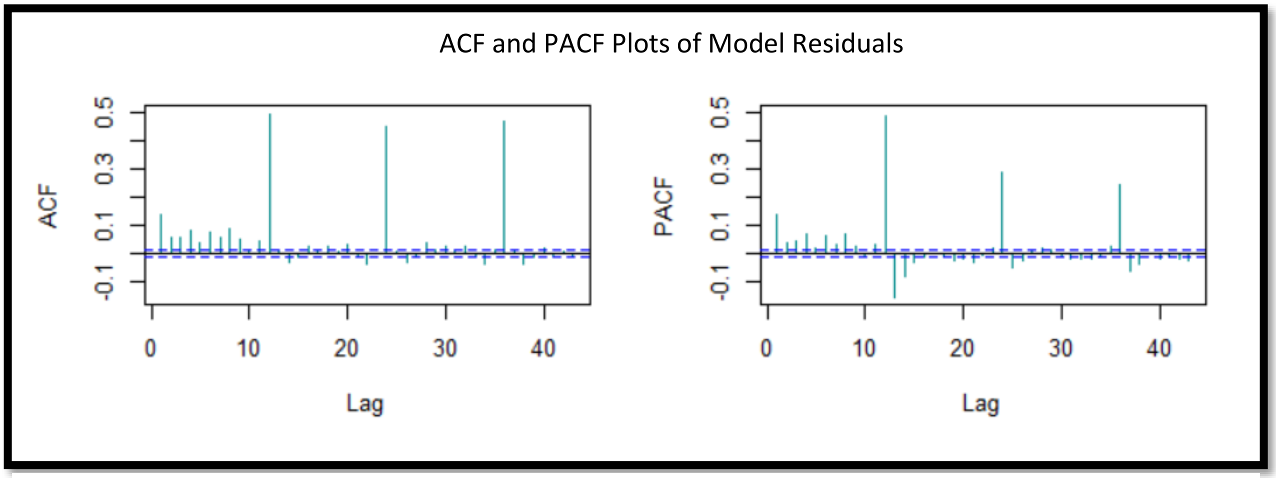

Figure 8ACF and PACF plots of ARIMA model residuals

The ACF and PACF plots of model residuals suggests that at every 12th lag the data is highly varied. This means there is high volatility present in the data. This explains that as compared to January, over the years Ireland has experienced the most amount of rainfall in the month in December. The positive estimate of height suggests that more amount of rainfall is recorded at higher stations of Ireland. The four AR components represents how a current value of rainfall is dependent on the past 4 values of series. AR1 is more than one and signifies that the amount of rainfall in the current month is highly affected by the amount of rainfall during the previous month. AR2 has a negative value which means that the amount of rainfall in the current month is negatively affected by the amount of rainfall recorded two months before the current time. AR3 does not have a significant effect as stated above and AR4 has a small coefficient value. The MA component means how the error terms affects a time series. From the selected model, the error from the previous one-month value negatively affects the current months response value.

The log-likelihood of the model is -111206.1, indicating that the model is complex and computationally heavy. ARMA(4,1) does not fit the data well as the spikes in ACF and PACF plots of residuals are not inside the critical line (Horizontal dashed blue lines). The ljung-Box tests a model fits well as the null hypothesis and alternatively it does not. The null hypothesis is rejected if the tests p-value is less than 5% significance level. When tested, ARMA(4,1) generated a Ljung Box p-value very close to 0 implying the model does not fit well. There might be many reasons why this model does not work properly. There are times when an ARIMA model does not fit a series due to presence of heteroskedasticity. In that case we use an ARCH model, where ARCH represents the Auto-Regressive Conditional Heteroskedasticity. Generalized ARCH (GARCH) have been found to be successful in capturing the volatility or the conditional variance structure of many time series. To check the presence of ARCH component, McLeod - Li test is used on the squared residuals of the ARIMA model. The GARCH components are chosen by examining ACF and PACF plots of squared residuals of fitted ARIMA model. ARCH theory was developed by Robert F. Engle and Clive Granger who also won **The Nobel Memorial Prize in Economic Sciences**. Their paper was about the methods of analyzing economic time series with time-varying volatility (ARCH).

