MATH60139/70139: Spatial Statistics
Assessed Coursework

Instructions:

This single piece of assessed coursework is worth 10% of the module and is **due at 1pm on Tuesday 18th March** via Blackboard the usual way. You should perform your analysis in R[1], but **submit a single PDF** with your answers and plots to each question. The final (mastery) question is only for those taking this as a 4th year or MSc module. **Submit your R code at the end of the PDF as an appendix**, please don't take screenshots of R but rather copy it over into your document such that your marker can copy it back into R and run it.

You have each been allocated your own unique data set to analyse, to find out which go to the "Project Allocations" document in the "Assessments and Mark Schemes" tab on Blackboard and match to your CID number. Then go into the "Project Data" folder and download the data (an Excel file) for your corresponding Project Number into the same working directory as your R script. You can then load the data using the following command:

```
install.packages("readxl")
library(readxl)
my_data <- read_excel("ProjectXX.xlsx")
```

In the above, replace XX with your actual Project Number. Your dataset is a monthly averaged climate variable taken from the UK Met Office historic record of station data, where you will study one of:

- Mean daily maximum temperature (tmax)
- Mean daily minimum temperature (tmin)
- Total rainfall (rain)
- Total sunshine duration (sun)

Your climate variable to be studied is given in the Project Allocations document. To view your spatial data at each weather station then enter one of `my_data$tmax`, `my_data$tmin`, `my_data$rain` or `my_data$sun`. To view the spatial coordinates of each weather station then enter `my_data$lat` and `my_data$lon` for the latitude-longitude coordinates for each corresponding weather station. For reference you can view the year and month that you are studying by entering `my_data$year` and `my_data$month` respectively. If your data is sunshine duration then it will have some values of "NA" - please remove these by running the following code:

```
my_data <- na.omit(my_data)
```

but do **not** do this if you are studying maximum/minimum temperature or total rainfall which have no NAs. For reference, the source to the data is here.

Finally, note that to find the Euclidean distance between two locations for this coursework simply assume this is given by the square norm of the differences in latitude and longitude respectively (i.e., you can model in $\mathbb{R}^2$ and ignore the effect of curvature of the Earth which is reasonable over small enough regions like the UK). Also, don't worry if you occasionally get small negative values for rainfall for your mapped predictions, you don't need to correct these. It's also fine for you to use and adapt code from the lab sessions. The total for Year 3 (Questions 1-10) is 20 marks, and for Year 4 / MSc (Questions 1-10 + Mastery Question) is 25 marks. Marks will then be scaled to 100 for everybody.

[1]Students on MSc Applied Maths may perform their analysis in Python, but all other students must use R

Questions. Using your dataset only:

1. Use *simple* Kriging to obtain a prediction of the value of your climate variable at the geographic location of Imperial College London (South Ken campus) with coordinates of latitude: 51.499, and longitude: -0.175. Use a Gaussian covariance kernel $\rho(s,t) = \sigma^2 \exp(-\beta||t-s||^2)$, with σ^2 equal to the sample variance across the stations, and set $\beta = 1.5$. Set the mean to be the sample mean across the stations. [1 mark]

2. Find the corresponding prediction error at this location. [1 mark]

3. Find also a prediction and prediction error at the geographic location of University College London (Bloomsbury campus) given by latitude: 51.525, and longitude: -0.134. [2 marks]

4. Explain in one sentence on why one of the prediction errors (at Imperial College London or University College London) is greater than the other. [1 mark]

5. Repeat Questions 1 and 2 to find the predictions and prediction error if $\beta = 2$. [1 mark]

6. Explain in one sentence on why one of the prediction errors ($\beta = 1.5$ or $\beta = 2$) is greater than the other. [1 mark]

7. Repeat Questions 1 and 2, with $\beta = 1.5$, but this time use *ordinary* Kriging assuming a constant but unknown mean. [2 marks]

8. Explain in one sentence on why one of the predictions errors (from simple or ordinary Kriging) is greater than the other. [1 mark]

9. Extend your ordinary Kriging predictions from Question 7 over the entire UK using a grid of coordinates. Plot your predictions and prediction errors on a map of the UK, where you can use packages such as `ggplot2` to produce nice maps. [2 marks]

10. In Questions 7–9 you used ordinary Kriging with a constant but unknown mean. For the UK, we do not expect the mean to be constant in various climate variables including temperature, rainfall and sunshine hours, as displayed in Figure 1 below. Motivated by this, build a *universal* Kriging predictor for your climate variable which varies spatially as a function of either latitude or longitude as follows, depending on your climate variable:

    - If your variable is max temperature, min temperature, or sunshine duration, then use the mean function: $\mathbb{E}(X_t) = \alpha_1 + \alpha_2 \text{lat}(t)$ where $\{\alpha_1, \alpha_2\}$ is unknown and $\text{lat}(t)$ is the latitude of location t. This mean function can therefore model temperatures and sunshine duration that change with latitude (going north or south) in the UK as depicted in Figure 1.
    - If your variable is rainfall then use the mean function: $\mathbb{E}(X_t) = \alpha_1 + \alpha_2 \text{lon}(t)$ where $\{\alpha_1, \alpha_2\}$ is unknown and $\text{lon}(t)$ is the longitude of location t. This mean function can therefore model rainfall that changes with longitude (going east or west) in the UK as depicted in Figure 1.

