Introduction to Environmental and Ecological Statistics

Environmental and Ecological Statistics is an incredibly broad term covering any form of statistics applied to environmental and ecological issues. Key themes include climate change, environmental regulation (e.g. water and air quality) and biodiversity monitoring. This course focuses on this theme rather than a particular type of statistical methodology. We will look at a variety of statistical methods, some of which you will know, and some which will be new.

The environment is (sometimes literally) a burning issue in the 21st century. This brings increased focus and interest in statistics as a subject, and how we are working to handle topics like climate change. Below, we see some examples of how climate change reporting can be presented.

BBC News article on climate change

On 14th January 2020, BBC News published an article titled "Climate change: Where we are in seven charts and what you can do to help". You can read this by clicking the link below:
https://www.bbc.co.uk/news/science-environment-46384067

Exercise 1

Please read the BBC News article linked to above, and consider critically how the information is presented. Specifically, what are the good and bad aspects of the graphs in the article? Does the article make you think about climate change any differently?

Channel 4 programme "2022: The Year from Space"

The Channel 4 programme "2022: The Year from Space" was first broadcast in 2022 and contains some useful illustrations of environment-related news stories.
https://www.channel4.com/programmes/2022-the-year-from-space/on-demand/74702-001

Exercise 2

Please watch the Channel 4 video from 38:41 to 44:00 minutes in. What type of data are being generated here? How might they be analysed? How might the results of these analyses be presented to governments and other stakeholders?

1 Where's the statistics?

We are interested in measuring, sampling or monitoring environmental and ecological data, including variation and uncertainty. This includes detecting and modelling environmental trends, including trends in time and space, modelling and understanding extreme data. We also wish to evaluate environmental regulation and policy, and risk assessment.

We want to understand changes in the environment, in either time, space or both. Are things getting better or worse? Where, when and by how much? What is going to happen next? Where do authorities need to take action, and how can we check if existing actions are working? Also, we should consider relationships between environmental variables (and other variables where necessary).

In general, there are no techniques that are unique to environmental statistics. However, the data used tend to be characterised by strong spatial and temporal elements, and often also high variability.

Our skills in presenting and communicating data are also crucial. We need to be able to explain our findings to the public and show them why our work is important. Below is an example of reporting of a winter storm in 2023, which makes use of plots to tell the story.

The Courier article on Storm Babet

On 20th October 2023, The Courier published an article titled "Storm Babet: Timeline of devastating rainfall in charts and maps". This contains some interesting examples of the use of plots to tell a developing environmental story. You can read this by clicking the link below:
https://www.thecourier.co.uk/fp/news/4788874/storm-babet-timeline-of-devastating-rainfall-in-charts-and-maps/

Exercise 3

Please read the Courier article on Storm Babet linked to above and think about the way that the information is presented.

Some examples of environmental & ecological statistics include:

- **Decision making:** Is it safe to eat fish from a particular river?
- **Prediction:** What is the trend in temperature? Can we predict its level in 2060?
- **Regulation:** Have emission control agreements reduced air pollutants?
- **Understanding:** How did sea levels change over the past 100 years?

Example: Air quality

We will illustrate this with an example relating to air quality.

Figure 1: Photo of smog in Shanghai.

Only **one person in ten** lives in a city that complies with the World Health Organisation Air quality guidelines.
Fine particular matter was associated with an estimated 2,000 premature deaths and 22,500 lost life years in Scotland in 2010. There are 38 "Air Quality Management Areas" (AQMAs) which breach or are likely to breach legal limits. The Cleaner Air for Scotland strategy seeks to reduce air pollution across Scotland. It aims to achieve the "ambitious vision for Scotland to have the best air quality in Europe".
There are 99 air quality monitoring stations in Scotland capturing $PM_{2.5}$, PM_{10}, NO_2, NO_X, SO_2 and O_3.

The map below shows the placement of the monitors. Live data are available at Scottish Air Quality.