ARIMA is known to give predictions that are linear in nature. But in real life data, there also exists some non-linear components to data. ANN models outperform the ARIMA models in being able to give predictions for non-linear components as well. As observed, rainfall is stochastic in nature and has some non-linearity attached to the prediction. Therefore, an ANN model is preferred in most cases when making predictions of weather components. A hybrid of ARIMA and ANN model, proposed by Zhang[4] has been proved to give better accuracy when predicting time series. It holds true as it is observed that almost all- real world time series contains both linear and non-linear correlation components. Zhang has been able to develop an important concept to model these correlations using the hybrid model which takes the capabilities of ARIMA to capture linearity and of ANN model to model the left-out nonlinearity, giving a more optimized model that outperforms both ANN and ARIMA individually.

But I will not be fitting a GARCH or an ANN or any hybrid of these model with ARIMA models. For the scope of this paper, I will use geostatistical models which would account spatial trends and coastal effects to understand rainfall of Ireland better.

8. Spatial Model

8.1 Introduction

Geo-statistics is a branch of applied statistics that deals with spatially correlated data based on the theory of the regionalized variables. It was initially addressed by George Matheron of the Centre de Morophologie Mathematicque in Fontainebleau, France in 1960s. The original purpose of geo-statistics is centered on estimating changes in ore grade within a mine. However, the principles have been applied to a variety of areas in geology and other scientific disciplines.[6] One such disciplines are studying how geography effects the weather and climate of a certain plalce.

Peter J. Dingle and *Paulo J. Ribeiro jr.* in their cook define spatial statistics as a wide range of statistical models and methods intended for the analysis of spatially referenced data. They theorized that within spatial statistics, the term geo-statistics refers to models and methods for data with characteristics such as, values Y_i : i = 1, . . . , n is observed at a discrete set of sampling locations x_i within some spatial region *A*. Each observed value Y_i is statistically related to the value of an underlying continuous spatial phenomenon, $S(x)$, at the corresponding sampling location x_i. [5]

Gaussian stochastic processes are widely used in practice as models for geostatistical data. These models rarely have any physical justification. Rather, they are used as convenient empirical models which can capture a wide range of spatial behavior according to the specification of their correlation structure.[6]

A gaussian spatial process,$\{S(x): x \varepsilon R^2\}$, is a stochastic process with the property that for any collection of locations $x_1,x_2,\ldots,x_n$ with each $x_i \varepsilon R^2$, the joint distribution of $S = \{S(x_1),\ldots,S(x_n)\}$ is multivariate Gaussian. Any process of this kind is completely specified by its mean function, $\mu(x) = E[S(x)]$, and its covariance function, $\gamma(x, x') = Cov\{S(x), S(x')\}$. [5]

The families of covariance functions that are already developed are *The Matérn family* and *The powered exponential family*.

The Matérn family is a two-parameter family function and is defined by the following correlation function:

$$\rho(u) = \{2^{\kappa-1}\Gamma(\kappa)\}^{-1}(u/\varphi)^{\kappa}K_{\kappa}(u/\varphi)$$

Where $K_\kappa(\cdot)$ is the modified Bessel function of order κ, $\varphi > 0$ is a scale parameter with the dimensions of distance, and $\kappa > 0$, called the order, is a shape parameter which determines the analytic smoothness of the underlying process S(x). Here κ takes the values 0.5, 1.5, and 2.5. [5]

The powered exponential family is defined by the following correlation function:

$$\rho(u) = \exp\{-(u/\varphi)^{\kappa}\}$$

Like the *Matérn* family, it has a scale parameter $\varphi > 0$, a shape parameter κ, in this case bounded by $0 < \kappa \leq 2$, and generates correlation functions which are monotone decreasing in *u*. [5]

Within our model-based framework, the nugget effect is the measurement error variance, τ^2, or equivalently the conditional variance of each measured value Y_i given the underlying signal value $S(x_i)$.

