**A Spatio-Temporal Analysis of UK Rainfall Data**

1. **Introduction and Data Description**
   1. Aims

The aim of the project is to analyse the spatio-temporal characteristics of historical UK rainfall data and to employ statistical modelling techniques, such as Autoregressive Integrated Moving Average (ARIMA) and Artificial Neural Networks (ANN), to determine whether the data shows attributes that enables it to be modelled based on its past values such that future values can be accurately forecast.

Time series analysis involves designing a model to enable extrapolation into the future based on the expectation that historic patterns will repeat over time. Several time series modelling methodologies have been developed of which ARIMA is one. ARIMA models have their roots in electrical engineering [1] and were first adapted for analysis of time series using statistical methods by Box and Jenkins in the 1970s [2]. However, there are limitations with ARIMA models, some of which will be discussed including the stationarity requirement and ability to deal with non-linear data. This has led to the emergence of artificial intelligence models such as ANN as powerful alternatives for time series forecasting [3].

* 1. UK Regional Rainfall Data

Monthly rainfall records for the UK were sourced from the UK Met Office [4]. The inspiration for using this data came from an article describing how the data was recently updated through the citizen science project 'Rainfall Rescue' [5]. The data covers the period January 1836 to December 2021 (2232 months) for each of the 10 district regions of the UK as defined by the UK Met Office [6].

Data for each region was downloaded as a .txt file, combined and indexed in an Excel file, and converted to .csv format (“uk\_rainfall\_data.csv”) to enable them to be joined with shapefiles. Shapefiles for the Great Britain district regions were sourced from the CEGE0042 Tutorial data. A further shapefile for Northern Ireland was sourced from OSNI Open Data [7]. These were combined into a single shapefile and then spatially joined with the rainfall data in ArcGIS Pro to create a geodatabase feature layer for importing as a dataframe into R: layer name “uk\_rain\_all\_districts” in file “UK Rainfall.gdb”.

* 1. UK Weather Station Rainfall Data

Chart, histogram

Description automatically generatedChart, histogram

Description automatically generatedMonthly UK rainfall data for weather stations sites across the UK was also sourced from the UK Met Office to give a point dataset to complement the areal dataset described in 1.2. The complete record contains monthly data for 37 weather station sites, with records of differing lengths going back as far as 1853 in some cases (see “Station Data.xlsx”). Due to not all the weather station sites having complete records, a subset of sites was identified that had near complete records for a significant period (Sep’64 to Aug’16 (445 months)) and the corresponding rainfall data was segregated for analysis (see worksheet “Selected Combined” in file “Station Data.xlsx”). Gaps in the data for individual months for individual stations were filled to create continuous record and to avoid the issue of NAs when working with the data in R (highlighted yellow in worksheet “rainfall\_by\_station” in file “Station Data.xlsx”). Gaps in month *t* are filled by taking the value for the same month in the previous year (*t-12*) for the same station as a ratio of the sum of the values for all other stations in the previous year (*St-12*) and applying that ratio to the current month’s data for all stations (*St*), as follows: (1.1)

Figure 1: Histograms of monthly rainfall data for UK regions (left) and UK weather stations (right)

Histograms of both the monthly regional data and the monthly weather station point data show non-normal distribution of rainfall values, bounded by zero with a positive skew due to small numbers of months with very high rainfall (Figure 1). Mean monthly rainfall for all regions for the full period of 1836-2021 is 88.55mm, with a standard deviation of 51.33mm. The ‘Scotland W’ region has the highest mean monthly rainfall at 131.78mm, with ‘East Anglia’ the lowest at 51.12mm.

1. **Exploratory Spatio-Temporal Data Analysis**
   1. Spatial and Temporal Characteristics

Summary visualisation of the full UK Regional Rainfall dataset is challenging due to the sheer number of months and due to there being 10 regions. A sample of the data for a shorter period for two of the regions was chosen and plotted as a time series (Figure 2). The data shows considerable variation from month to month, little evidence of a distinct long-term trend, which is consistent with the findings of Lee (2020) [7].

A picture containing graphical user interface

Description automatically generatedGraphical user interface

Description automatically generated with medium confidence

Chart, line chart, histogram

Description automatically generatedFigure 2: Plots of monthly rainfall time series for Scotland N and East Anglia regions for Jan 2000 to Dec 2021

Plotting the monthly UK weather station data reveals a similar pattern to the regional data, with significant variation from month-to-month, no obvious long-term trend up or down, and suggestions of seasonality. Plotting the annual averages showed no obvious patterns (Figure 3), indicating there has been no overall change in UK annual rainfall levels on the past 50-60 years, which is consistent with the findings of Jenkins, et al. (2009) [8].

2D, 3D and dynamic scatterplots of the UK weather station point data do not show any clear relationships between average rainfall levels and latitude, longitude, or altitude, other than a hint that rainfall is higher the further west and north.