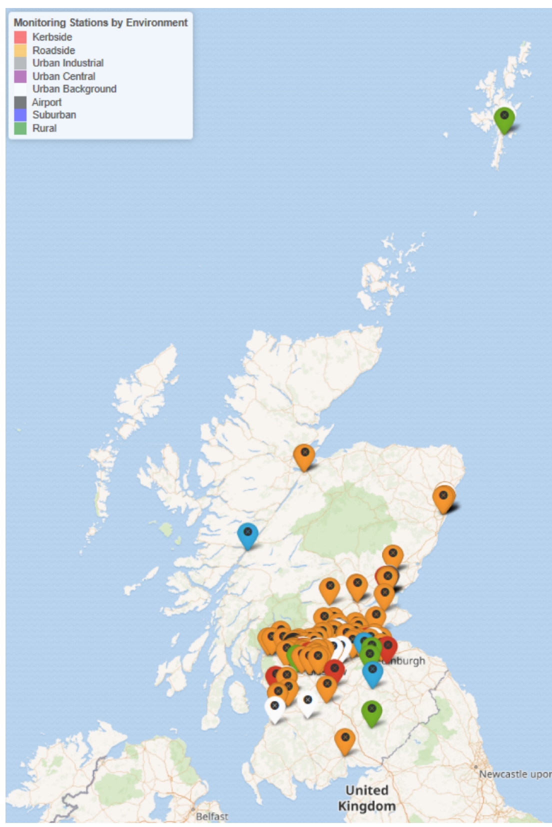


Figure 2: Monitoring station map.

The data from these monitors can be used to estimate the pollution levels across Scotland.

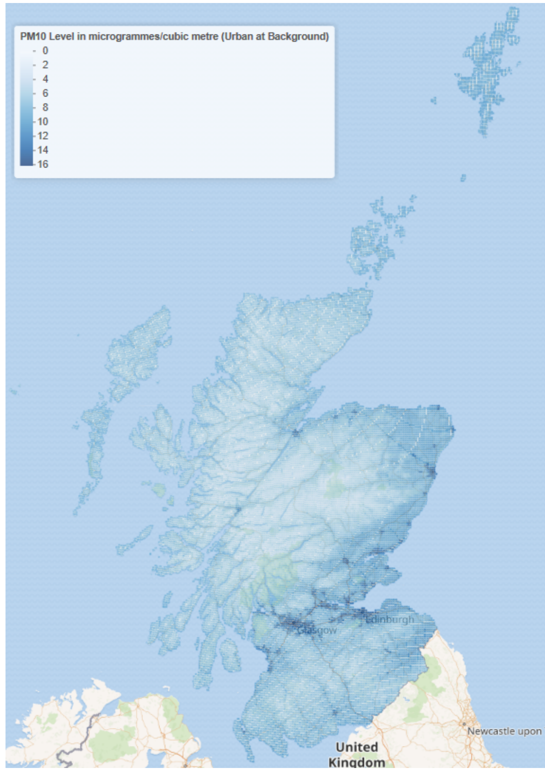


Figure 3: Estimated PM_{10} levels.

Exercise 4

The example above shows a map of estimated PM_{10} levels across Scotland. What information do we not have, which we would normally expect to see here?
Solution
The main thing missing here is the lack of uncertainty information. It seems that most of the monitoring stations are in the Central Belt of Scotland (where most of the pollution might be expected), so we'd expect to see lower uncertainty there than in areas further from a monitoring station.
We might also like to know more about the temporal aspects of the data. Is the prediction map for the same time that the data were collected, or do frequencies of collection vary across the monitoring stations?
Also, are these predictions just from the station data, or do we have other sources that are combined with these? (For at least some variables, we might have satellite data available that provide us with measurements across a fine spatial grid, which we can combine with

