Pave the World? The soil ecosystem under threat by expanding cities and towns.

Freddy Marten, BSc (Hons) Economics and Mathematics

School of Computing and Communications

Lancaster University

A dissertation submitted in partial fulfillment of the degree of

Master of Science in Data Science

September, 2024

Pave the World? The soil ecosystem under threat by expanding cities and towns.

Freddy Marten, BSc (Hons) Economics and Mathematics.

School of Computing and Communications, Lancaster University

A dissertation submitted in partial fulfillment of the degree of *Master of Science* in Data Science. September, 2024.

Abstract

The release and intake of soil greenhouse gases (GHG) is forgone when urban areas expand and soil is sealed over. Here, we present an initial estimation of the forgone emissions of CO_2 and N_2O as well as the additional emissions of CH_4 in the UK by 2018. We found that sealed surfaces are causing effective intake of 64.5 ± 24.7 Mt CO_2 yr^{-1} and 14.5 ± 4.98 kt N_2O yr^{-1} along with emissions of 1.83 ± 1.79 kt CH_4 yr^{-1}. At the 20-year time horizon, the CO_2 equivalent emissions of CH_4, 0.15 ± 0.15 Mt CO_2e yr^{-1} and equivalent intake of N_2O, 3.97 ± 1.36 Mt CO_2e yr^{-1} were an order of magnitude lower. Greater seasonal savings were seen in the winter and summer months with more extreme weather. 90.3% of the overall GHG CO_2e saving from sealing came from CO_2 in either cropland (31.1 ± 18 Mt CO_2 yr^{-1}, 45.6%) or grassland (30.5 ± 17.3 Mt CO_2 yr^{-1}, 44.7%). To reach an understanding of the overall effect of soil sealing on GHG emissions, additional side effects of urban area expansion must be accounted for.

Acknowledgements

I would like thank Dr. France Gerard, Dr. Andrew Singer and Dr. David Hofmeyr for their consistent - and extremely helpful - support throughout this project. Also, many thanks to Thea Moores for her enduring support.

Declaration

I declare that the work presented in this dissertation is, to the best of my knowledge and belief, original and my own work. The material has not been submitted, either in whole or in part, for a degree at this, or any other university. This dissertation does not exceed the maximum permitted word length of 20,000 words including appendices and footnotes, but excluding the bibliography. A rough estimate of the word count is: 11609

Freddy Marten

Contents

1 Introduction **1**

1.1 Terminology and definitions 1

1.2 Outline 2

2 Background **3**

2.1 Related work 3

2.2 Related Data 7

2.2.1 Soil Greenhouse Gas Flux databases 7

2.2.1.1 Soil Respiration Database 7

2.2.1.2 COSORE 8

2.2.2 Land Cover 10

2.2.3 Imperviousness 11

3 Methodology **13**

3.1 Method overview 13

3.2 Simulating GHG fluxes 15

3.2.1 Analytical results 18

3.3 Creating the unsealed map 20

3.4 Spatial resampling 23

3.4.1 Nearest-neighbour 26

3.4.2 Bilinear 27

3.4.3 Cubic 28

3.4.4 Aggregation 28
3.5 Accounting for roads 29
3.6 GHG flux relationship with temperature 30
3.7 Annual sealed (and unsealed) soil GHG fluxes 32

4 Results **35**
4.1 Annual saving in GHG fluxes from sealing 39
4.2 Seasonal saving in flux from sealing 44
4.3 Correlation Scenarios 49

5 Discussion **51**
5.1 Additional side-effects of paving 52
5.2 Future research 54

6 Conclusion **57**

Appendix A Appendices **58**

References **64**

List of Tables

2.1 Global GHG fluxes, from Hashimoto (2012) 4

2.2 Responses of global GHG fluxes to a 1 °C increase in temperature. From Yan et al. (2022). 6

3.1 Values of $\boldsymbol{\mu}$ for the five broad land cover classes used. These are just of the (lower + upper)/2, where lower and upper are the typical bounds of soil GHG fluxes from Oertel et al. (2016). Units for CO_2 are $\mu mol\, CO_2\, m^{-2}\, s^{-1}$. If X is either CH_4 or N_2O, units are $\mu mol\, X\, m^{-2}\, h^{-1}$. 16

3.2 Standard deviations in CO_2, CH_4 and N_2O. 17

3.3 A list of CLC classes (and their codes) that we reclassified as NA. . . 21

3.4 Spatial products used, their resolution and the year they represent. We derived the unsealed land cover map using the CLC in Section 3.3. 24

3.5 Summary of the Belfast imperviousness data following different kinds of resampling from 10 to 1000 squared m resolution. 29

3.6 Q_{10} used for GHGs values. Taken from Oertel et al. (2016), Dalal and Allen (2008), and Abdalla et al. (2009). 30

3.7 The unsealed area of each land cover class and the proportion of this they account for. The Avg. IMD is the average inverted imperviousness for each class. Finally, the sealed areas are the element-wise sum of the UK imperviousness data and the UK proportional land cover data (ϕ = Sealed area in Equation 3.6). We used this column to adjust the fluxes for sealing in the linear case. The sealed area is the effective area of the UK that the land covers account for, when each square kilometre is scaled by IMD. 34

4.1 Global warming potentials (GWP) of the three GHGs we used here in terms of the carbon dioxide equivalent (CO_2e). The data here are at the 20-year time horizon from report 6 of the International Panel for climate change assessment (AR6-20, IPCC (2021)). The data we used were the output of the LICA formatter from Young et al. (2021) 40

List of Figures

2.1 Key drivers of soil GHG emissions (from Oertel et al. (2016)). 5

2.2 Soil Respiration Database data (SRDB) from Jian et al. (2021). Only unique sites with annual soil respiration are shown. We have grouped arctic with boreal, and tropical with subtropical to allow the ecosystems to be more easily differentiated. 8

2.3 Number of observations of each ecosystem type in the SRDB. A log base-10 scale is used for the x-axis. 9

2.4 Continuous soil respiration data sources from the COSORE database, Bond-Lamberty, Christianson, et al. (2020). 10

2.5 OSM (Mark Padgham et al., 2017) road data alongside 2018 IMD in Ceredigion, Wales. 12

3.1 Flowchart showing an overview of methodology. 14

3.2 On the left we have a UK CLC map after reclassifying all sealed classes to NA. Then we show the results of moving window calculation using a square and circular weighted kernel. Some land cover class names have been shortened. 22

3.3 2018 IMD in Belfast following resampling. Different resampling methods are shown by aggregated (A), bilinear (B), cubic (C), cubic spline (D), and nearest-neighbour (E). (A)-(E) are all at 1000 m resolution. The 10 m resolution raw data is shown in (F). 25

3.4 OSM (Mark Padgham et al., 2017) road data alongside 2018 IMD in Lancaster and Morecambe. The thick dashed line is the west coast of the England, north of Morecambe. We zoom in on a portion of motorway to illustrate that this is classified correctly as impervious. . 31

4.1 The annual saving in CO_2 (tonnes) from sealed soil in the UK. The y-axis is ${CO_2}^{\text{unsealed}} - {CO_2}^{\text{sealed}}$ and shown on a base-10 logarithmic scale. The error bars are the 95% confidence intervals based on 1000 simulations. The bars are filled in based on the type of impervious surface causing the difference in emissions. 36

4.2 The annual saving in CH_4 (kilotonnes) from sealed soil in the UK. For CH_4, the magnitude of these mostly negative savings are really emissions caused by sealing. The y-axis is ${CH_4}^{\text{unsealed}} - {CH_4}^{\text{sealed}}$. The error bars are the 95% confidence intervals based on 1000 simulations. The bars are filled in based on the type of impervious surface causing the difference in emissions. 37

4.3 The annual saving in N_2O (kilotonnes) from sealed soil in the UK. The y-axis is ${N_2O}^{\text{unsealed}} - {N_2O}^{\text{sealed}}$. The error bars are the 95% confidence intervals based on 1000 simulations. The bars are filled in based on the type of impervious surface causing the difference in emissions. 38

4.4 The annual saving in CO_2e (tonnes) from sealed soil in the UK. The y-axis is ${CO_2e}^{\text{unsealed}} - {CO_2e}^{\text{sealed}}$ and shown on a base-10 logarithmic scale. The mean estimate of each land cover is as white text on the bars. CO_2e is calculated using the AR6-20 GWP for CO_2, CH_4 and N_2O as shown in Equation 4.2. The error bars represent a 95% confidence interval. Data used here is based on the same simulated fluxes with 0 correlation used in Figure 4.1. 41

4.5 The annual saving due sealed soil in the UK for CO_2, CH_4 and N_2O and all three GHGs . The y-axis is in terms of the CO_2e (megatonnes) of the GHGs. These annual savings have been summed across the 5 land cover types. The text on the bars represents the annual flux (megatonnes) and the orange error bars are the 95% confidence interval based on 1000 UK GHG flux simulations. 42

4.6 The annual savings due to sealed soil for CO_2, CH_4 and N_2O, split according to land cover. The y-axis is in terms of the CO_2e (megatonnes) of the GHGs for ease of comparison. The orange error bars are the 95% confidence interval based on 1000 UK GHG flux simulations. 43

4.7 The monthly saving in CO_2 (megatonnes) from sealed soil in the UK. The light blue dashed line shows the mean total monthly GHG saving in the UK. The lower and upper grey dashed lines show the lower and upper bounds of a 95% confidence for this average. 45

4.8 The monthly saving in CH_4 (tonnes) from sealed soil in the UK. For CH_4, the magnitude of the negative savings here represent the emissions from sealing. The light blue dashed line shows the mean total monthly GHG saving in the UK. The lower and upper grey dashed lines show the lower and upper bounds of a 95% confidence for this average. 46

4.9 The monthly saving in N_2O (tonnes) from sealed soil in the UK. The light blue dashed line shows the mean of the total monthly GHG saving due to sealing; the lower and upper grey dashed lines show the lower and upper bounds of a 95% confidence interval about this. . . . 47

4.10 The monthly saving in CO_2e (megatonnes) from sealed soil in the UK. We have summed the CO_2e of seasonal savings in CO_2, CH_4 and N_2O to see the overall effect of these GHGs. The light blue dashed line shows the mean total monthly GHG saving in the UK. The lower and upper grey dashed lines show the lower and upper bounds of a 95% confidence for this average. 48

4.11 The annual saving in GHG soil fluxes from sealing given different correlation scenarios. The contours represent the 2D kernel density estimation of the saving in CO_2 and N_2O over the 1000 simulations. In these scenarios $\rho_{1,2} = \text{corr}(CO_2, CH_4)$, $\rho_{1,3} = \text{corr}(CO_2, N_2O)$ and $\rho_{2,3} = \text{corr}(CH_4, N_2O)$. 49

4.12 Contours showing the 95% confidence interval of the 2D kernel density estimates of CO_2 and N_2O emissions under different correlation scenarios. In these scenarios $\rho_{1,2} = \text{corr}(CO_2, CH_4)$, $\rho_{1,3} = \text{corr}(CO_2, N_2O)$ and $\rho_{2,3} = \text{corr}(CH_4, N_2O)$. 50

Chapter 1

Introduction

The impact of paving over land on greenhouse gas (GHG) emissions is largely absent from existing research. This dissertation aims to produce estimates for the GHG emission of carbon dioxide (CO_2), methane (CH_4), and nitrous oxide (N_2O) from soil in different types of land in the UK. Then, by estimating the difference in these GHG emissions from paving we can begin to appreciate how we have inadvertently impacted climate change, and possibly better inform future efforts to mitigate this impact.

1.1 Terminology and definitions

'Fluxes' describes both the uptake and emission of GHGs from soil. Later, when we report the 'saving' in GHG emissions from sealing we mean GHGs that would have been emitted if the concrete were not present. For ease of comparison between GHGs, we sometimes report these savings as negative; in this case sealing the soil is effectively emitting the GHG in question. We use 'sealed' to describe our world where some soils are paved over by artificial materials like concrete, asphalt or gravel. Therefore, soil covered by ice or snow is not considered sealed. 'Unsealed' refers to land that has not been covered by some impermeable material, like concrete or asphalt. In our world, soils can be fully sealed, unsealed or partially sealed. We refer

to this as the degree of 'imperviousness' of the soil, taking a value $x \in (0, 100)\%$. So soil with an imperviousness value of x interacts with the soil atmosphere at $\beta = (100 - x)\%$ of the unsealed equivalent. We also will consider the GHG fluxes across different 'land covers'. A land cover map is a representation of Earth's surface using a set of discrete classes. These classes might include, for example, *grassland*, *urban* or *shrubland*. There are many different ways of representing the Earth's surface in this way and different land cover maps can better represent certain features.

1.2 Outline

We begin by providing a background to this dissertation in Chapter 2. Next in Chapter 3, we show and justify the steps taken to produce our results. Then, in Chapter 4 we give, analyse and interpret our results. We discuss the challenges, limitations, and results from our method in a wider context in Chapter 5. Finally, we conclude in Chapter 6 with a summary of the implications of our research.

Chapter 2

Background

2.1 Related work

Existing literature on the area mainly considers relatively few sites and aims to understand the relationship between soil GHG fluxes and the conditions in soils in a particular biome.

Hashimoto (2012) estimate the global total, spatial distribution and seasonality of the soil fluxes of CO_2, N_2O and CH_4 using a simple data-oriented model. We see their primary results in 2.1. Generally CO_2 contributed the most to the global GHG flux, with tropical regions the home of the highest fluxes for all three GHGs. They also found the seasonal pattern in the fluxes of CO_2 and N_2O to be significantly more pronounced. We will check our seasonal results for a pattern like this later. Globally, they considered the global flux at a $0.5^\circ \times 0.5^\circ$ resolution between 1980 and 2009, presenting the average of these maps. Hashimoto (2012) also stresses the importance of building global soil GHG databases - which we examine in 2.2.1.

Qiu et al. (2024) examine the global effect of expanding imperviousness surface area (ISA) on carbon emissions due to the depletion of terrestrial carbon sinks. However, they do not consider the GHG emissions forgone due to expanding ISA. Following expansion in ISA, they report significant carbon losses from top soil and biomass of 46-75 Mt C between 1993 and 2018. Such a result is crucial to the overall

GHG	Units	Global soil flux	95% Monte Carlo CI
CO2 emission	Gt C yr^{-1}	78.0	(64, 95)
CH4 uptake	Mt C yr^{-1}	18.0	(11,23)
N2O emission	Mt N yr^{-1}	4.4	(1.4,11.1)

Table 2.1: Global GHG fluxes, from Hashimoto (2012)

effect of sealing land on our environment. We will therefore couple our findings with this later to avoid ignoring the wider picture of sealing soil. To calculate the expansion in urban surfaces, they averaged four different global ISA products. A range of carbon density maps were combined with this ISA expansion map to calculate carbon losses from urban expansion. This too was combined with a 300 m European Space Agency climate change initiative land cover map Bontemps et al. (2013). They went further than this though, adapting the Kaya-identity method in Kaya (1989) to attribute the change in ISA-driven carbon sink losses to the different driving factors of ISA.