Figure 3: Average annual rainfall for all UK weather stations, 1965 to 2015

Text

Description automatically generated with medium confidenceA picture containing text

Description automatically generatedThis relationship is reinforced by examining heatmaps ordered by latitude and longitude, which confirm highest rainfall levels in the north and west (Figure 4). Latitude appears to be a greater factor than longitude. This could be due to the shape of the UK, which has a greater north-south extent than east-west. These heatmaps also show that the spatial variation in rainfall is greater than the temporal, as most of the rows are consistent in colour.

Figure 4: Heatmaps of rainfall for UK weather stations ordered by latitude (left) and longitude (right)

Chart, scatter chart

Description automatically generatedMapping of the annual average rainfall figures for the weather station point data also confirmed the pattern of wetter in the north and west, drier in the south and east and away from the coasts. Eskdalemuir in Scotland records the highest annual average rainfall (140mm); Cambridge records the lowest (46mm). This is in line with expectations based on knowledge of the south-westerly prevailing weather pattern in the UK, which brings rain in from the Atlantic with the central hills and mountains of England and Scotland creating a rain shadow effect to the east. Choropleths for each month of 1984, which was identified by analysis in Excel to be year with each region most closely matching their annual average rainfall, also clearly shows that the western and northern regions of the UK are the wettest, although ‘N. Ireland’ is relatively drier than nearby ‘Scotland W’. The driest regions are towards the east and the south. Looking just at 1984, there is an indication of seasonality with more rain in the autumn/winter than spring/summer.

* 1. Spatial and Temporal Autocorrelation

Global Moran’s I was calculated to test the spatial autocorrelation of rainfall levels across the regions of the UK. This had to be done for GB, i.e., excluding N. Ireland, which otherwise created an empty neighbour set as it is not connected to the rest of the UK. A Global Moran’s I value of 0.516 was calculated within a range of -0.593 to 1.018. The moran.test and moran.mc functions both gave p-values <0.05, confirming statistically significant autocorrelation for the regional data, leading to the conclusion that rainfall in one region is more similar in neighbouring regions than those further away.

Figure 5: Annual average rainfall for UK weather stations, 1965 to 2015

The lowest local spatial autocorrelation values are in the ‘S Wales & England SW’ and ‘England NW & N Wales’ regions. These are wet regions in the west with long borders with dry regions in the east, hence low autocorrelation. The highest local Moran’s I values are seen in ‘East Anglia’, a dry region in the east bordering other dry regions, and ‘Scotland N’, a wet region in the north-west bordering other wet regions. Based on unadjusted p-values, only ‘Scotland W’ and ‘East Anglia’ have significant local Moran’s I, but adjusting the p-values using the Bonferroni method shows no regions with statistically significant local Moran’s I.

Spatial autocorrelation in the UK weather station point data was analysed using a semivariogram. Results show a scattered result but there is an indication that rainfall levels at weather stations that are closer are more similar than those further away. No clear results were seen from the directional variograms, although there are hints of anisotropy in that not all the semivariograms look the same, with the semivariance varying more with distance in the 0o and 135o plots than the 45o and 90o plots. This gives a weak indication that spatial autocorrelation is stronger in the north-south and northwest-southeast directions than in other directions, which is broadly consistent with the findings from the analysis of spatial characteristics.

Chart, scatter chart

Description automatically generatedTemporal autocorrelation within the regional dataset shows it to be generally weak, with one month’s rainfall less strongly correlated to previous month’s rainfall than seen with UK temperatures. The ‘Scotland N’ region shows the highest temporal autocorrelation with a PMCC of 0.306 (Figure 6); Midlands shows lowest PMCC of 0.081 (1 month lag interval).

The annual data for UK weather stations also shows week temporal autocorrelation (PMCC = 0.121), indicating one year’s rainfall is not significantly related to the previous year’s rainfall.

Figure 6: PMCC for Scotland N region (one month lag)

1. **Methodology and Results**
   1. ARIMA

A picture containing text, people

Description automatically generatedTo analyse the data further, the ‘Scotland N’ regional data was selected as this showed the strongest temporal dependency. A shortened version of the ‘Scotland N’ data covering the period 1990 to 2021 is used to make the analysis more manageable. This time series is decomposed into its trend, seasonal and residuals components to determine stationarity and then the Box-Jenkins approach to ARIMA modelling is followed to try to fit and test a model for forecasting.

It is expected that the ‘Scotland N’ time series is not stationary due to it containing an element of seasonality. Decomposition of the ‘Scotland N’ time series was done using Seasonal and Trend Decomposition using Loess (STL). Various values for the t.window parameter were tried, with a value of 25 chosen as this resulted in the smallest remainders. The results (Figure 7) show a trend component with no clear pattern, but a clear seasonal component. However, remainders are still high at approximately +/-100mm, compared to the seasonal component, which is +/-50 to 60mm. Confirmation of a clear seasonal component indicates that seasonal differencing will be required for ARIMA. With the first step – Exploratory Data Analysis – already completed, the Box-Jenkins approach moves directly to the next step with analysis of the autocorrelation factor (ACF) and partial autocorrelation factor (PACF) plots and differencing to provide more insight into potential ARIMA model parameters.

