Course Project Data Science
A study to assess the impact of nitrate concentrations on river water in Northern Ireland

Adina Asim (adina.asim@student.uhasselt.be)
Mohammad Sazegar (mohammad.sazegar@student.uhasselt.be)
Aarvan Kaushik (aaryan.kaushik@student.uhasselt.be)
Rahmanda Wibowo (rahmanda.wibowo@student.uhasselt.be)

UHASSELT
KNOWLEDGE IN ACTION

Abstract

This study tried to analyze the association between rainfall and nitrate level (NO3-N) in river water in Northern Ireland, which is the region focused on agriculture and can attributed to a higher utilization of nitrate. The dataset was achieved by combining two data source from DAERA UK and open-meteo.com. Based on exploratory analysis, it is observed that higher amount of rainfall leads to higher amount of NO3-N. From linear mixed model, we observed that 1mm increase of rainfall can cause an increase of 1 mg/L of NO3-N in river water.

Background

In the last two decades, there have been many reports in the British Isles of increased concentrations of nitrate in surface waters. In 1991, the EU introduced the Nitrates Directive, which aimed to reduce water pollution caused or induced by nitrate from agricultural sources [1]. This is important because high nitrate levels in water (> 11.259 mg NO3-N/L) can cause serious health effects on human and aquatic ecosystem. There have been many studies in the past that observed a high level of nitrate in water that was caused by high precipitation of rainfall. Due to the fact that 75% of lands in Northern Ireland is used for agriculture, it can be an interesting study to analyze the same phenomenon in this specific region.

Objective

This study aims to analyze the association between rainfall and nitrate concentrations (NO3-N) in river waters, and to assess its overall impact on the quality of river water in Northern Ireland.

Dataset

The water quality dataset was retrieved from DAERA, UK [2]. The data was **repeated nitrate measurements from various monitoring sites in Northern Ireland from 1990 to 2018**, covering 141,431 rows and 17 variables in a JSON format which was converted into a CSV file. At the start of the study, the data was only recorded from 129 monitoring sites. From the year 2000 onward, 624 sites were present, but most of them had missing NO3-N (5% missing values of NO3-N). We wanted to see how many sites had observation every month over 19 years, and we found out only 94 sites had completed the study between 2000 and 2018 (21354 observations). **Rainfall dataset was retrieved from open-meteo.com historical weather API** which is based on ERA5 grid dataset [3]. To enrich the water quality dataset with this rainfall data, first we summarised the longitude, latitude, and the first and last date measurement of NO3-N for each monitoring site. Then, we run the script to make multiple batches of API requests with these parameters to open-meteo.com, export them as multiple CSV files, and perform some data manipulation techniques to combine these rainfall data with the water quality dataset.

Table 1. Data description of the final dataset.

Variable	Description
Site Code	Code for river monitoring sites (character)
latitude & longitude	Coordinates of river monitoring sites (numeric)
Date	Date of the measurement (date string with format yyyy-mm-dd)
NO3_N_MGL	Nitrate nitrogen level (mg/L) in the monitoring site (numeric)
rainfall	Amount of rain (mm) at the day of the measurement (numeric)

Methodology

Initially, we identify in which regions very high or very low values of nitrate and rainfall lie. This assessment was made with the help of **hotspots** (regions where the concentration is high) and **coldspots** (regions where the concentration is low). To observe the cluster pattern of nitrate concentrations and rainfall in the given regions, **K-means clustering** was performed. Finally, to draw a marginal inferences by taking into account multiple measurements for each site given in the study, a **linear mixed model** has been taken into account. We estimated random and fixed effects for mean nitrate in a month for each site specifically. The statistical model is formulated as follow:

$$log(Y_{ij}) = \beta_0 + \beta_1 MeanRain_{ij} + b_{0i} + b_{1i} MeanRain_{ij} + \epsilon_{ij}$$

where $i = 1, ..., n$ is the monitoring site index, $j = 1, ..., 12$ is the month, Y is the NO3-N outcome (log transformed), β_0 and β_1 are the fixed effect for their respective regressors, $b0$ and $b1$ are random intercept and random slope for mean rain, and $\epsilon_i \sim \mathrm{N}(0, \sum_i)$.