In this dissertation we considered data on both heterotrophic respiration, R_h, the microbial respiration of CO_2 from bacteria living in the soil, and autotrophic respiration, R_a, which concerns CO_2 respiration of plant parts like roots Rankin et al. (2023). Both of these types of respiration emit a greenhouse gas - CO_2 - from soil. Therefore, the total respiration, $R_s = R_a + R_h$ is an appropriate way of measuring the overall CO_2 flux from soil. Some researchers, like Kumari et al. (2023), make it clear that the impact of soil microbes on the climate is not solely coming from their respiration. They highlight how soil microbes improve soil health, support plant growth, and decontaminate soils. So, sealing will have additional impacts on soil microbes, impacts that will also affect the environment.

Soil GHG flux measurements do not include the fluxes of plants on top of the soil, but only the GHG fluxes of the soil itself. For instance, the CO_2 intake from trees in a Forest is not of interest. Only the GHG fluxes of soil lying below the canopy are of interest. Existing research into soil GHG emissions were compiled by

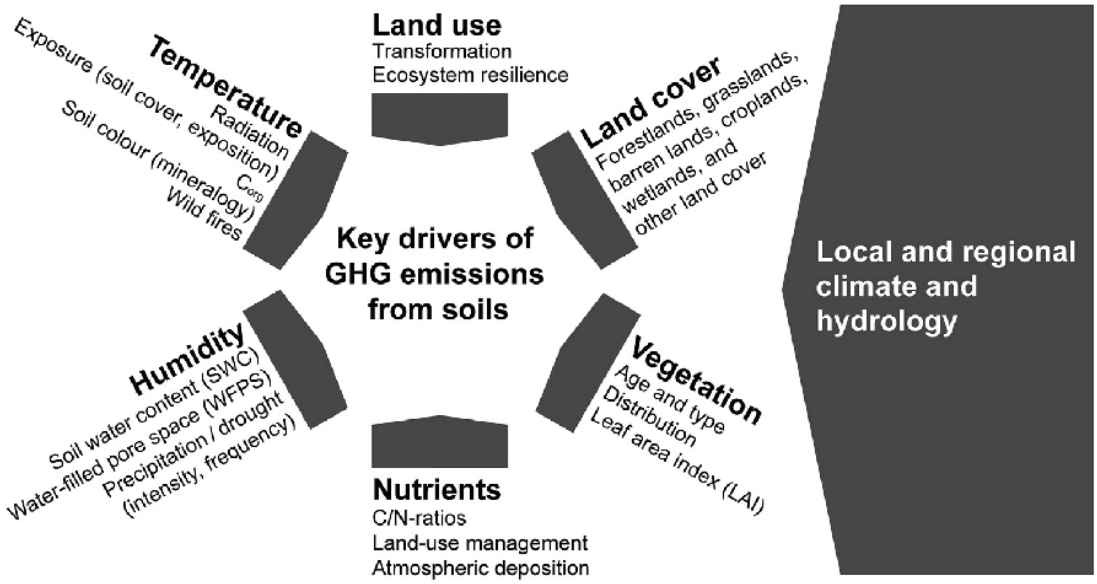

Figure 2.1: Key drivers of soil GHG emissions (from Oertel et al. (2016)).

Oertel et al. (2016). This was the primary source of soil greenhouse gas emissions we used in this dissertation. Oertel et al. (2016) pay particular attention to chamber systems in their analysis. Here, chambers are placed directly on soil, allowing the effects of the soil to be measured - without other nearby sources like plants. Maier et al. (2022) present a guideline for measurements of GHG fluxes using a type of chamber system called the non-steady-state chambers. They stress that, to obtain reliable measurements from these chambers, many details need to be accounted for.

Oertel et al. (2016) review the key drivers of GHG emissions from soil, summarising them in Figure 2.1. Clearly, a large number of climatic conditions affect GHG fluxes. We will not consider all of these in this dissertation - focusing only on land cover and temperature. However, Oertel et al. (2016) highlight that many other factors will influence soil GHG emissions, and the relationships between these drivers are incredibly important to consider. Next, we will consider the influence of temperature on soil GHG emissions.

The effect of temperature on soil GHG fluxes has often been estimated to try and discern how global warming will effect fluxes over time. Yan et al. (2022) executed a meta-analysis using results from 2423 observations. Given that in lowland regions,

GHG	Change in flux (kg ha^{-1} year^{-1})	Percentage change
CO_2	2289.0	9.9
CH_4 lowland	31.0	10.1
CH_4 upland	0.5	7.5
N_2O	-0.1	17.9

Table 2.2: Responses of global GHG fluxes to a 1 °C increase in temperature. From Yan et al. (2022).

CH_4 is emitted whereas in upland regions CH_4 is stored, they split CH_4 fluxes into these two groups. In Table 2.2 we see their results, with the magnitude of each GHG flux increasing following a 1 °C increase in temperature.

In reality, the temperature sensitivity of GHG emissions is not the same for different GHGs, land cover classes or soil moisture levels. Q_{10} is a useful measure of the temperature sensitivity of biological systems - in our case the microbes in soil - to a 10°C change in temperature (Meyer, Welp, and Amelung, 2018). Q_{10} is defined by Equation 2.1 as follows, where R_1 is GHG flux rate prior to adjustment is; R_2 is the temperature adjusted GHG flux; T_1 is the temperature for R_1, and T_2 is the temperature for R_2:

$$Q_{10} = \left(\frac{R_2}{R_1}\right)^{10°C/(T_2-T_1)}. \tag{2.1}$$

The differences in temperature sensitivity across GHGs are addressed by Oertel et al. (2016), Dalal and Allen (2008), and Abdalla et al. (2009) who together provide ranges and point estimates for the Q_{10} temperature coefficients (Equation 2.1, Mundim et al. (2020)) of CO_2, CH_4 and N_2O fluxes. There can be complicated patterns in temperature sensitivity that might not be captured by Q_{10} values. For example, Oertel et al. (2016) explains that for N_2O, emissions rise with temperatures up to 37°C before beginning to fall. When we consider temperature in the UK, this effect will not come in to play, as the UK's average monthly temperatures are not this high. Yan et al. (2022) stress the importance of considering the biome alongside

the responses of GHG respiration and temperature sensitivity to warming. They also provide the data they used to calculate the temperature sensitivity for each GHG and biome, allowing us to derive Q_{10} values for comparison with those in Oertel et al. (2016). Meyer, Welp, and Amelung (2018) also explore the relationship between moisture, land cover and Q_{10} by considering the temperature sensitivity of soil CO_2 respiration.

2.2 Related Data

Before we describe the sources of data related to our subject matter, we first provide some background information on the different types of data we used. First, we have raster data, which represents some region of space using a grid of rectangles (Robert J. Hijmans, 2024). Bernardo et al. (2013) describes this grid as a matrix with each element (or cell) representing a spatial attribute. Rasters can have multiple, identically aligned, layers. Each layer contains cells relating to one spatial attribute. These layers are really just individual rasters aligned in the same way. Raster data was the main source of the spatial data we used in this dissertation. Portable network graphic images (PNGs) are a form of raster data, so when describing methods relating to our spatial rasters sometimes the word 'image' is used.

2.2.1 Soil Greenhouse Gas Flux databases

2.2.1.1 Soil Respiration Database

Jian et al. (2021)'s Soil Respiration Database, (SRDB version 5) is the largest and furthest reaching source of soil respiration data. Annual soil respiration is the key column of the SRDB for research in this area, and also with the most observations in the database. There is also data regarding temperature, seasonal soil respiration, drainage and precipitation. Currently, it only contains data relating to CO_2. Without CH_4 and N_2O included, it could not be used as the only source

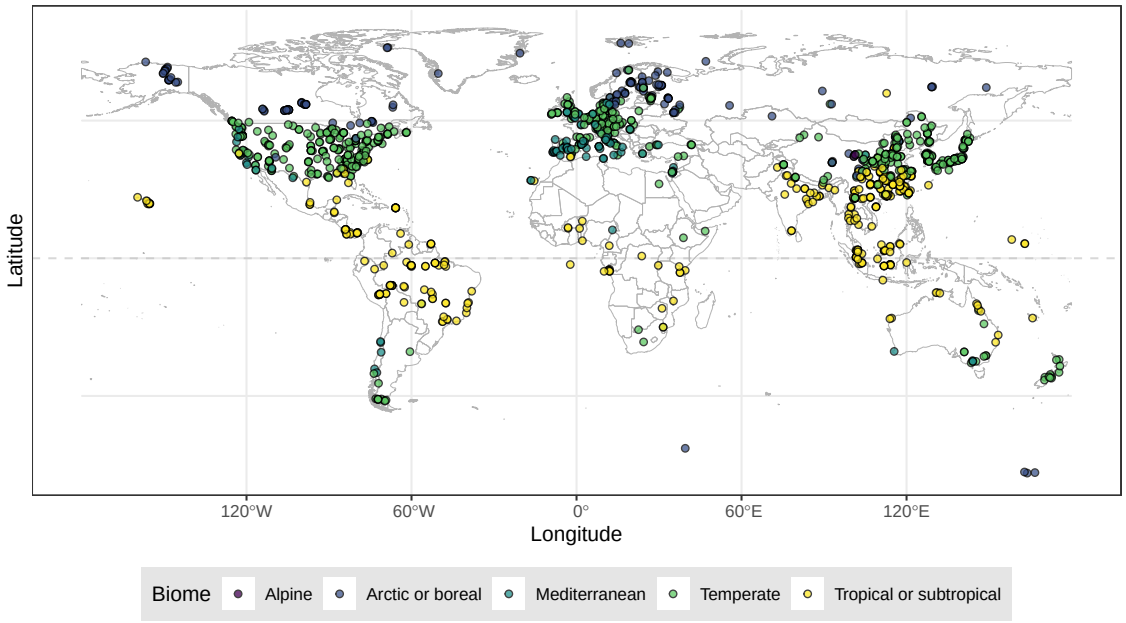

Figure 2.2: Soil Respiration Database data (SRDB) from Jian et al. (2021). Only unique sites with annual soil respiration are shown. We have grouped arctic with boreal, and tropical with subtropical to allow the ecosystems to be more easily differentiated.

of GHG flux data in this Dissertation. In Figure 2.2 we show the all of the sites with annual CO_2 respiration recorded in the SRDB. The vast majority of sites are in the northern hemisphere, specifically Europe, North America and East Asia. Also, in Figure 2.3, we show the quantity of observations in each ecosystem across the SRDB.

2.2.1.2 COSORE

Another good source of soil respiration data is the community-sourced COSORE database. COSORE, accessed via an R package of the same name, contains continuous soil respiration data and has observations of both CO_2 and CH_4 at

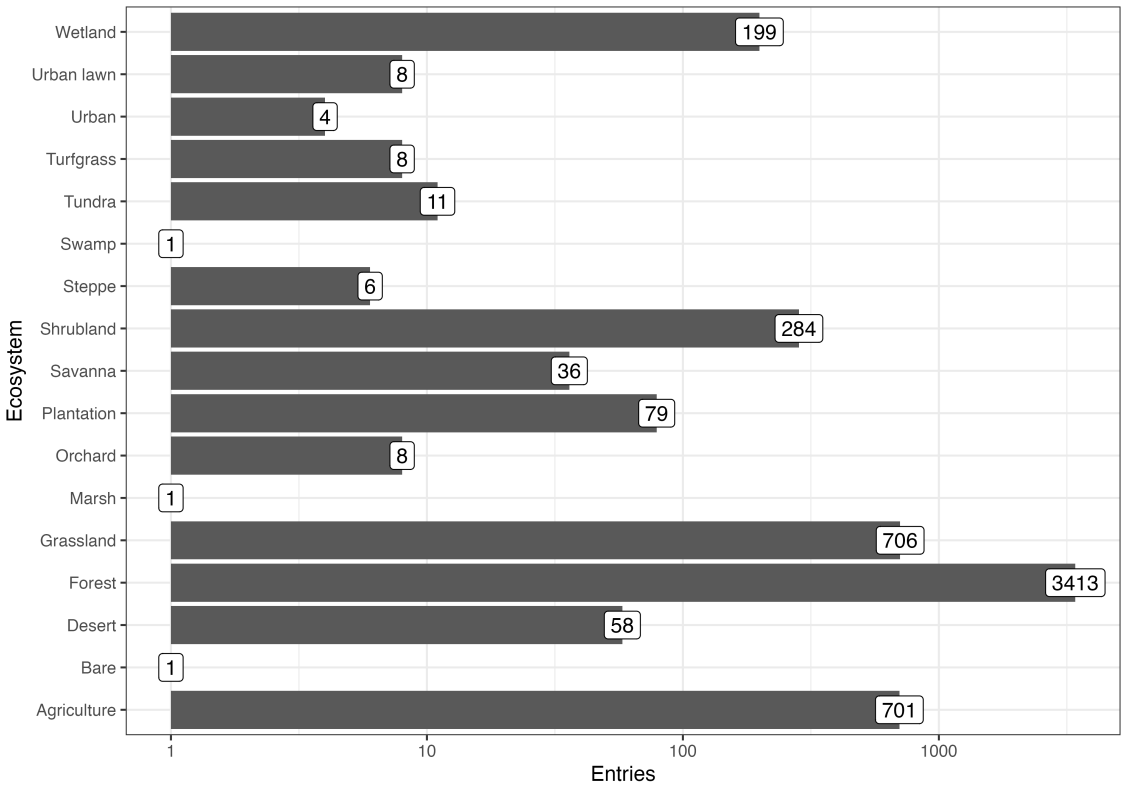

Figure 2.3: Number of observations of each ecosystem type in the SRDB. A log base-10 scale is used for the x-axis.