8.2 Alteration of Rainfall data

The **geoR** package in R has a function called likfit which fits a gaussian spatial model to a geo-data. The family of covariance matrix to be used is given as an argument with some initial values for τ^2, σ^2, and κ. Since the function only read a spatial data, the original data when transformed into a spatial data was very heavy. Likfit function is memory intensive and hence could not run on a large dataset. Solutions to tackle this was to either run the code using Virtual Machines provided by Google Cloud Platform(GCP) or to transform data in a way that would reduce the no. of observations. As this paper is time sensitive and learning GCP could take some time, the data was altered. The original data was transformed by taking average of amount of rainfall of each month over the years for each station. Averages were taken for two intervals 1960-1990 and 1991-2020, so that two models could be built to compare and see the change in rainfall. This way, the data now had 12*28 = 336 observations and 4 variables. Easting and northing were considered as the coordinates to capture spatial trends, while height and distance from the sea was used to model elevation and coastal effects respectively. The following Figure shows the transformed data:

	Station ID	Station Name	County	Easting	Northing	Height	Distance	Avg_1960_1990	Avg_1991_2020	month
1	108	Foulkesmill (Longraigue)	Wexford	284185	118335	71	4480	106.81613	98.47000	1
2	108	Foulkesmill (Longraigue)	Wexford	284185	118335	71	4480	82.62258	73.39333	2
3	108	Foulkesmill (Longraigue)	Wexford	284185	118335	71	4480	79.83548	71.98000	3
4	108	Foulkesmill (Longraigue)	Wexford	284185	118335	71	4480	63.52581	73.89000	4
5	108	Foulkesmill (Longraigue)	Wexford	284185	118335	71	4480	65.70968	68.04000	5
6	108	Foulkesmill (Longraigue)	Wexford	284185	118335	71	4480	59.76774	79.48000	6
7	108	Foulkesmill (Longraigue)	Wexford	284185	118335	71	4480	60.91290	75.93333	7
8	108	Foulkesmill (Longraigue)	Wexford	284185	118335	71	4480	84.78387	84.29667	8
9	108	Foulkesmill (Longraigue)	Wexford	284185	118335	71	4480	83.50000	79.84000	9
10	108	Foulkesmill (Longraigue)	Wexford	284185	118335	71	4480	110.65484	118.42000	10
11	108	Foulkesmill (Longraigue)	Wexford	284185	118335	71	4480	101.60968	114.01333	11
12	108	Foulkesmill (Longraigue)	Wexford	284185	118335	71	4480	103.21290	109.38333	12
13	518	Shannon airport	Clare	137900	160300	15	10	96.70645	103.12333	1
14	518	Shannon airport	Clare	137900	160300	15	10	71.29032	89.56667	2
15	518	Shannon airport	Clare	137900	160300	15	10	70.94194	77.84333	3
16	518	Shannon airport	Clare	137900	160300	15	10	55.99677	62.75333	4
17	518	Shannon airport	Clare	137900	160300	15	10	61.72258	63.24000	5
18	518	Shannon airport	Clare	137900	160300	15	10	62.97097	69.69000	6

Figure 9First 18 observations of the transformed data for Spatial model

8.3 Model Summary

Model was developed by using both *Matérn* and *Exponential* family as covariance matrix. The *Matérn* covariance function was tested for κ values of 0.5, 1.5, and 2.5. It was seen that for κ = 0.5, the *Matérn* correlation function reduces to the exponential function $\rho(u) = \exp(-u/\varphi)$, whilst as $\kappa \to \infty$, $\rho(u) \to \exp\{-(u/\varphi)^2\}$ which is also called the Gaussian correlation function or Gaussian model.

Diggle and Ribeiro suggest that 'Scale parameters corresponding to different orders of *Matérn* correlation is not directly comparable. The relationship between the practical range and the scale parameter φ therefore depends on the value of κ'. It was seen the practical range is approximately 3φ, 4.75φ and 5.92φ for the *Matérn* functions with κ = 0.5, 1.5 and 2.5, respectively. The practical range increases as $\kappa \to \infty$. The initial values for σ^2, τ^2, and φ were chosen as stated in Dingle and Riberio example data set s100. Then with some hit and trial values, their final values were selected. Model which gives the minimum negative log-likelihood is selected.

Two model were developed, one each for the LTAs of the years 1960-1990 and 1991-2020.