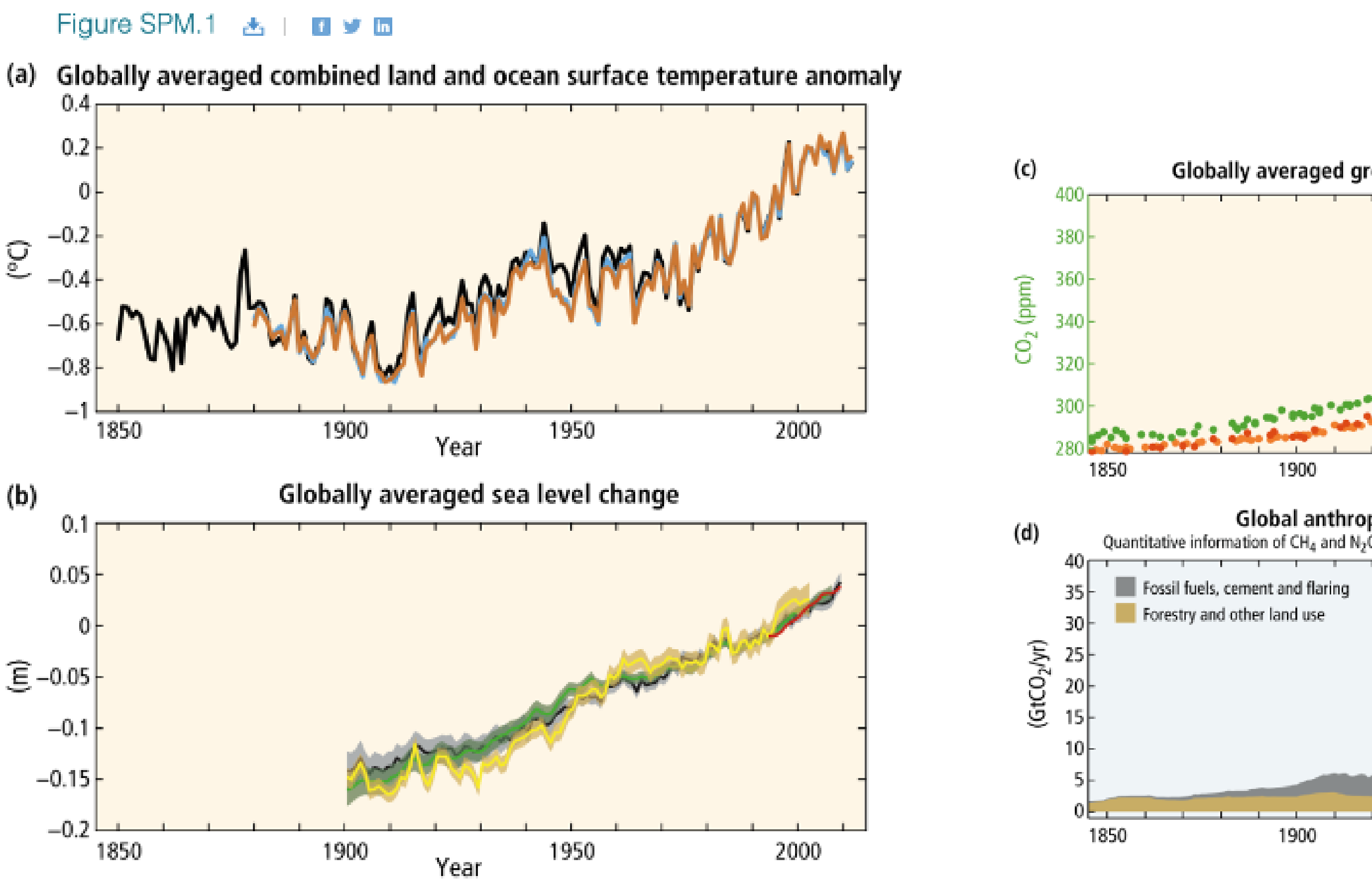
monitoring station data through "data fusion".) In any case, we'd like to know how these predictions are made.
You may have thought of some other ideas too. We can discuss these in the lectures.

2 Quantification

Understanding and measuring quantities is a fundamental part of all science, not just statistics. As scientists, we use data to understand the process which we are investigating. These data have two main sources of uncertainty or error:

- **Inherent variability** of the process itself (the thing we are measuring is variable).
- **Imprecise knowledge** of the process (our measurements may not be accurate).

The plots below illustrate the trends in several climate change measures. Both sources of variability will be present, but how much of each?



A big part of our role as statisticians is to ask questions of both our data and our models. How were our data collected? Are they representative of the population? How much uncertainty do we have? Are our models valid? Are the assumptions reasonable? Does the model make sensible predictions? How much uncertainty do we have in our results? These skills are particularly crucial in applied areas such as environmental and ecological statistics.

Example: Arctic sea ice cover

Submarines have been used to measure Arctic sea ice. Over time, the ice has shrunk both in terms of thickness and extent. We may soon see ice-free summers, which will have a devastating impact on sea life.

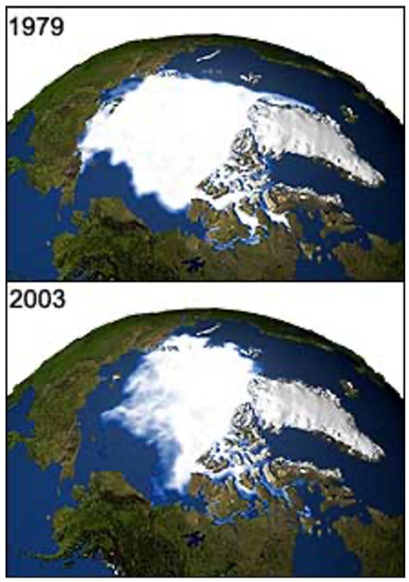


Figure 4: Arctic ice coverage in 1979 and 2003.