Data Visualization

Figure 1. Hotspots visualization of average NO3-N level and rainfall in Northern Ireland

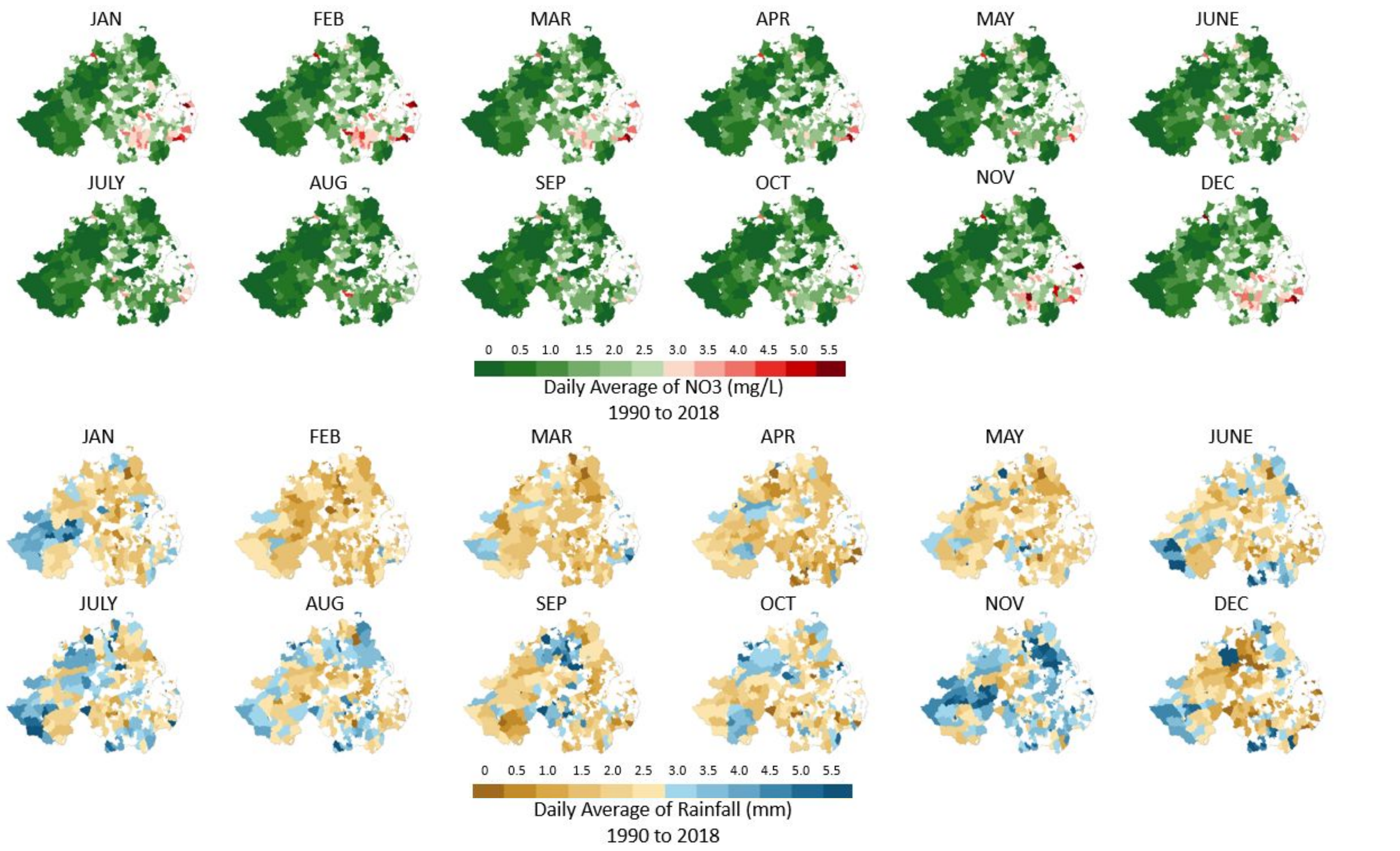


Figure 2. Topology map of Northern Ireland