The model summaries are as follows:

Model with Linear trend									
Cov. Fun	Model	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\sigma}^2$	$\hat{\varphi}$	$\hat{\tau}^2$	Log L	Years
Matérn	$\kappa = 2.5$	93.8869	0.2639	-0.0005	667.8982	20000	682.1933	-1601	1960-1990
Matérn	$\kappa = 2.5$	99.1297	0.2614	-0.0005	696.1862	20000	750.0779	-1619	1991-2020

Table 3Estimated parameters of Spatial Model

$\hat{\beta}_0$: Intercept; $\hat{\beta}_1$: Estimated Co-efficient for Height; $\hat{\beta}_2$: Estimated Co-efficient for Distance.
$\hat{\sigma}^2$: Variance of Gaussian Model; $\hat{\tau}^2$: Variance of error terms; $\hat{\varphi}$: Scale Parameter

8.4 Model Interpretation

As seen earlier in section 5 of this paper, the geographical variables were statistically correlated to amount of rainfall, but the values were small. Using the gaussian model, we see φ has a very large value it suggests that probability density function of model is widely spread. For our data, it means the data is spatially correlated. Φ has the value of 20000m = 20km, which means the model suggests that amount of rainfall is affected by spatial trends which are in 20km radius of a station.

Model interpretation for the LTAs of years 1960-1990:

- A 1% increase in **height of station**, holding other variables as constant, is associated with a **26.39% increase** in the **amount of rainfall**.

- A 1% increase in **distance from the sea**, holding other variables as constant, is associated with a 0.05% **decrease** in the **amount of rainfall**.

Model interpretation for the LTAs of years 1991-2020:

- A 1% increase in **height of station**, holding other variables as constant, is associated with a **26.14% increase** in the **amount of rainfall**.

- A 1% increase in **distance from the sea**, holding other variables as constant, is associated with a 0.05% **decrease** in the **amount of rainfall**.

The model shows areas near the coast experience more rainfall, than the areas away from rainfall. More amount of rainfall is recorded at more elevated regions of Ireland. For example: Three stations in Dublin with heights 152m, 13m, and 21m and distance from the sea 16560m, 2679m, and 3270m respectively, have the average amount of rainfall from 1960 to 2020 as 96mm, 59mm, and 61mm respectively. The stations with height difference of 8m and distance difference of 591m records only 2mm lesser amount of rainfall, but the station with max height but far from the sea has a difference of 35-37mm of rainfall. We can see the effect of distance from the sea has affected the average amount of rainfall for the station with maximum height as suggested by the model.

Both models gives very similar results, but the variance of both model are very different. The model variance increases by 28.288, as height and distance from the sea has not changed over the years it means the environmental factors affecting the rainfall has changed. On taking the difference of averages of 1960-1990 and 1991-2020, there was a 6.15% increase in the amount of monthly rainfall in Ireland.

9. Summary and Conclusions

The three models that were fitted indicate similar results, but not all models meet the assumptions associated with model diagnostics. The following Tablele shows the Log-likelihood of all the models, this processes the shape and curvature of the likelihood surface representing information about the sTableility of the estimates, which is why the likelihood function is often plotted as part of a statistical analysis. The likelihood associated with each fitted model is as follows.

Model	Log-Likelihood
Weighted Least square Linear Model	-109473 with 24 degrees of freedom.
ARIMA Model	-111206.1
Gaussian Model	Model 1: -1601
	Model 2: -1619

Table 4Model comparison Table

The Gaussian Model maximizes the Log-likelihood (Minimizes the negative log-Likelihood) for Ireland long-term rainfall averages. Therefore, results from Gaussian model should be considered.

The model suggests that there is more rainfall recorded at higher station which are closer to the sea. So, this would mean that stations even with smaller heights but are situated on mountains will record more rainfall. While Ireland is an Island, it has mountains on its western, Northwestern, and Southwestern regions as seen in Figure1. So, counties like Mayo, Galway, Cork, and Donegal experience most amount of rainfall. This is also inferred by the data. Model also suggests that when moving away from the sea and towards more plainer regions the amount of rainfall is lesser than the other regions.

The long-term monthly rainfall averages of Ireland with the geographical effects of stations indicates that there has been a significant increase in the amount of rainfall over the period of 61 years. Since the height of the station, distance between the sea and stations, and location of stations do not change over time, it illustrates the environmental features associated with these geographical variables has changed over the years causing the amount of rainfall to increase. The environmental features are the coastal effects, elevation, and the spatial trends associated with the distance, height, and location of the stations respectively. There was a 6.15% increase in the monthly amount of rainfall recorded for the period 1991-2020 as compared to monthly amount of rainfall for the period 1960-1990.