Figure 7: Decomposition of Scotland N regional rainfall data 1990 to 2021 using STL

*Undifferenced:*

* *ACF peaks at lags of 12, 24, 36, 48 months, negative peaks at 6, 18, 30, 42 months: non-stationary with peaks at fixed intervals – seasonal differencing with order 12 required*
* *PACF significant positive results at lag 12 and 24, negative at lags 5, 6, 18 – again demonstrates seasonality and suggests need to include a seasonal autoregressive term in ARIMA / STARIMA*

*Seasonal Differenced*

* *ACF seasonal pattern essentially removed beyond lag 12; significant autocorrelation remains at lags 1 and 12 – suggests seasonal MA = 1 and nonseasonal MA = 1*
* *PACF significant at lags 1, 12, 24, 36, 48*

*Non Seasonal Differenced*

* *ACF significant negative at lags 1, 6, 18; significant positive at lags 12*
* *PACF multiple negative at lags 1 to 11*

*Undifferenced ACF suggest seasonal differencing*

The ACF plot of the ‘Scotland N’ regional data shows a seasonal pattern with positive peaks at lags of 12, 24, 36, 48 months, and negative peaks at 6, 18, 30, 42 months. This is the hallmark of a non-stationary time series and strongly suggests seasonal differencing with order 12 is required to make the time series stationary, which is a prerequisite of a linear regression model such as ARIMA. This was done and the ACF re-run, producing a plot showing that seasonal differencing has largely removed the seasonal component from the data. Significant ACF remain at lags 1 and 12

Seasonal ACF suggests ARIMA(0, 0, 1)(0, 1, 1)12 or ARIMA(0, 0, 2)(0, 1, 1)12

Back to regional data, ACF for Scotland N shows seasonality with statistically significant positive autocorrelation

PACF for Scotland N region shows

No statistically significant PACF for annual UK weather station data, suggesting annual rainfall is random from one year to the next (not looked at individual weather stations)

Spatio-Temporal Analysis

Spatio-temporal ACF for GB regions shows similar seasonal pattern as for temporal analysis. No statistically significant STPACF

ST semivariogram point UK weather station point data shows the semivariance increases rapidly with increasing temporal separations and then becomes more complex with some peaks and troughs. No equivalent increase in semivariance with increasing spatial separations, but spatial and temporal separations are not comparable.

* 1. Artificial Neural Networks

Form of Supervised Learning – labelled training data used to predict labels for unseen data

1. **Discussion and Conclusions**

Spatial and temporal analysis of the UK rainfall time series data shows it to be quite weakly correlated spatially and temporally. The wettest regions are in the north and west of the UK and there is some seasonality with the wettest months in the autumn and winter, but with no apparent long-term trend.

References

[1] Wiener, N. (1949). Extrapolation, interpolation, and smoothing of stationary time series, with engineering applications. Technology Press of the Massachusetts Institute of Technology

[2] Box, G.E.P. & Jenkins, G. M. (1970). Time series analysis : forecasting and control. Holden-Day

[3] Mitrea, C. A., Lee, C. K. M., Wu, Z. (2009). A Comparison between Neural Networks and Traditional Forecasting Methods: A Case Study. International Journal of Engineering Business Management, Vol. 1, No. 2, p 19-24

[3] Met Office (2022). UK and regional series. Available at https://www.metoffice.gov.uk/ research/climate/maps-and-data/uk-and-regional-series. Accessed 04/04/2022

[4] Hawkins, E., Burt, S., McCarthy, M., Murphy, C., Ross, C., Baldock, M., et al (2022) Millions of historical monthly rainfall observations taken in the UK and Ireland rescued by citizen scientists. Geoscience Data Journal, 00, 1– 16. Available from: https://doi.org/10.1002/gdj3.157

[5] Met Office (2022). UK climate districts map. Available at https://www.metoffice.gov.uk/research/climate/maps-and-data/about/districts-map. Accessed 04/04/2022

[6] Open Data NI (2022). OSNI Open Data – 50K Boundaries – NI Outline. Available at [https://www.opendatani.gov.uk/dataset/osni-open-data-50k-boundaries-ni-outline. Accessed 04/04/2022](https://www.opendatani.gov.uk/dataset/osni-open-data-50k-boundaries-ni-outline.%20Accessed%2004/04/2022)

[7] Lee, E.M. (2020). Statistical analysis of long-term trends in UK effective rainfall: implications for deep-seated landsliding. Quarterly Journal of Engineering Geology and Hydrogeology, 53, 587-597

[8] Jenkins, G.J., Perry, M.C. and Prior, M.J. (2009). The Climate of the United Kingdom and Recent Trends. Met Office Hadley Centre, Exeter