    As in Questions 1 and 7, use a Gaussian covariance kernel with $\beta = 1.5$ and σ^2 equal to the sample variance across the stations.

    (a) Find the prediction and prediction error at Imperial College London (as in Questions 1 and 2). [2 marks]

    (b) Plot the prediction and prediction error across the UK (as in Question 9). [3 marks]

    (c) Plot the difference between the universal and ordinary Kriging predictors. [1 mark]

    (d) Comment on the key differences between your plots using ordinary and universal Kriging (two sentences max). [2 marks]

11. **Mastery Question (Year 4 and MSc only):** Repeat Question 10(b) (universal Kriging) to produce a map of predictions and prediction errors in the UK, but this time use an *anisotropic* covariance function given by: $\rho(s,t) = \rho_0(||A(t-s)||)$ where $\rho_0(d) = \sigma^2 \exp(-\beta d^2)$ is the Gaussian covariance kernel as before ($\beta = 1.5$ and σ^2 equal to the sample variance), and where $A = DR$ is the product of a diagonal matrix D and a rotation matrix R given by:

$$D = \begin{pmatrix} \rho & 0 \\ 0 & 1/\rho \end{pmatrix}, \rho = 2, \quad R = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

If your variable is max temperature, min temperature or sunshine duration then use $\theta = \pi/3$. If your variable is rain then use $\theta = -\pi/6$. These choices are motivated from Figure 1 below. Comment on how and why your predictions and prediction errors have changed compared with Question 10(b) (two sentences max). [5 marks]

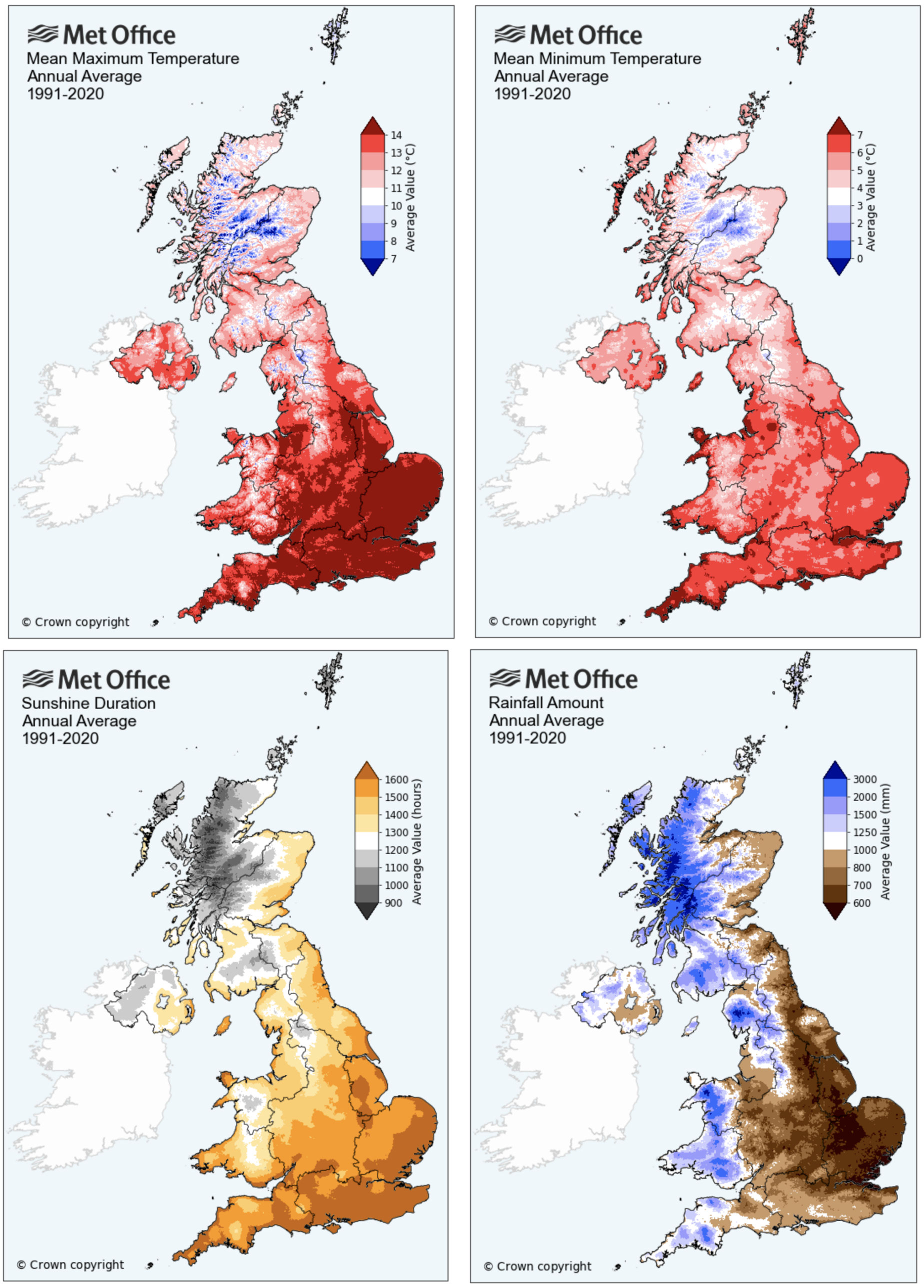

Figure 1: Average climate variables between 1991-2020 in the UK.