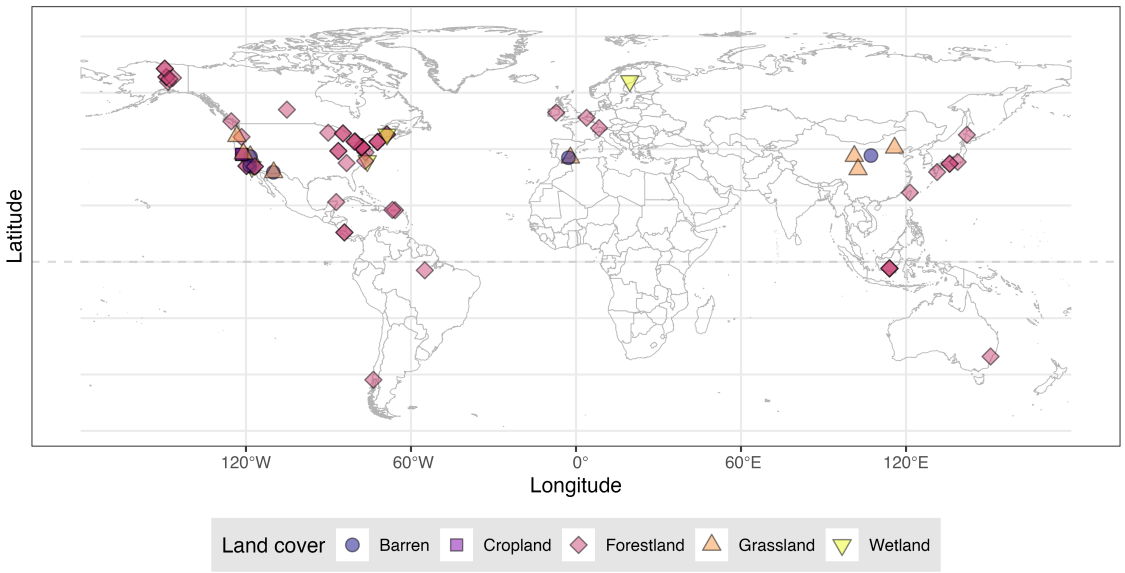

Figure 2.4: Continuous soil respiration data sources from the COSORE database, Bond-Lamberty, Christianson, et al. (2020).

certain sites. It does not currently contain any data regarding N_2O, but if long-term soil respiration observations are added from new, and existing, sites this could be extremely helpful for global estimations of soil GHG fluxes. COSORE certainly has a significant amount of potential, enabling different kinds of research using soil GHG fluxes than the SRDB.

2.2.2 Land Cover

Corine (coordination of information on the environment) land cover from the European Environment Agency (2020a) is not just a land cover map but goes into some detail on land use as well. We will henceforth call the 100m Corine land cover map the CLC. Büttner et al. (2021) give us a full breakdown of the nomenclature of the different classes used, referring to this as 'AD01'. On two occasions, the CLC has

been verified to have a thematic accuracy of at least 85%. The distinction between thematic accuracy, or how correct a map's classification of a region is to reality, from positional accuracy concerning the fit of features in maps or images to reality is made by Congalton (2005). They add that these are concepts whilst different are related to each other and are both important. The CLC raster layer was derived from Sentinel-2 satellite data. It conforms fully with the European Environment Agency (EEA-39) reference grid (Agency, n.d.), which still includes the UK as well as non-EU member countries and cooperating countries.

Another option we considered was the UK land cover map (UKLCM, Morton et al. (2022). This product has far fewer classes and a pre-aggregated version from 2018. It also only as two urban classes - urban and suburban. Much like the CLC, the UKLCM was derived using Sentinel 2 satellite data. 9 different spectral bands (each a different region of the electromagnetic spectrum) from all four seasons were used to generate the UKLCM. Cloud cover was persistent for some regions of the UK in certain seasons, but never in one region for the full year. This allowed an algorithm to be used to fill in the gaps in the season. To avoid issues with different land covers with similar spectral properties context rasters were used to avoid confusion.

2.2.3 Imperviousness

Imperviousness data from European Environment Agency (2020b) provides an indication of how sealed soil in different regions is. Henceforth, we will calls this dataset IMD. Certain features like permanent greenhouses can be difficult to detect using current Sentinel-2 satellite images so are sometimes underestimated. Currently it has a 10 m resolution, which captures large urban areas well but struggles to account for roads (European Environment Agency, 2020b). This potentially poses a significant problem for our estimate. In 2.5 we see that for Ceredigion in Wales, an area with no major roads, IMD does not capture any of the roads in the region. Since the majority of roads in the UK these minor roads (UK Department for Transport, 2022), we would miss out on a large area of sealed land in the UK if we only used

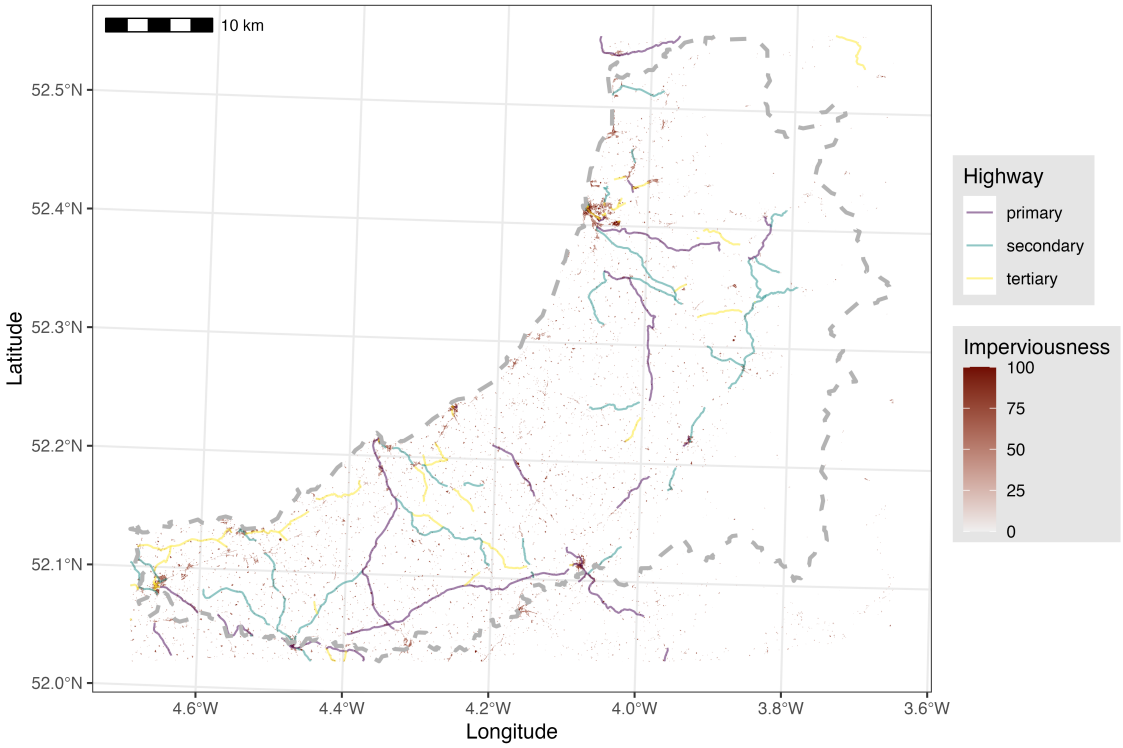

Figure 2.5: OSM (Mark Padgham et al., 2017) road data alongside 2018 IMD in Ceredigion, Wales.

the IMD data.

We defined user and producer accuracy (Story and Congalton, 1986), in terms of true positives, TP, false positives, FP and false negatives FN in 2.2. For IMD, they give $a_{\text{user}} = 96.14\%$ and $a_{\text{producer}} = 93.68\%$. Therefore, there is a 96.14% chance that sealed land is also sealed in their reference map, which was the very high resolution image dataset used to produce the final map.

$$\text{Specificity} = a_{\text{user}} = \frac{TP}{TP + FP} \tag{2.2a}$$

$$\text{Sensitivity} = a_{\text{producer}} = \frac{TP}{TP + FN} \tag{2.2b}$$

Chapter 3

Methodology

3.1 Method overview

An overview of the steps used in the analysis is shown in Figure 3.1. In order to calculate the total UK flux for each GHG and land cover we: (1), created a new unsealed UK land cover map; (2), assigned GHG flux rates to classes in the unsealed map (3), combined this map with the imperviousness over each squared kilometre (4), created a new sealed map where we scale each flux in the sealed map based on the inverted imperviousness; (5), summed the fluxes over each land cover for the sealed map and, finally (6), repeat step (5) for the unsealed map.

Soil GHG flux lower and upper bounds were acquired from Oertel et al. (2016). There were many different types of land cover accounted for in this review, but we only considered the temperate bounds for five broad land cover classes: grassland, cropland, bare soil, wetlands and forestland. Two types of bounds were given in the paper - extreme outcomes and then other, typical ranges. We used the typical bounds for each land cover and GHG. GADM administrative areas helped to examine the major UK cities in greater detail.

All analyses were performed using the R Statistical Software (v4.4.1; R Core Team, 2024). Spatial processing and analysis was carried out via the R packages terra (v1.7.78; Robert J. Hijmans, 2024) and sf (v1.0-16; Pebesma, 2018). Plots

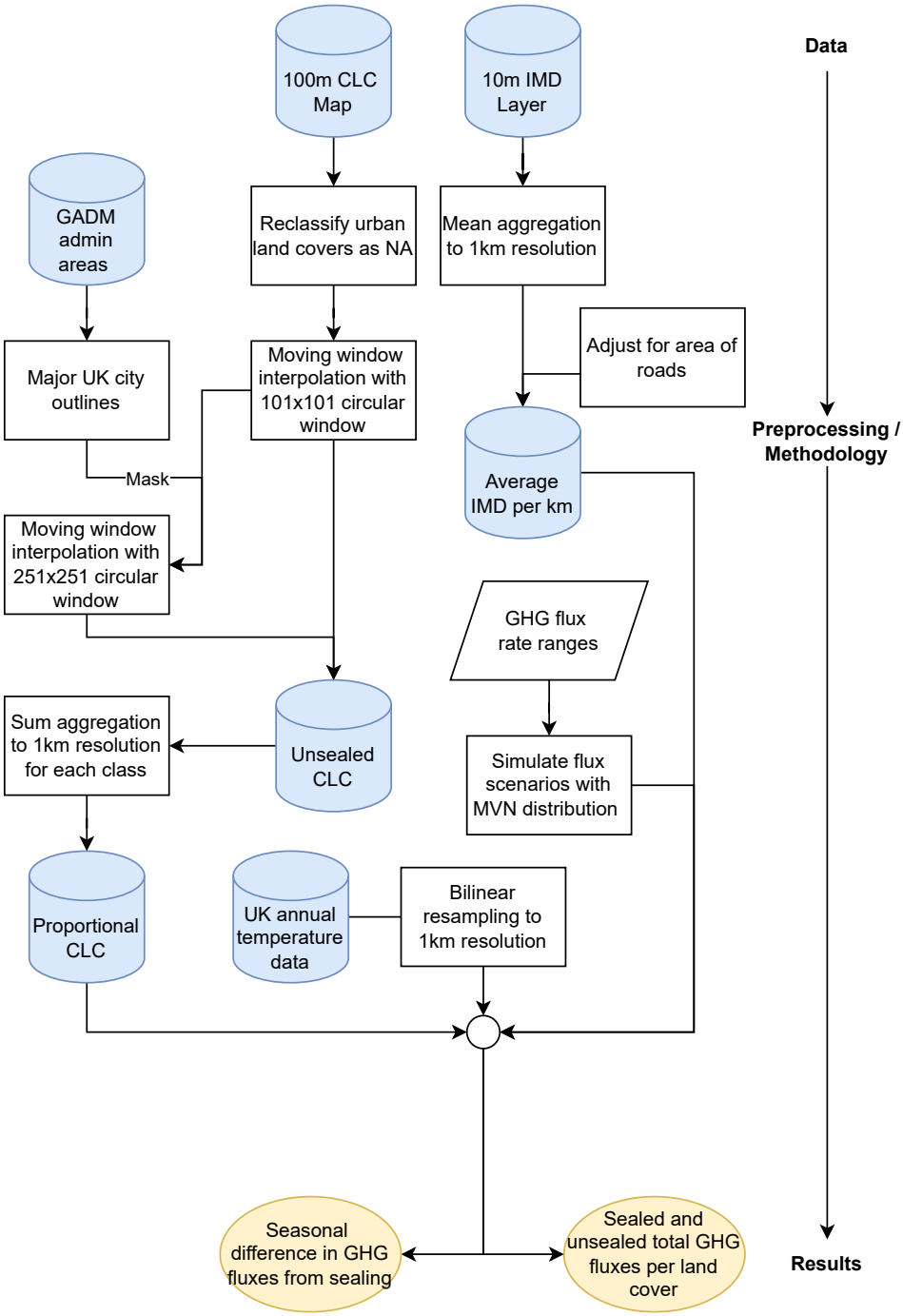

Figure 3.1: Flowchart showing an overview of methodology.

were produced using the ggplot2 package (v3.5.1; Wickham, 2016). Seasonal results were summarised using the duckdb R api (v1.0.0-2; Mühleisen and Raasveldt, 2024). The R package MASS (v7.3.61; Venables and Ripley, 2002) was used for multivariate normal simulations.

3.2 Simulating GHG fluxes

We decided to use simulated data based on the typical upper or lower bounds of fluxes provided in Oertel et al. (2016). The aim of the simulation was to provide a more detailed view of the distribution of soil GHG fluxes. Using the multivariate normal (MVN) distribution also allows us to consider correlation between the soil fluxes of CO_2, CH_4 and N_2O, which otherwise we are assuming to be completely uncorrelated - despite evidence in Kumari et al. (2023) against this for CO_2 and CH_4 and in Dijkstra et al. (2012) for all three GHGs. Whilst we cannot consider all of the different correlation scenarios possible, by comparing some scenarios we could still gain a better idea of possible distribution of results than were we to only assume the uncorrelated scenario.

When generating the simulations we derived the parameters of our MVN distribution using the typical bounds (Oertel et al., 2016), assuming they were within 3 standard deviations of the midpoint between the upper limit and lower limit. These midpoints were used as the values of $\boldsymbol{\mu}$ shown in Table 3.1. Together, this setup ensures that, were the GHG gas fluxes normally distributed, around 99.73% of the data would fall within these typical bounds.