Effects of global warming can be seen and hence require attention by the society and government. The government can contribute towards saving the environment by incorporating changes in regulations as suggested by The UN's 17 Sustainable Development Goals (SDGs). It was seen that the environment was healing itself slowly in the past two years when all of humanity was on lockdown, it hints that human needs to develop ways to coexist with nature and not induce harm towards the environment. I would like to end this paper with a quote by *Dr. Wangari Maathai* "The future does not exist in the future. Rather, it is born only through our actions in the present, and if we want to realize something in the future, we must take action toward it now."

10. Future directions for research

Virtual machines on Google Cloud Platform can be used to model the whole data, which might give better results. The book *Model-based Geostatistics by Peter J. Diggle and Paulo J. Ribeiro jr.* have a lot of information about geo-statistical modeling. They provide with many different methods of estimations of parameters of a gaussian model. Therefore, using different methods to model rainfall on virtual machines can be developed which has a better log-likelihood than the model suggested in this paper. If virtual machines are being used more stations can be studied. Methods suggested by *Mr. Séamus Walsh*, in his paper titled *"New Long-Term Rainfall Averages for Ireland"* can be used to deal with missing values if more stations are to be studied. Further methods for spatial predictions can be studied and model can be used to for future predictions.

11. References

1. https://earthobservatory.nasa.gov/world-of-change/global-temperatures#:~:text=According%20to%20an%20ongoing%20temperature,2%C2%B0%20Fahrenheit)%20since%201880

2. http://www.askaboutireland.ie/enfo/irelands-environment/Change/impacts-of-climate-change/

3. http://edepositireland.ie/bitstream/handle/2262/76135/Climatological%20Note%20No.%2015%20-.pdf?sequence=1&isAllowed=y

4. Zhang, G.,P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing 50, 159-175.

5. Peter J. Diggle and Paulo J. Ribeiro jr (2007): Model-Based Geostatistics

6. Erfan Hosseini, Raoof Gholami, and Farzad Hajivand: Geostatistical modeling and spatial distribution analysis of porosity and permeability in the Shurijeh-B reservoir of Khangiran gas field in Iran

12. Appendices

A. List of stations with their height(m), distance from the sea(m), and location(m).

Station.Name	County	Easting	Northing	Height	Distance
Foulkesmill (Longraigue)	Wexford	284185	118335	71	4480
Shannon airport	Clare	137900	160300	15	10
Greencastle	Donegal	264600	440800	53	250
Creeslough (Brockagh)	Donegal	201625	425840	119	9980
Loughglinn	Roscommon	163400	286000	98	51090
Castle Island (Coom)	Kerry	107400	109900	157	26690
Tulla	Clare	148865	179955	57	18000
Costelloe Fishery	Galway	97455	226775	12	590
Omeath	Louth	314155	316690	12	250
Meelick (Victoria lock)	Offaly	194600	212900	39	57650
Cloone lake	Kerry	71200	78400	122	13010
Drumshanbo	Leitrim	196000	312500	54	36140
Glasnevin	Dublin	315000	237000	21	3270
Athlone OPW	Westmeath	203900	241300	37	71990
Dromahair (Market st.)	Leitrim	180600	331500	27	4160
Cuilcagh MTNS.	Cavan	212990	324125	290	52310
Hacketstown (Voc. Sch.)	Carlow	297500	179900	189	28860
Eskeragh	Mayo	104100	318965	85	18560
Glen Imaal (for. Stn.)	Wicklow	297915	194835	213	33330
Glenamaddy	Galway	162795	261085	84	44450
Cloonacool (Lough Easkey)	Sligo	144600	320700	204	19680
Slieve Bloom MTNS. (Nealstown)	Laois	219900	193600	219	78590
Carrigadrohid (gen. stn.)	Cork	140640	71950	65	29910
Macroom (renaniree)	Cork	120100	72600	198	29810
Ballymore Eustace DCWW	Kildare	293300	209200	172	36270
Merrion square	Dublin	316400	233500	13	2670
Silvermines MTNS. (Curreeny)	Tipperary	190100	164700	312	41730
Glenasmole (Supts lodge)	Dublin	309200	222200	152	16560