This interactive map provides an illustration of the changes.

Exercise 5

How might we quantify the trend in Arctic sea ice cover, and what problems might we encounter in aiming to do this?

Solution

We might want to present this as a graph of total estimated sea ice cover over time, across many years. Some problems with this approach could be data availability over time (satellites or submarines not covering the whole area, or only at certain timepoints, and lack of data in earlier years), and the difficulty in estimating coverage at any timepoint from these measurements — statistical approaches would be vital to estimate this and

the associated uncertainty.
Note that the use of before and after images (like in the above example) might not be thought of as the best idea in statistics (since they could be misused, with particular years cherry-picked to back up an incorrect claim), but such images can be useful in illustrating a trend to the general public, as long as this is also presented with additional complete information about the general patterns.

3 Understanding our data: trends

Much of the statistical analysis of environmental and ecological data will focus on identifying **trends**. The statistical definition of a trend is *a long-term change in the mean level.* Trends aren't restricted to being linear, and we will also look at examples of non-linear trends, and changepoints.

Trends generally tend to focus on the **mean** of the data. However, we may also be interested in observing other aspects of the statistical distribution. **Extreme value theory** looks specifically at the limits of our distribution, and focuses on rare (or extreme) events.

We'll cover this more later in the course.

4 Policy and legislation

A great deal of environmental statistical research is funded by governments and regulatory bodies. They need to know where the biggest challenges lie so that they can allocate their resources appropriately. Evidence-based policy relies on measuring changes and also evaluating the impacts of existing policies. However, the environment will be one of many competing policy areas, and every government will prioritise it differently.

Example of international cooperation and negotiations: COP

Figure 5: Team photograph from COP28 in 2023.

COP is the United Nations Climate Change Conference that takes place annually (held in Glasgow in 2021), where most countries' governments come together to try to agree on goals to limit global temperature rise and achieve net zero emissions. The most recent conference was COP29, which took place in Baku, Azerbaijan, in November 2024.
One of the key achievements of COP29 was an agreement by wealthier countries to pay at least US $300 billion per year to less wealthy countries to help finance their mitigation of climate change impacts, and to transform their economies away from fossil fuels. However, there was limited progress in other sectors.
A report on COP29 and its implications for the UK is available from the UK's Climate Change Committee here.

Environmental policy tends to use very specific language: *objectives, targets, guide values, standards, reductions relative to a baseline.* Policy often prescribes monitoring quantities of interest over space and/or time. Quantities of interest include *water, air and noise pollution, waste management, radioactive substances, biodiversity and animal and plant species.*

Legislation is the legal framework used to implement policy. Most legislation focuses on setting targets or safe levels for the pollutants. A number of regulatory bodies exist specifically to monitor such things, e.g. Scottish Environment Protection Agency (SEPA, for Scotland), the

Environment Agency (EA, for England) and the European Environment Agency (EEA, for the EU). As well as regulatory bodies, the Climate Change Committee advises governments within the UK on reducing emissions and addressing climate change impacts.

Much of our data come from *routine monitoring* of our environmental quantities of interest. Government agencies often make these data available to researchers and/or the public. These data are used to assess compliance with legislation as well as to identify environmental trends and their potential impacts on society.

Example of Natural Capital and Ecosystem Assessment (NCEA) programme

The UK Government's Natural Capital and Ecosystem Assessment (NCEA) programme aims to provide a comprehensive understanding of the state of the natural environment to inform better decision-making for its protection and enhancement. By integrating data collection, mapping, standardized protocols, and detailed reporting, the NCEA seeks to halt biodiversity loss, support climate adaptation and mitigation, and ensure sustainable management of natural capital for future generations. This evidence-based approach helps identify environmental trends, risks, and opportunities to guide policy and conservation efforts.
The brochure below provides an overview of some of the key resources produced by the NCEA.

Exercise 6

Download the NCEA. brochure and read to three case studies and identify (1) the problem that is being address (2) the proposed solution and its (3) policy impact. Then, think about how would you present the data.

References

Some general reference books that you may find useful include the following:

- Piegorsch, W. W., & Bailer, A. J. (2005). *Analyzing environmental data*. Wiley. (Available from the University Library as an e-book here.)
- Barnett, V. (2004). *Environmental statistics: Methods and applications*. Wiley. (Available from the University Library as an e-book here.)
- Manly, B. F. J. (2001). *Statistics for environmental science and management*. Chapman & Hall/CRC. (No e-book available, but a physical copy is available from the University Library.)