For each of the five land cover classes shown in Tables 3.1 and 3.2, we use the MVN distribution to simulate

$$\boldsymbol{X} = \begin{bmatrix} x_{1,1} & x_{1,2} & \dots & x_{1,n} \\ x_{2,1} & x_{2,2} & \dots & x_{2,n} \\ x_{3,1} & x_{3,2} & \dots & x_{3,n} \end{bmatrix}. \tag{3.1}$$

GHG	Symbol	Cropland	Barren	Grassland	Wetland	Forestland
$\mathbf{CO_2}$	μ_1	4.325	1.95	4.250	4.65	0.97
$\mathbf{CH_4}$	μ_2	-0.285	-1.76	-1.730	13.10	-6.29
$\mathbf{N_2O}$	μ_3	2.155	4.96	4.865	1.40	0.65

Table 3.1: Values of $\boldsymbol{\mu}$ for the five broad land cover classes used. These are just of the (lower + upper)/2, where lower and upper are the typical bounds of soil GHG fluxes from Oertel et al. (2016). Units for CO_2 are $\mu\text{mol}\,CO_2\,m^{-2}\,s^{-1}$. If X is either CH_4 or N_2O, units are $\mu\text{mol}\,X\,m^{-2}\,h^{-1}$.

In 3.1, $\forall j = 1 \dots n$ the values of $x_{1,j}$, $x_{2,j}$ and $x_{3,j}$, represent the jth simulation of the CO_2, CH_4 and N_2O fluxes respectively. So $\boldsymbol{X}$ only accounts for one land cover class but will be in the same form whichever class we simulate. Since the method is identical for all the land cover classes we will only show the method for one here.

We use the following notation when referring to the correlation between the annual fluxes of carbon dioxide, methane and nitrous oxide:

$$\rho_{1,2} := \text{corr}(CO_2, CH_4), \tag{3.2a}$$

$$\rho_{1,3} := \text{corr}(CO_2, N_2O), \tag{3.2b}$$

$$\rho_{2,3} := \text{corr}(CH_4, N_2O). \tag{3.2c}$$

These correlation measures were used to calculate different candidates for the variance covariance matrix, $\boldsymbol{\Sigma}$ shown in Equation 3.3. We used five different versions of $\boldsymbol{\Sigma}$ to account for the five different standard deviation vectors composed of $\sigma_1, \sigma_2, \sigma_3$ for cropland, barren, grassland, wetland and forests. The values used for these 5 different versions of $\boldsymbol{\Sigma}$ are shown in Table 3.2 and Equations 3.2.

Now with both the standard deviations of the GHGs and correlations specified we can define the variance-covariance matrix of the GHG fluxes, $\boldsymbol{\Sigma}$, as follows in line 3.3:

GHG	Symbol	Cropland	Barren	Grassland	Wetland	Forestland
CO_2	σ_1	1.292	0.583	1.250	1.150	0.073
CH_4	σ_2	0.042	0.413	0.590	4.967	2.070
N_2O	σ_3	0.715	1.680	1.612	1.000	0.190

Table 3.2: Standard deviations in CO_2, CH_4 and N_2O.

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho_{1,2} & \sigma_1\sigma_3\rho_{1,3} \\ \sigma_1\sigma_2\rho_{1,2} & \sigma_2^2 & \sigma_2\sigma_3\rho_{2,3} \\ \sigma_1\sigma_3\rho_{1,3} & \sigma_2\sigma_3\rho_{2,3} & \sigma_3^2 \end{bmatrix} \tag{3.3}$$

.

In line 3.4, $\boldsymbol{\mu} = [\mu_1, \mu_2, \mu_3]$ are the midpoints in the same fashion. These are shown in Table 3.1.

$$\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \tag{3.4}$$

.

In each simulation, the same GHG flux was assigned to each particular land cover. For example, a simulated value of 3.82 µmol $CO_2\,m^{-2}\,s^{-1}$ for grassland would be assigned to all grassland in the first simulation. The same applies to other land covers and their associated simulated GHG fluxes. This assumes the highest possible level of spatial dependency between GHG fluxes, and will cause the variance of our final estimates to be higher than were we to simulate a different flux for each of the 244321 different km^2 in our maps. Given the context of climate change, the cost of underestimating the confidence interval of our total UK fluxes is higher than that of overestimating them. Therefore, assuming a high level of this spatial dependency - here we assume the maximum - reduces the risk of overestimating the confidence intervals of our results. This is a real risk when using this kind of site data in simulations (Hashimoto, 2012).

3.2.1 Analytical results

Without including temperature, the relationship between the simulated GHG fluxes, the imperviousness data and the total area is linear in our method. For the unsealed fluxes, we scale the fluxes (see 3.21 later) multiplying each of the n simulated fluxes by the total area (kilometres squared) taken up by the land cover class. We derive the sealed map by computing the sum the element-wise product of, the imperviousness values for each land cover and areas of each land cover 3.5. In our case, we used 1000 simulations so this left us with 1000 annual estimates for each GHG and land cover class.

For a single land cover class, we denote the imperviousness data by $\boldsymbol{m} = [m_1, \ldots, m_p]^\top$ and the proportion of each class by $\boldsymbol{a} = [a_1, \ldots, a_p]^\top$. We have imperviousness values and class proportions for each km^2 in the UK so, in our case, $p = 244321$. We denote the mean imperviousness for our given land cover class by $\bar{m}$. The total area of the class is:

$$\sum_{i=1}^{p} a_i = p\bar{a}. \tag{3.5}$$

We denote the constant the total sealed area of each class by

$$\phi = \sum_{i=1}^{p} a_i m_i = \boldsymbol{a}^\top \boldsymbol{m}. \tag{3.6}$$

From here we can evaluate the expectation and variance of our transformed simulated fluxes. For the map of sealed results we therefore have:

$$\mathbb{E}\left[\phi \boldsymbol{X}\right] = \phi\, \mathbb{E}[\mathbf{X}] \tag{3.7}$$

$$= \phi \boldsymbol{\mu}, \tag{3.8}$$

and variance:

$$\text{Var}\left[\phi \mathbf{X}\right] = \phi^2 \text{Var}[\mathbf{X}] \tag{3.9}$$

$$= \phi^2 \boldsymbol{\Sigma}. \tag{3.10}$$

Similarly, for the unsealed results we have:

$$\mathbb{E}\left[p\bar{a}\boldsymbol{X}\right] = p\bar{a}\,\mathbb{E}[\mathbf{X}] \tag{3.11}$$

$$= p\bar{a}\boldsymbol{\mu}, \tag{3.12}$$

and variance:

$$\mathrm{Var}\left[p\bar{a}\mathbf{X}\right] = p^2\bar{a}^2\mathrm{Var}[\mathbf{X}] \tag{3.13}$$

$$= p^2\bar{a}^2\boldsymbol{\Sigma}. \tag{3.14}$$

The only 'randomly sampled' terms here are the normally distributed fluxes for each land cover class. Therefore, for the sealed estimates, our transformation, $\boldsymbol{Y} = \phi\boldsymbol{X}$, is just a linear combination of ϕ and $\boldsymbol{X}$, we know that $\boldsymbol{Y}$ has a multivariate normal distribution as follows:

$$\boldsymbol{Y} \sim \mathcal{N}\left(\phi\boldsymbol{\mu}, \phi^2\boldsymbol{\Sigma}\right). \tag{3.15}$$

Similarly, we denote the unsealed map by the matrix $\boldsymbol{Z} = p\bar{a}\boldsymbol{X}$. This gives:

$$\boldsymbol{Z} \sim \mathcal{N}\left(p\bar{a}\boldsymbol{\mu}, p^2\bar{a}^2\boldsymbol{\Sigma}\right). \tag{3.16}$$

Both the means of $\boldsymbol{Y}$, and $\boldsymbol{Z}$, are $3 \times n$ matrices with the first, second and third rows containing estimates of the total UK flux from CO_2, CH_4 and N_2O respectively. Each of the $n = 1000$ is a different run of the simulation. This is the simulation of one land cover. By stacking the simulated total UK fluxes for each different land cover, it follows that 1000 simulations of all 3 GHGs and 5 land covers is a 15×1000 matrix with each row corresponding to a particular land cover and GHG. Each column is, again, a run of the simulation. Scaling the GHG fluxes by $\boldsymbol{m}$, exclusively separates the sealed, and unsealed, versions of the UK.

3.3 Creating the unsealed map

We considered using either the CLC and UKLCM. The CLC covers all of Europe, so has more potential for larger scale discoveries than the UK land cover map (UKLCM). The CLC's 47 classes - as opposed to the UKLCM's 21 classes - add to this. The greater thematic detail in the CLC's cover could allow researchers to assign cover specific GHG fluxes, some of which are significantly different. Using a global map with a higher level of thematic detail is a next step from with this research. The CLC can be more easily expanded to include this detail. Thus, it can make for better comparisons with future, global research that considers more cover specific GHG fluxes.

To estimate the effect of soil sealing on UK GHG emissions, we need to consider some world without any paved soil. For example, a housing estate in the middle of a forest is sealed, but we want to consider this a forest in our unsealed map to allow us to isolate the effect of sealing. We began this by reclassifying urban land cover types (see Table 3.3) in the CLC land cover map with a value of NA. This gave the left hand plot in Figure 3.2 where we can see the vast majority of London has been erased from the land cover map due to being largely sealed. From here, the problem becomes populating these NA values with unsealed land cover types. There are different approaches to this task.

We strongly considered using a potential natural vegetation map (PNV) to assign urban regions a natural land cover class based on climactic conditions in the region. We looked into using the potential vegetation map produced and given to us by Hinze, Albrecht, and Michiels (2023) to achieve this. Their PNVs were too different from reality, so were not used in this research. Unsealed areas like forests, gardens and crop fields are not unmanaged but most PNVs, by their nature, do not consider these kinds of human management. Comparing the soil GHG fluxes of a PNV map with the soil GHG fluxes of our world might, however, be an interesting option for further research.

Assigning unsealed land cover types in the NA regions based on nearby

Label	Code
Continuous urban fabric	1
Discontinuous urban fabric	2
Industrial or commercial units	3
Road and rail networks and associated land	4
Port areas	5
Airports	6
Mineral extraction sites	7
Dump sites	8
Construction sites	9
Green urban areas	10
Sport and leisure facilities	11

Table 3.3: A list of CLC classes (and their codes) that we reclassified as NA.

information from the CLC map is another option. We implemented the latter option using a moving window technique. This involves moving a weighting kernel (or window) between each NA pixel in the raster and replacing the NA value based on some calculation consisting of the non-NA values within the weighting kernel. Using NA values in the weighting kernel prevents specific cells from being used. The choices of both the weighting kernel and the operation performed on the non-NA values included will influence the resulting map.

In the raster, since we have discrete land cover classes represented by integers, we have to use a function that allows only for an output from one of these same neighbouring land cover categories. This already gave us a short list of options like the mode and median. We chose the mode because the land cover classes in the CLC map are ordered but definitely not perfectly smoothly. Therefore, choosing an intermediary value could assign rare land cover classes to a cell only because there are two similarly common classes, that are both more likely candidates, in the nearby region. The mode alleviates this issue somewhat, but still could assign

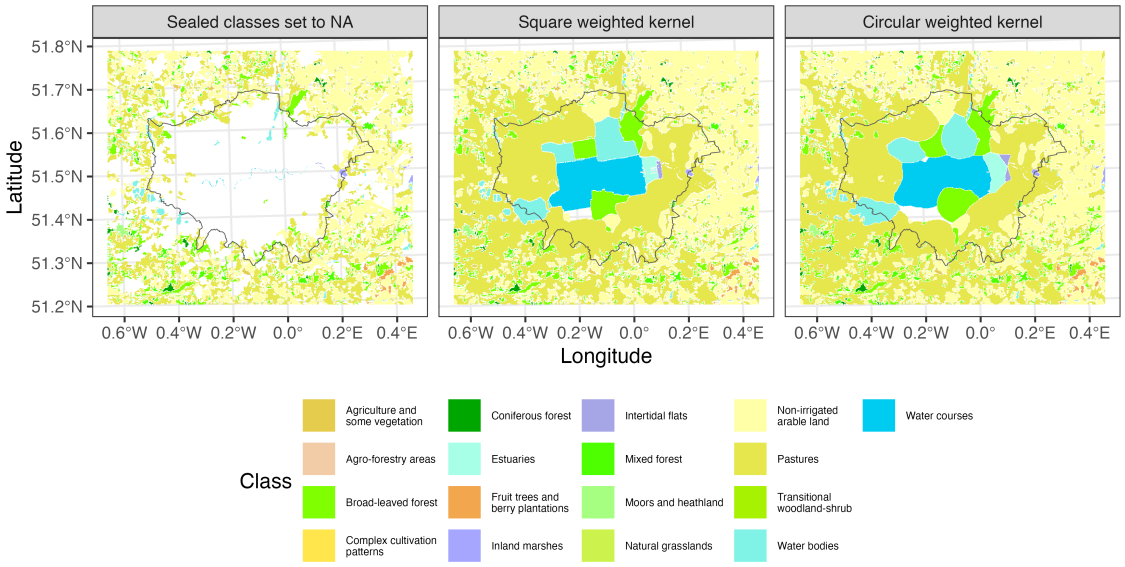

Figure 3.2: On the left we have a UK CLC map after reclassifying all sealed classes to NA. Then we show the results of moving window calculation using a square and circular weighted kernel. Some land cover class names have been shortened.

unlikely values to cells.

The weighting kernel determines how important the cells neighbouring NA regions are when estimating their value via the mode. The circular weighting kernel we used was similar to the matrix W_{9x9} in Equation 3.17 but of a larger size - 101x101. The rectangular matrix used as a comparison in Figure 3.2 was identical to the circular weighting kernel but with all elements equal to 1. Within a large urban area like London, the square weighting kernel creates less natural, linear boundaries between land covers. The boundaries created by the circular weighted kernel are preferable, as we expect vegetation to grow outward in all directions equally from a point. It is this vegetation that we are reclassifying NA regions with, to produce an unsealed version of the UK.