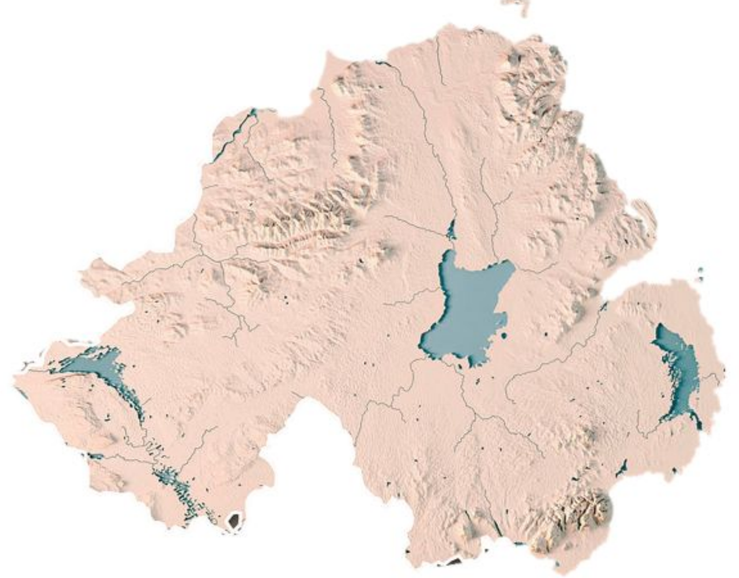


Figure 3. Individual site plot mean NO3-N in 2018.

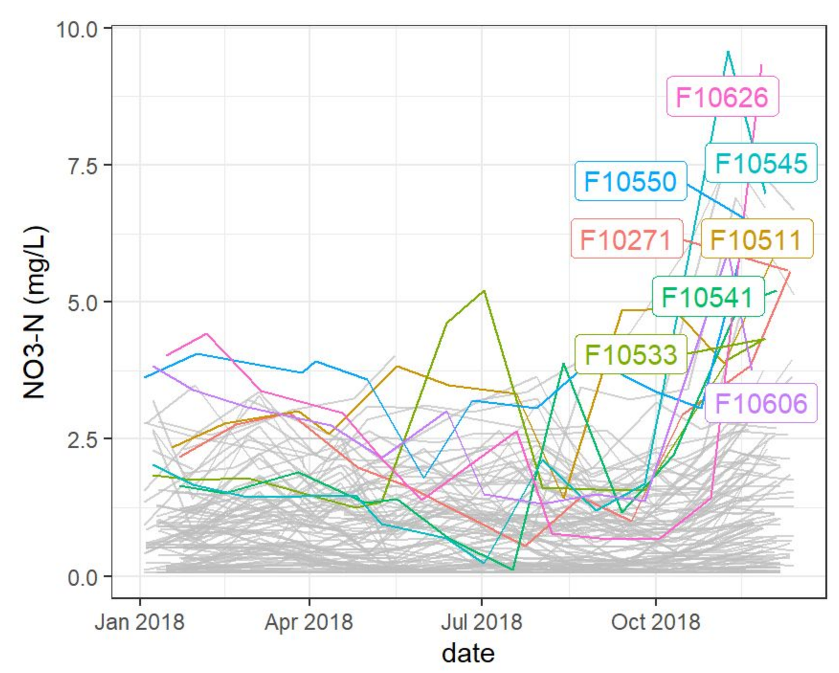
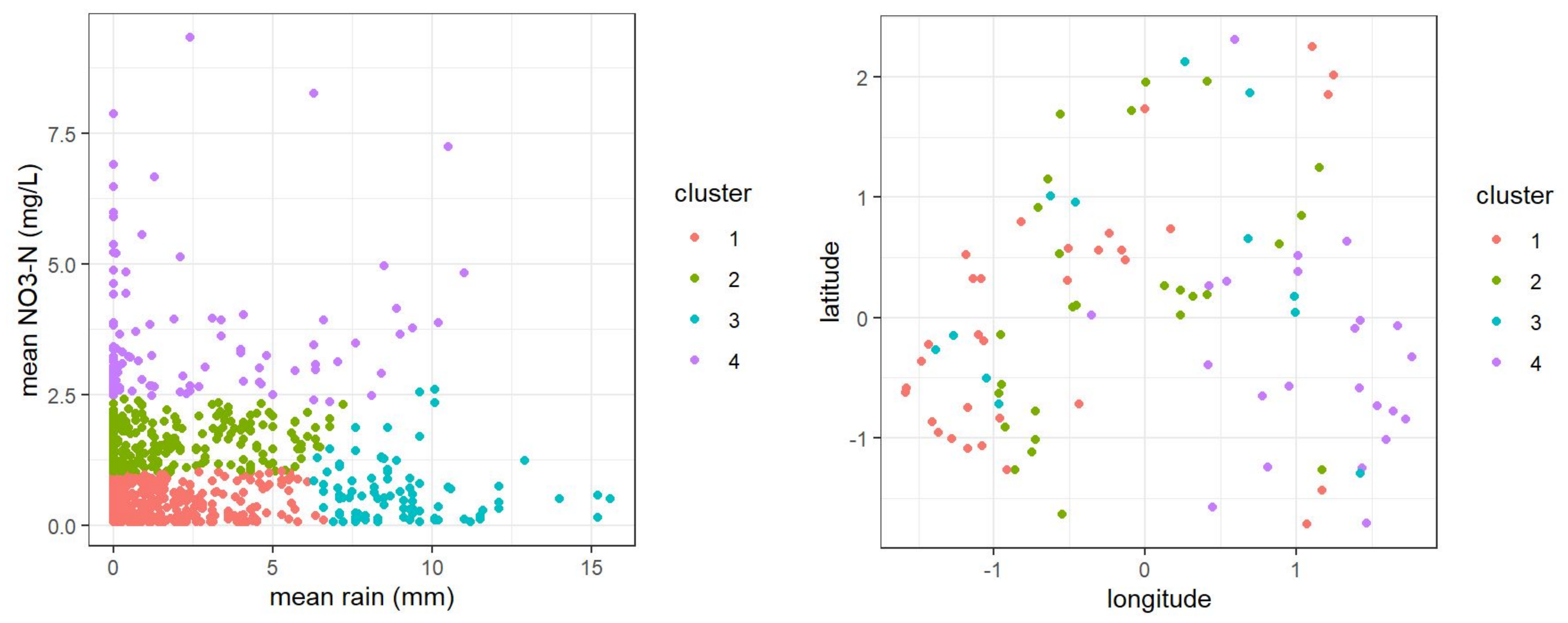