There are still issues with the map of London created by the circular weighted kernel in Figure 3.2. The most striking of these is the unclassified region below the large water course at the centre of London. This water course itself is also not desirable as it is assuming that River Thames would be occupying a huge region of sealed soil in London. We might expect the Thames to be larger without sealed land, but certainly not this much larger. The problem here is that the Thames is the dominant non-sealed land cover in central London at 100 m resolution. We address these problems for the UK's largest cities by reclassifying water courses in these regions as NA and then repeating the moving window technique but only for the UK's largest cities - London, Birmingham, Manchester and Glasgow. We masked this layer containing natural versions of the major cities over the first map produced by the circular weighted kernel (the circular weighted kernel in Figure 3.2). This gave us an unsealed land cover map at 100 m resolution.

$$W_{9x9} = \begin{bmatrix} NA & NA & NA & NA & 1 & NA & NA & NA & NA \\ NA & NA & 1 & 1 & 1 & 1 & 1 & NA & NA \\ NA & 1 & 1 & 1 & 1 & 1 & 1 & 1 & NA \\ NA & 1 & 1 & 1 & 1 & 1 & 1 & 1 & NA \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ NA & 1 & 1 & 1 & 1 & 1 & 1 & 1 & NA \\ NA & 1 & 1 & 1 & 1 & 1 & 1 & 1 & NA \\ NA & NA & 1 & 1 & 1 & 1 & 1 & NA & NA \\ NA & NA & NA & NA & 1 & NA & NA & NA & NA \end{bmatrix} \tag{3.17}$$

3.4 Spatial resampling

The three raster products we started with had different grid cell sizes (Table 3.4). We need to manipulate these grid sizes so that these products can be combined effectively allowing us to calculate the effect of sealing. Generally, to consider as much information as possible it makes sense to resample to the coarsest resolution

Product	Resolution	Year
Unsealed land cover	100	2018
IMD Degree of Imperviousness	10	2018
WorldClim UK monthly temperature	700	2018

Table 3.4: Spatial products used, their resolution and the year they represent. We derived the unsealed land cover map using the CLC in Section 3.3.

dataset available. This was the 700m resolution (from the WorldClim temperature data). However, we chose the 1km instead of 700m resolution. We did this to allow additional soil characteristics, which often have a 1km resolution, to be easily integrated into the method in future research. Producing a new representation of rasters at this 1km resolution is where spatial resampling (henceforth resampling) - came in.

Parker, Kenyon, and Troxel (1983) splits the process of image resampling into two parts. First, there is the fitting of an interpolating function based on the signal, or original image, alone. Keys (1981) highlights that this interpolating function has the fundamental property of intersecting with every element of the input signal. Secondly, the sampling function, based on the resolution desired, samples the interpolating function at a particular number of evenly spaced points. The interpolating function is continuous, and therefore can be sampled using any grid desired to produce a new representation of the input raster, now with a different grid cell size.

There are a selection of different options we can use for resampling. Different resampling techniques could significantly alter results. We will now go into further detail on some common resampling techniques: nearest-neighbour, bilinear, cubic, and aggregation.

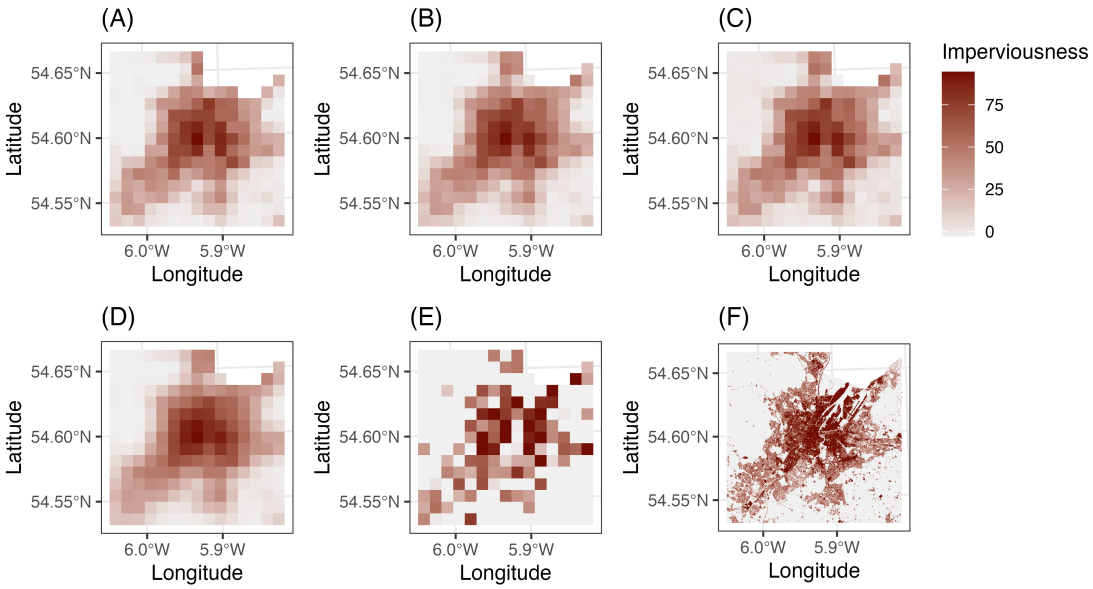

Figure 3.3: 2018 IMD in Belfast following resampling. Different resampling methods are shown by aggregated (A), bilinear (B), cubic (C), cubic spline (D), and nearest-neighbour (E). (A)-(E) are all at 1000 m resolution. The 10 m resolution raw data is shown in (F).

3.4.1 Nearest-neighbour

Nearest-neighbour resampling simply assigns cells in the new grid the value of the nearest cell in the original raster. The 'nearness' here is measured using Euclidean distance. So for new point, $a = (x_a, y_a)$ and four neighbouring points: $B = \{b_1, b_2, b_3, b_4\}$, where $b_i = (x_i, y_i)$, we have:

$$j := \arg\min_i \left(\sqrt{(x_a - x_i)^2 + (y_a - y_i)^2} \right). \tag{3.18}$$

The value assigned to the new point a is then the value of point b_j (the point with the smallest Euclidean distance from a).

This property is ideal for discrete data because it will only assign pre-existing values to new raster cells. For instance, if we have two classes represented by 1 and 2, there is no way of assigning some value 1.1 to a new raster when using nearest-neighbour. Figure 3.3(E) shows this, with a particularly distinct blocky output and no smoothing. Table 3.5 also shows that nearest neighbour is the only algorithm that classifies pixels in the new raster as 100 - totally impervious. This would be very unlikely using the resampling methods to come.

When transitioning from suburban areas to the countryside, a nearest neighbour technique might be more desirable because the transition is somewhat crude. For example, the imperviousness degree of soil in both the middle of the countryside and in a city's green belt should be around 0. Provided these regions are sufficiently large to be captured by the 1 km^2 cell, they should both be 0. Nearest neighbour will classify these as 0, but other interpolation techniques are more likely to estimate the grid cell in a city's green belt above 0, since they consider the values of points a little further than the nearest point. In this way the large number of 0 values in the IMD data is likely better represented by nearest neighbour than the other methods. However, this does not make nearest neighbour a good choice for IMD. Nearest-neighbour's average imperviousness in the Belfast region is most different to the raw data.

Since it is this overall representation of the 10 m resolution raster that is really of

value in our case, nearest-neighbour is not appropriate for the imperviousness data. Furthermore, another technique (aggregation) can capture more of the complexity of discrete datasets than nearest neighbour. Thus, nearest-neighbour was not used to resample the unsealed map to 1km resolution either.

3.4.2 Bilinear

As the name suggests, bilinear resampling relates to a linear relationship between points in two (perpendicular) directions - say x and y. So for each x value in the input raster, a linear relationship across the full range of y values can be interpolated. However, this only gives 2-dimensional slices of a 3-dimensional surface. To guarantee that we can resample a new set of points at any linear combination of x and y, we need this 3-dimensional surface. This is derived by interpolation again, but now between these 2-dimensional slices. Sampling points in this 3-dimensional surface with a new grid of cells then produces a new, resampled, raster.

One issue with bilinear interpolation shown by Parker, Kenyon, and Troxel (1983) is that isolated details in an image can be smoothed out of the resampled image. For example, they show that details in the iris of a human eye are completely smoothed out following bilinear interpolation. In our case, we can see from Figure 2.5 that IMD data can be comprised of small patches (i.e. small farms in rural areas). Bilinear interpolation risks removing these patches entirely.

When representing 700m resolution monthly temperature data at 1km resolution we used bilinear resampling. The spatially continuous nature of temperature data suits bilinear resampling well. Plus, bilinear resampling has an added benefit. Values in the output raster are guaranteed to fall within the range of values in the input raster. In our case, all the average monthly temperatures in the input raster are in the interval [-2.5°C, 18.4°C]. With bilinear resampling, we can be assured that no grid cells will take values outside this interval. Therefore, we will not have any extreme temperatures in the 1km raster, avoiding inaccuracies in our temperature-

adjusted fluxes.

3.4.3 Cubic

Keys (1981) describes cubic interpolation methods as ideal when using continuous spatial data. Like bilinear, cubic spline interpolation smooths rasters (Parker, Kenyon, and Troxel, 1983) so it makes most sense if the aspect of Earth's surface it approximates is continuous. Parker, Kenyon, and Troxel (1983) adds that cubic spline interpolation, is quite capable at accounting for small details. Cubic methods differ from bilinear interpolation by considering a larger number of nearby pixels - 16 (Han, 2013). With a cubic relationship first assumed between rows (or columns) points in the signal. Accordingly, some cubic resampling algorithms can produce results with values outside the allowed range of the signal. This characteristic deterred us from using a cubic method with the seasonal temperature raster. We see an example of this for the imperviousness data in Table 3.5, where the cubic method produces a minimum imperviousness of -2.92 in its 1km version of Belfast. The more important property we desire when resampling the imperviousness data is retaining as much information from the input raster as possible.

3.4.4 Aggregation

Aggregation involves computing some function, often the sum, mean, maximum or minimum on all pixels within a new, larger grid cell. This result is allocated to the new grid cell. If we use the mean, this technique crucially preserves most of the average information provided by the original raster. For discrete rasters, we can sum the occurrences of each class within $1km^2$ grid cells. These sums are assigned to the $1km^2$ grid cells, giving 1 km layers for each class.

We used aggregation with this sum technique to convert the 100 metre resolution unsealed map into a proportional raster. This gave us 33 different 1 km resolution rasters. Combining these rasters gave us the unsealed map that we would assign GHG fluxes to.

Resampling	Mean	Minimum	Maximum	Median
Aggregated	25.79	0.00	94.32	18.09
Bilinear	25.91	0.00	91.41	19.33
Cubic	25.90	-2.92	94.48	19.34
Nearest-neighbour	24.00	0.00	100.00	0.00
IMD 10m	25.51	0.00	100.00	0.00

Table 3.5: Summary of the Belfast imperviousness data following different kinds of resampling from 10 to 1000 squared m resolution.

Aggregation was also implemented on the 10 metre resolution imperviousness data. This time, a different type of aggregation was used - the mean of the 10000 pixels within each new $1km^2$ cell. We therefore have the same average imperviousness value as in the original 10 m resolution dataset - except for incomplete grid cells containing NA values (i.e. on the coast). Consequently, the proportional land cover raster we produced retains much of the average information of the original, higher resolution rasters.

3.5 Accounting for roads

As we showed in Figure 2.5, due to its 10 m resolution, IMD data does not account for the majority of roads. Roads missing from IMD data were therefore included by assigning a sample of the 1km IMD data a value corresponding to completely impervious land. We had 2514 km^2 of roads and our spatial data frame has rows corresponding to a 1 km^2 cell. So, the number of elements in this sample corresponded to roughly the area taken up by roads in the UK. This area was calculated using data from UK Department for Transport and Ministry of Housing, Communities & Local Government (2007) and UK Department for Transport (2022). Motorways - and most roads surrounded by an urban area do still get picked up in the IMD data. Figure 3.4 shows this, where the M6 through Lancaster is identified

GHG	Q_{10}
$\mathbf{CO_2}$	2.4
$\mathbf{CH_4}$	4.0
$\mathbf{N_2O}$	6.0

Table 3.6: Q_{10} used for GHGs values. Taken from Oertel et al. (2016), Dalal and Allen (2008), and Abdalla et al. (2009).

as impervious in the zoomed in window. Furthermore, as with Figure 2.5, the imperviousness data does not capture Lancaster's country roads.

3.6 GHG flux relationship with temperature

Adjusting our estimates for temperature required data on temperature in the UK over time, and the sensitivities of each GHG, and perhaps land cover, to a change in temperature - likely using Q_{10}. For the monthly temperature data we used the climate research unit time-series 4.06 (CRU-TS; Harris et al. (2020)) downscaled with WorldClim 2.1 Fick and Robert J Hijmans (2017) and acquired via the R package, geodata (v0.6-2; Robert J. Hijmans et al. (2024)). We exclusively used the temperature data from 2018 to correspond with our the 2018 versions of CLC and IMD. Like Bond-Lamberty and Thomson (2010), we use air temperature rather than soil temperature data due to accessibility. We expect that the difference between average yearly temperature (T_1) and the temperature in each month (T_2) is likely to be quite similar for both soil and air temperature datasets. Thus, this seems unlikely to significantly alter results. It is this difference, $(T_2 - T_1)$, that we are primarily concerned with in Equation 3.19, which is just a rearranged form of Equation 2.1.

We experimented with using the supplementary data from Yan et al. (2022) to calculate Q_{10} values for particular land covers. We filtered these values to be as close to the UK as possible. However, with many Q_{10} values significantly outside of the range of values for Q_{10} provided in Oertel et al. (2016), it seems using their data to

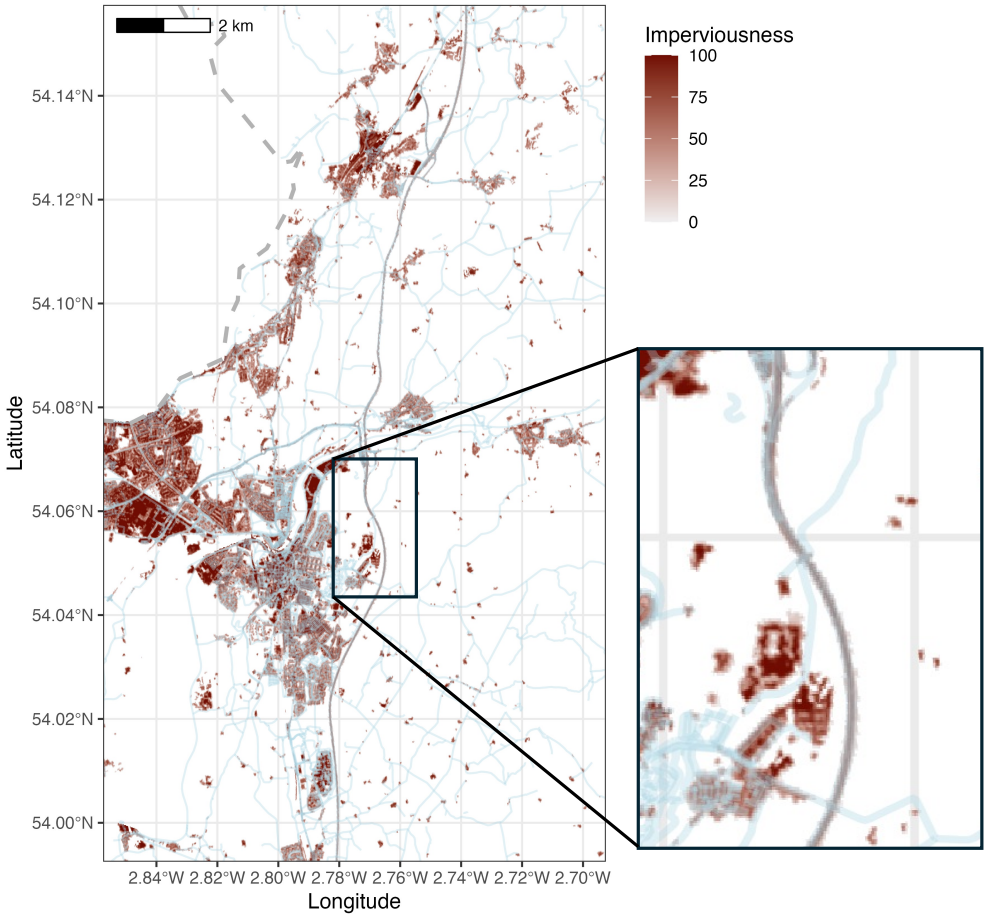

Figure 3.4: OSM (Mark Padgham et al., 2017) road data alongside 2018 IMD in Lancaster and Morecambe. The thick dashed line is the west coast of the England, north of Morecambe. We zoom in on a portion of motorway to illustrate that this is classified correctly as impervious.

produce Q_{10} like this might require further research. Consequently, for our measure of temperature sensitivity we proceeded with Q_{10} values shown in Table 3.6.

In Equation 3.19, the GHG flux rate prior to adjustment is R_1. We can then obtain R_2, the temperature adjusted GHG flux for a given month in 2018 using Equation 3.19. The average yearly temperature in 2018, T_1, will remain constant for whichever GHG and month. T_2 will change according to the month and the Q_{10} value corresponding to the GHG.

$$R_2 = R_1 Q_{10}^{(T_2 - T_1)/10°C} \tag{3.19}$$

3.7 Annual sealed (and unsealed) soil GHG fluxes

We have now described the majority of the preprocessing we undertook. We have produced the unsealed land cover map; converted this to a proportional representation at 1 km^2 resolution; averaged and inverted the imperviousness data and finally simulated GHG fluxes for different land covers using the typical GHG flux bounds from Oertel et al. (2016). Now, we had to bring these elements together into one dataset and can calculate the annual sealed, and unsealed fluxes for the UK.

This process involves carefully scaling our fluxes so that the units are in a desirable form. Now, the fluxes from Oertel et al. (2016) are measured in either:

$$A = \mu\text{mol}\,CO_2\,m^{-2}\,s^{-1}\,,$$
$$B = \mu\text{mol}\,CH_4\,m^{-2}\,h^{-1}\,,$$
$$C = \mu\text{mol}\,N_2O\,m^{-2}\,h^{-1}\,.$$

We needed the flux rates in g yr^{-1} km^{-2} because these units allow us to easily calculate the total UK fluxes. So, we must convert from micromoles to grams, metres to kilometres and minutes (or seconds) to years. We show this conversion

for CO_2 and CH_4 or N_2O in Equation 3.21, where the scaling factors for each GHG are calculated using the molar masses, and units of CO_2, CH_4 and N_2O respectively.

$$\text{X g km}^{-2}\ \text{yr}^{-1} = \begin{cases} 3600 \times 24 \times 365 \times 44.009 \times A & \text{if } X \text{ is } CO_2 \\ 24 \times 365 \times 44.013 \times B & \text{if } X \text{ is } CH_4, \\ 24 \times 365 \times 16.043 \times C & \text{if } X \text{ is } N_2O. \end{cases} \tag{3.20}$$

$$= \begin{cases} 1387867824 \times A & \text{if } X \text{ is } CO_2, \\ 385553 \times B & \text{if } X \text{ is } CH_4, \\ 140536 \times C & \text{if } X \text{ is } N_2O. \end{cases} \tag{3.21}$$

In line 3.20, we cancelled the scaling factors for the SI units of micromoles (1×10^{-6}) and squared metres to squared kilometres (1×10^{6}).

From here, our method depends on how we treat the relationship between the fluxes and other variables in each square kilometre. In the case linear case, we do not consider seasonal adjustment of soil GHG fluxes. Therefore, for each land cover, we calculated the element-wise sum of the area they cover and the average degree of imperviousness (Table 3.7). For the sealed map, we then multiplied the fluxes (scaled as shown in Equation 3.21) by these two values, whereas in the unsealed case we simply did not multiply by the average imperviousness for each land cover.

When considering seasonal adjustment of soil GHG fluxes we had a non-linear relationship between the fluxes and the average monthly temperature. These results, therefore, were much more difficult to produce computationally. We had to calculate everything per squared kilometer, combining only once our fluxes had been adjusted by the temperature.

Land cover	Unsealed area (km^2)	Proportion	Avg. IMD	Sealed area (km^2)
Cropland	77252.17	0.324	0.940	72133.32
Barren	24107.53	0.107	0.984	23806.34
Grassland	90208.86	0.382	0.956	85063.61
Wetland	22292.01	0.100	0.994	22250.92
Forestland	20185.46	0.088	0.964	19552.45

Table 3.7: The unsealed area of each land cover class and the proportion of this they account for. The Avg. IMD is the average inverted imperviousness for each class. Finally, the sealed areas are the element-wise sum of the UK imperviousness data and the UK proportional land cover data (ϕ = Sealed area in Equation 3.6). We used this column to adjust the fluxes for sealing in the linear case. The sealed area is the effective area of the UK that the land covers account for, when each square kilometre is scaled by IMD.

Chapter 4

Results

We are interested in the overall impact of soil sealing on total UK GHG emission. To measure this, we used Equation 4.1, the flux not emitted due to sealing as follows:

$$\text{diff}^{\text{sealing}} = \text{GHG}^{\text{unsealed}} - \text{GHG}^{\text{sealed}}. \tag{4.1}$$

We chose to subtract the terms in this way for convenience, solely because the result was more often positive like this. The value of $\text{diff}^{\text{sealing}}$ for CH_4 was overall negative, representing an effective emission in CH_4. Alongside estimates, we consider the uncertainty of $\text{diff}^{\text{sealing}}$ using confidence intervals computed via the sealed and unsealed simulations. Therefore, these confidence intervals are dependent on the way that we setup our simulation initially. We will address the implications of this in the Discussion later. We will contextualise the results for CO_2 and the CO_2 equivalence (CO_2e) using the emissions database for global atmospheric research (EDGAR) from Crippa et al. (2020), specifically data for emissions from specific industries in countries.

We split our results in three ways: the average difference from sealing in Section 4.1, then the seasonal difference in Section 4.2 and finally the difference depending on different levels of correlation between CO_2, CH_4 andN_2O in Section 4.3.

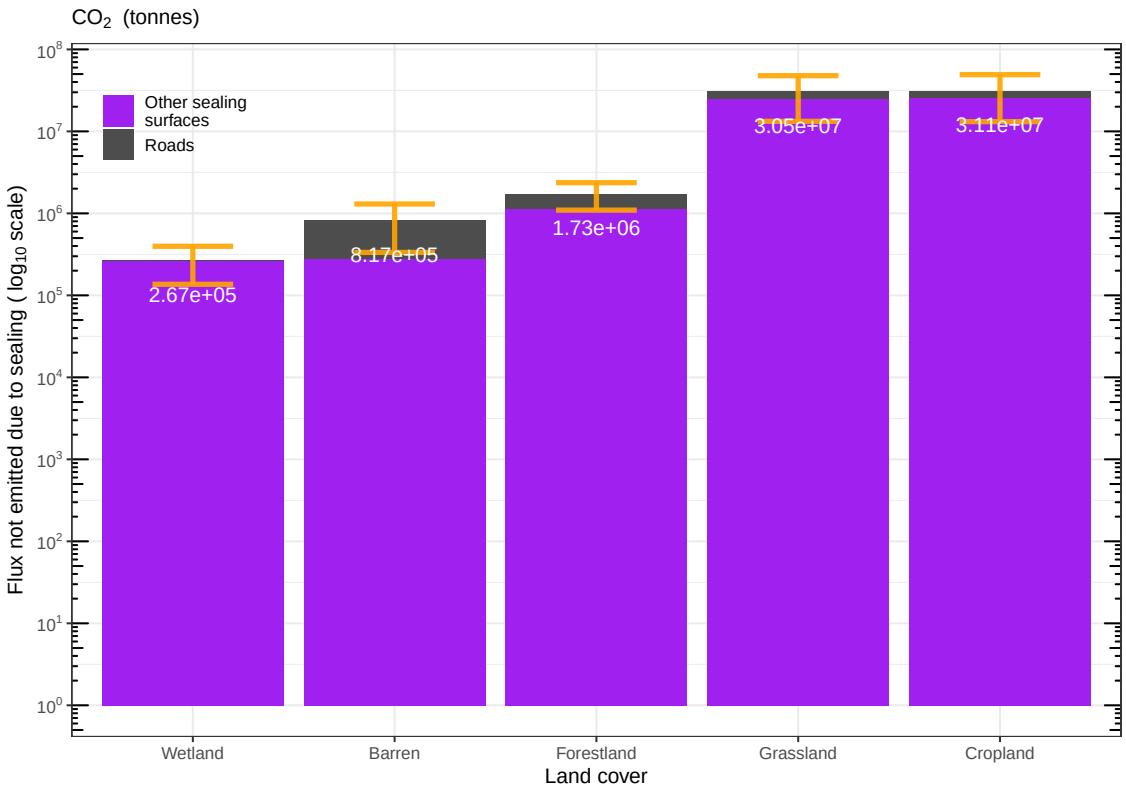

Figure 4.1: The annual saving in CO_2 (tonnes) from sealed soil in the UK. The y-axis is ${CO_2}^{\text{unsealed}} - {CO_2}^{\text{sealed}}$ and shown on a base-10 logarithmic scale. The error bars are the 95% confidence intervals based on 1000 simulations. The bars are filled in based on the type of impervious surface causing the difference in emissions.

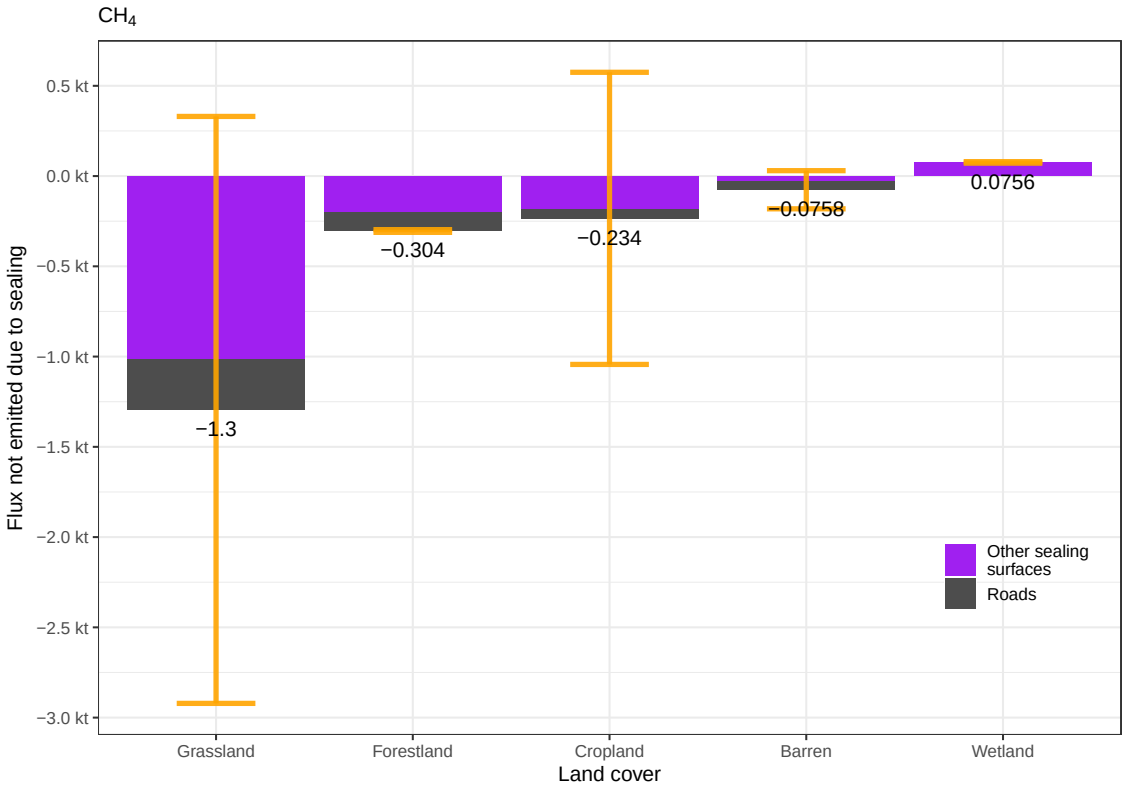

Figure 4.2: The annual saving in CH_4 (kilotonnes) from sealed soil in the UK. For CH_4, the magnitude of these mostly negative savings are really emissions caused by sealing. The y-axis is ${CH_4}^{\text{unsealed}} - {CH_4}^{\text{sealed}}$. The error bars are the 95% confidence intervals based on 1000 simulations. The bars are filled in based on the type of impervious surface causing the difference in emissions.