Figure 4. Result of k-means clustering for observations in 2018.



Linear Mixed Model

From the table, it is evident that for every increase of 1 mm of average daily rain in a month cause increase of exp(-0.01115) = 1 mg/L of nitrate level in river water at 5% of significance level.

Effect	Estimate	Std. Err	Pr > |t|
Intercept	-0.3937	0.1132	0.0008
mean_rain	-0.01105	0.001667	<.0001

Conclusion & Discussion

- From the hotspots analysis, it is observed that the region with a moderate level of rainfall also have a moderate level of nitrate in water. The region with a high level of rainfall has a low level of nitrate, but around these specific regions the nitrate level is high in a distance. Interestingly, cluster 4 from k-means consists of these regions.
- Without including coordinates information in k-means, the clusters still show the topography differences of Northern Ireland.
- Based on the linear mixed model result, it is shown that there is a positive relationship between the average daily rain and the nitrate level in river water.
- Spatial analysis method is another option to analyze this type of dataset to take into account the correlation between clusters of regions.
- Cluster 3 and cluster 4 have different characteristics in terms of nitrate level and rainfall level. Taking into account these differences by adding topography information to the dataset might provide more meaningful insight to the study.

References

1. Northern Ireland Statutory Rules (2019). The Nutrient Action Programme Regulations (Northern Ireland) 2019 No. 81. https://www.legislation.gov.uk/nisr/2019/81/contents/made.
2. DAERA UK (2022). River Water Quality Monitoring 1990 to 2018 - Nitrate, Northern Ireland. Accessed 15 December 2022. https://opendata-daerani.hub.arcgis.com/.
3. Hersbach et al (2018). ERA5 hourly data on single levels from 1959 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). (updated daily), 10.24381/cds.adbb2d47.