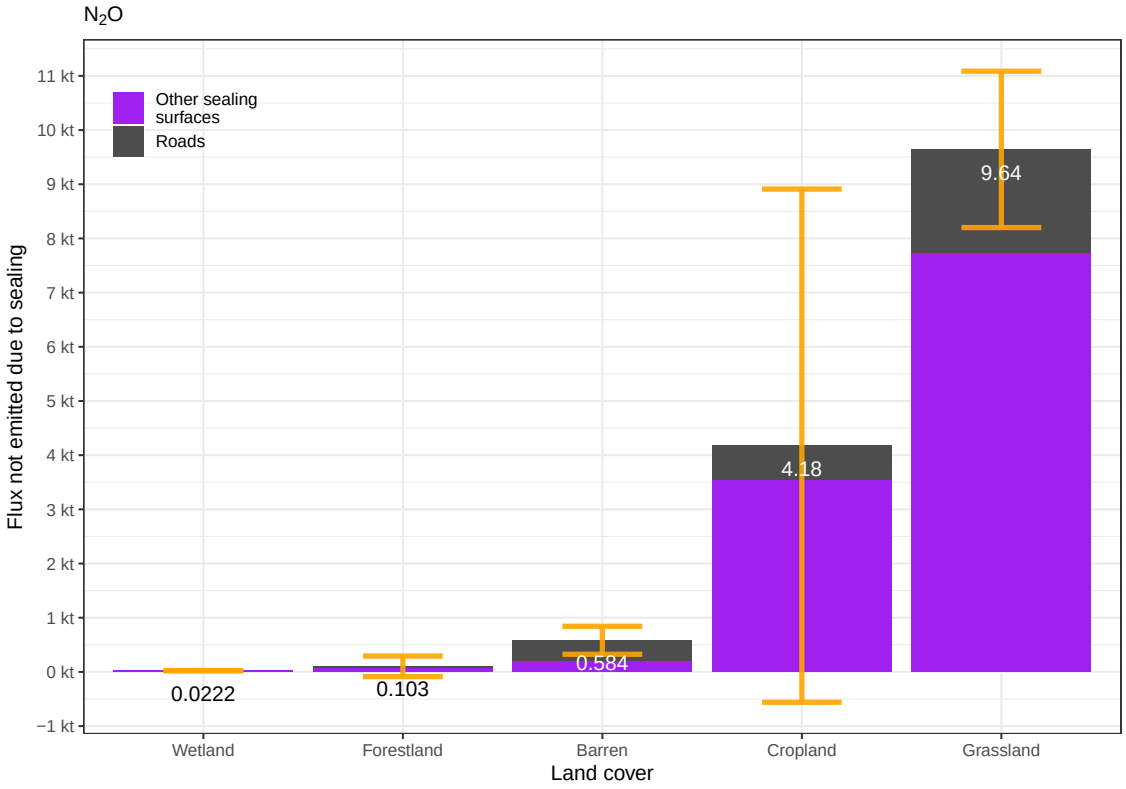

Figure 4.3: The annual saving in N_2O (kilotonnes) from sealed soil in the UK. The y-axis is $N_2O^{\text{unsealed}} - N_2O^{\text{sealed}}$. The error bars are the 95% confidence intervals based on 1000 simulations. The bars are filled in based on the type of impervious surface causing the difference in emissions.

4.1 Annual saving in GHG fluxes from sealing

In Figures 4.1 to 4.3, the mean estimate of each land cover is as a number written on the bars. The portion of the bars labelled with 'other' is the saving from all imperviousness data. The portion of the bars labelled with 'roads' is the saving from the inclusion of non-major roads using OSM data. The error bars represent a 95% confidence interval. All data shown here is based on 1000 simulations with 0 correlation between CO_2, CH_4, N_2O ($\rho_{1,2} = \rho_{1,3} = \rho_{2,3} = 0$).

So far, our results give an idea of the CO_2, CH_4 and N_2O fluxes resulting from sealing over land, but do not account for the overall global warming potential of these three gases. To address this, we also used the CO_2 equivalents (CO_2e) of the GHGs - shown in Table 4.1. It should be noted that these global warming potential (GWP) values change as new IPCC assessment reports are released, so these values are certainly not set in stone. Moreover, using different time horizons will change results. These GWP values are then paired with the corresponding GHGs as coefficients in Equation 4.2 (Yan et al., 2022; Brander and Davis, 2012) to calculate the CO_2e of all three gases. So, converting some quantity of a GHG, X, to its CO_2e simply involves multiplying X by its GWP. From here it is comparable to CO_2 and other GHGs that have been converted to their CO_2 equivalent emission.

$$CO_2e = CO_2 + 81.2 \times CH_4 + 273 \times N_2O \tag{4.2}$$

Figure 4.5 provides a clearer comparison of the CO_2e saving in soil GHG emissions due to sealing. The importance of the CO_2 saving is an order of magnitude greater than either the CH_4 emission or N_2O saving. For context, the total CO_2e saving of all 3 GHGs due to sealing is roughly the same as the 2018 GHG emissions from fossil fuels in Greece (68.4 Mt CO_2 yr^{-1}, Crippa et al., 2020).

We can also see in Figure 4.5 that the overall confidence interval for CO_2 is an order of magnitude greater than the other fluxes. This is then the primary source of the uncertainty when considering for all three GHGs. Relative to the magnitude of

GHG	GWP $kgCO_2e/kg$ GHG
CO_2	1.0
CH_4	81.2
N_2O	273.0

Table 4.1: Global warming potentials (GWP) of the three GHGs we used here in terms of the carbon dioxide equivalent (CO_2e). The data here are at the 20-year time horizon from report 6 of the International Panel for climate change assessment (AR6-20, IPCC (2021)). The data we used were the output of the LICA formatter from Young et al. (2021)

the point estimates, both CO_2 and N_2O have similar uncertainties. The uncertainty of CH_4, however, is about same magnitude as its point estimate.

The main takeaway from Figure 4.4 is the similarity between it and the results for just CO_2 in 4.1. This means that, relative to CO_2, the annual flux from CH_4 and N_2O are quite small. All of the land cover classes except Forest had an increased annual flux when we looked at the CO_2e. Forest land cover type, however, was around 1 million CO_2e lower. This comes from Forest soils having a significant negative flux (e.g., sink) of -6.29 $\mu mol\, CH_4\, m^{-2}\, h^{-1}$.

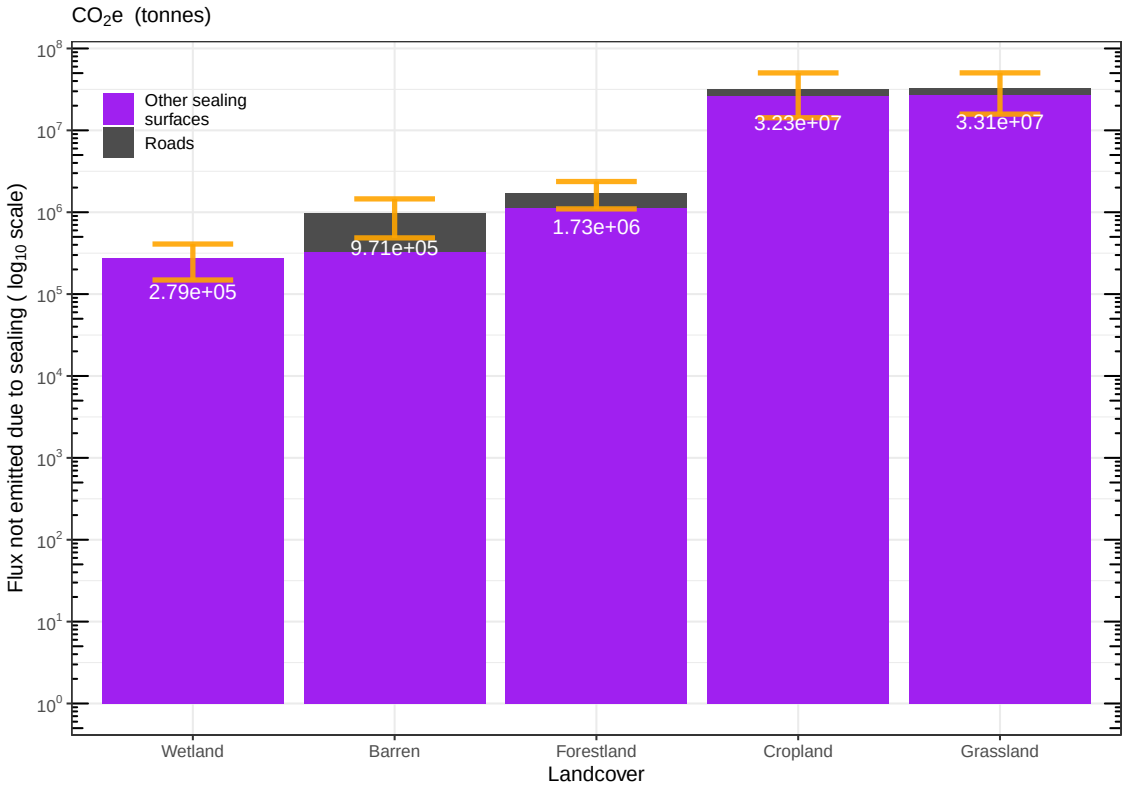

Figure 4.4: The annual saving in CO_2e (tonnes) from sealed soil in the UK. The y-axis is $CO_2e^{\text{unsealed}} - CO_2e^{\text{sealed}}$ and shown on a base-10 logarithmic scale. The mean estimate of each land cover is as white text on the bars. CO_2e is calculated using the AR6-20 GWP for CO_2, CH_4 and N_2O as shown in Equation 4.2. The error bars represent a 95% confidence interval. Data used here is based on the same simulated fluxes with 0 correlation used in Figure 4.1.

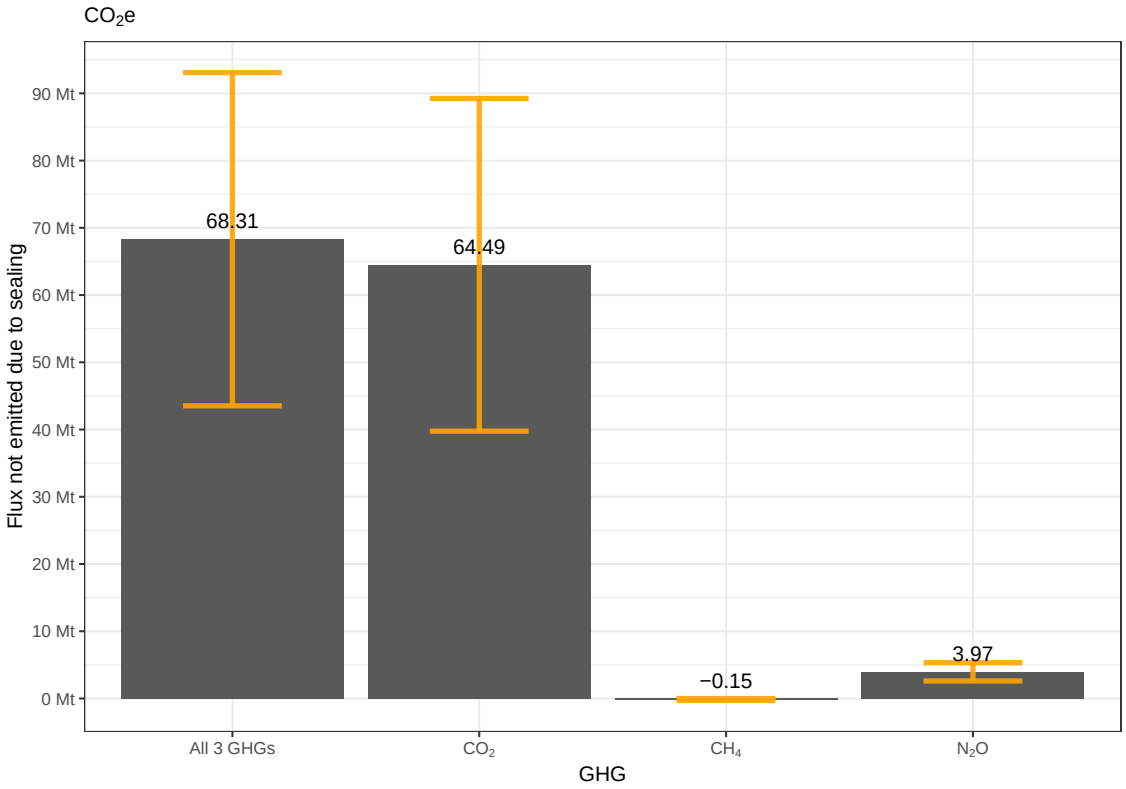

Figure 4.5: The annual saving due sealed soil in the UK for CO_2, CH_4 and N_2O and all three GHGs . The y-axis is in terms of the CO_2e (megatonnes) of the GHGs. These annual savings have been summed across the 5 land cover types. The text on the bars represents the annual flux (megatonnes) and the orange error bars are the 95% confidence interval based on 1000 UK GHG flux simulations.

Further context is added to this in Figure 4.6, which considers the differences in CO_2e emission savings across land covers. This highlights the relative importance of CO_2 emission savings from both cropland and grassland. Even the lower bound of the confidence intervals for these two bars are dominant. For CO_2, the yearly savings for the other land covers are in line with N_2O savings for grassland and croplands. Savings from the other N_2O land covers are hardly effected by soil sealing when compared to the CO_2. The same applies to CH_4, with the emissions caused by sealing grassland, forestland and cropland, completely counteracted by CO_2 savings due to sealing.

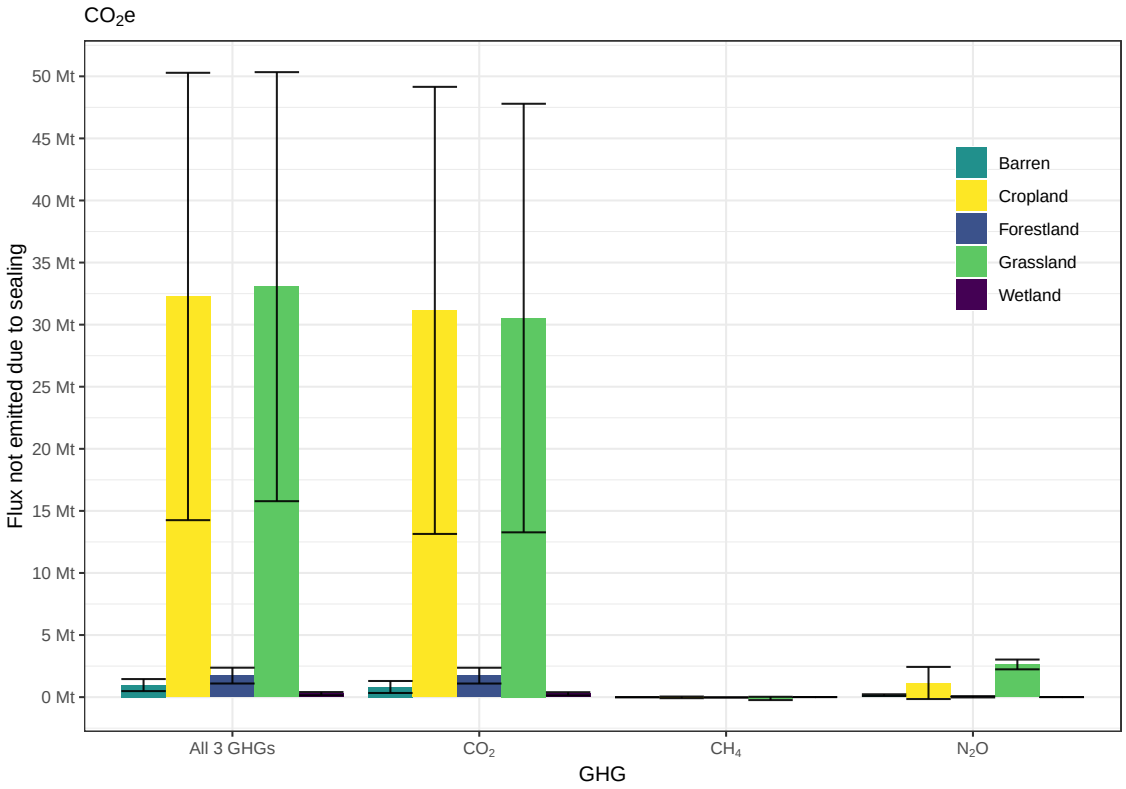

Figure 4.6: The annual savings due to sealed soil for CO_2, CH_4 and N_2O, split according to land cover. The y-axis is in terms of the CO_2e (megatonnes) of the GHGs for ease of comparison. The orange error bars are the 95% confidence interval based on 1000 UK GHG flux simulations.

4.2 Seasonal saving in flux from sealing

Our seasonal results account for the change in average temperature throughout the UK in the year 2018. We show how the soil GHG flux for each land cover is affected by temperature in Figures 4.7 to 4.10.

For CO_2 and N_2O, we observe more extreme savings in the seasons with more extreme weather. For CH_4, this saving is now an intake, with a more extreme intake of CH_4 during the winter and summer.

However, Figure 4.4 shows that the CO_2e of seasonal CH_4 intake - and the corresponding confidence interval - is very small in magnitude relative to N_2O and almost unnoticeable compared to CO_2.

Figure 4.2 indicates that the uncertainty of seasonal CH_4 emissions specifically is very extreme relative to the point estimates. This is particularly true for grassland with the 95% confidence interval is extremely wide at $-1.3 \pm 1.62 CH_4$ kt yr^{-1}. However, if we put this in terms of the CO_2e this might have a very little effect next to CO_2 - or even N_2O - savings from sealing.

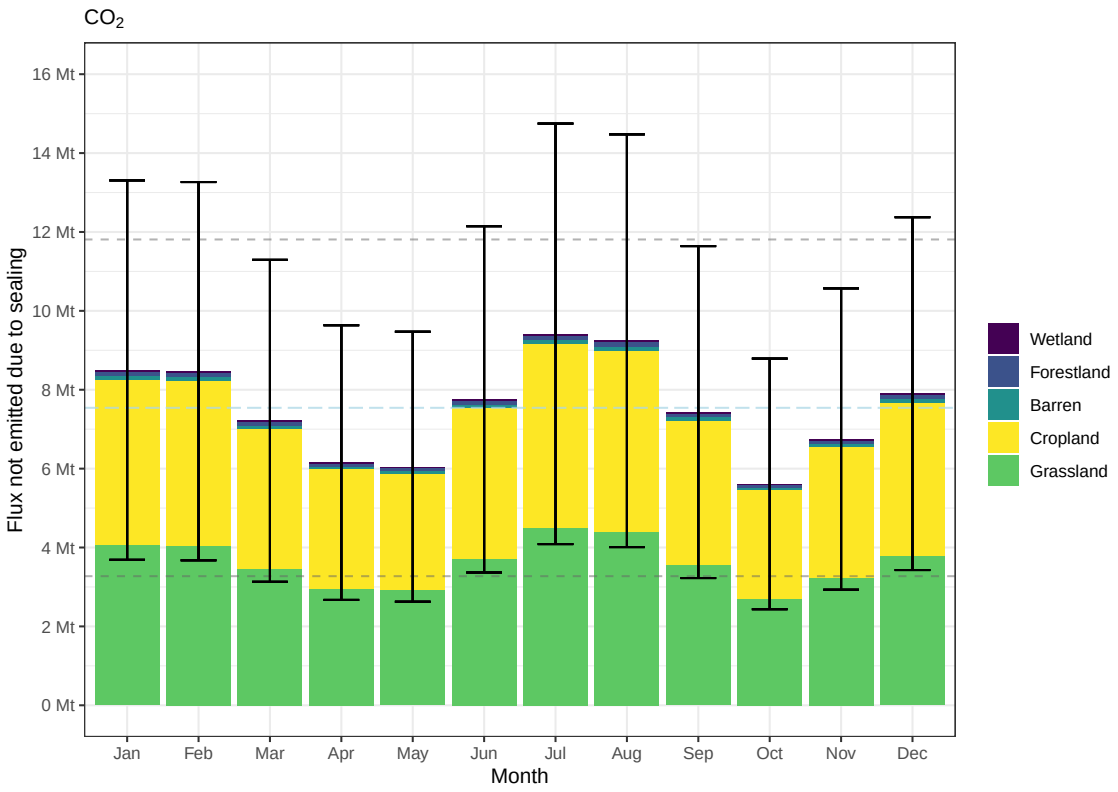

Figure 4.7: The monthly saving in CO_2 (megatonnes) from sealed soil in the UK. The light blue dashed line shows the mean total monthly GHG saving in the UK. The lower and upper grey dashed lines show the lower and upper bounds of a 95% confidence for this average.

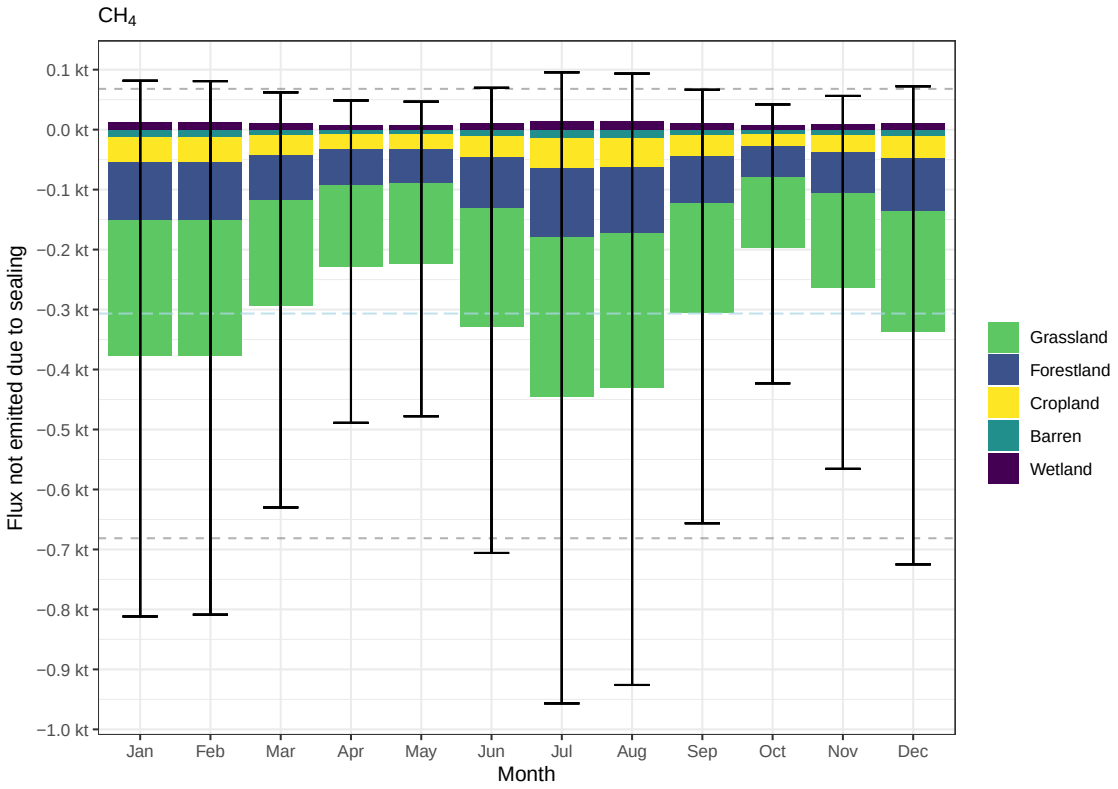

Figure 4.8: The monthly saving in CH_4 (tonnes) from sealed soil in the UK. For CH_4, the magnitude of the negative savings here represent the emissions from sealing. The light blue dashed line shows the mean total monthly GHG saving in the UK. The lower and upper grey dashed lines show the lower and upper bounds of a 95% confidence for this average.

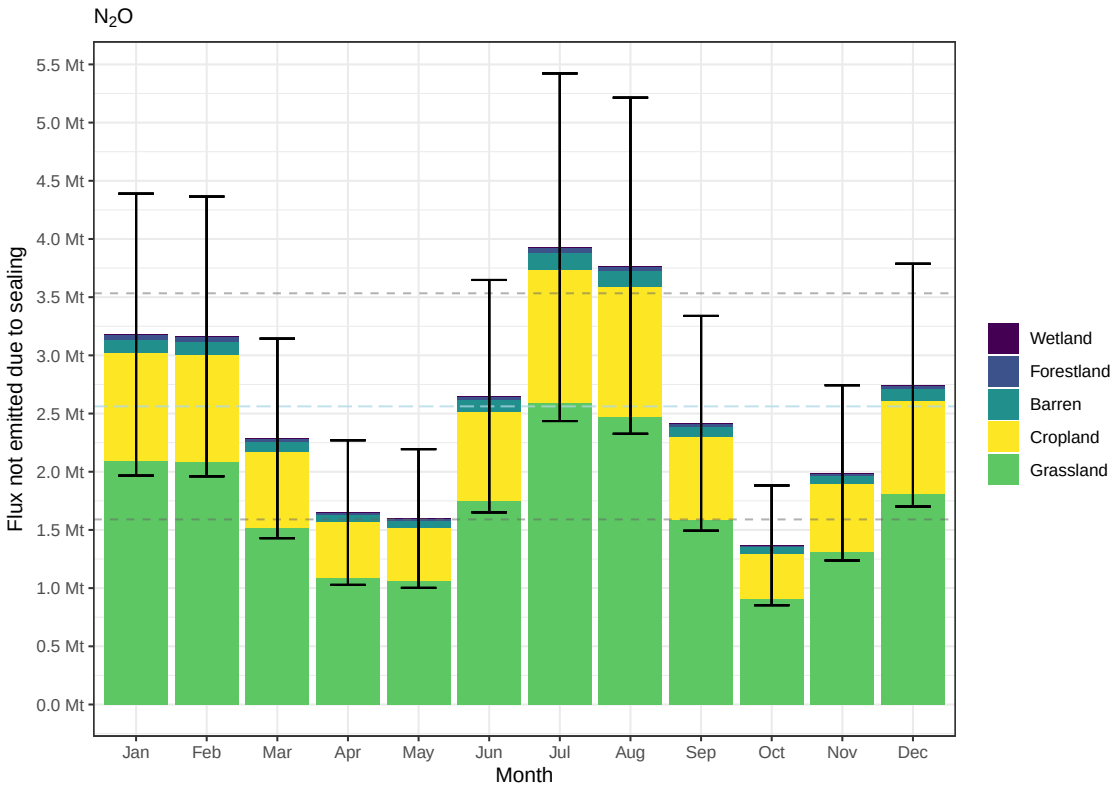

Figure 4.9: The monthly saving in N_2O (tonnes) from sealed soil in the UK. The light blue dashed line shows the mean of the total monthly GHG saving due to sealing; the lower and upper grey dashed lines show the lower and upper bounds of a 95% confidence interval about this.

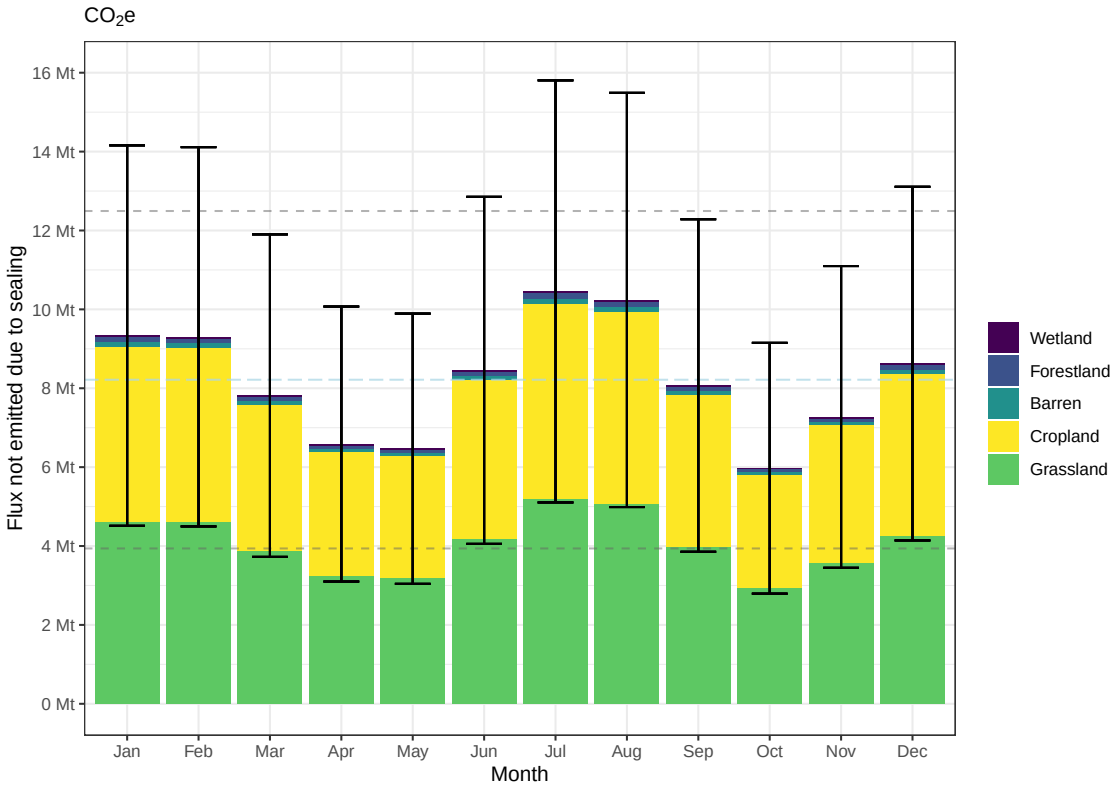

Figure 4.10: The monthly saving in CO_2e (megatonnes) from sealed soil in the UK. We have summed the CO_2e of seasonal savings in CO_2, CH_4 and N_2O to see the overall effect of these GHGs. The light blue dashed line shows the mean total monthly GHG saving in the UK. The lower and upper grey dashed lines show the lower and upper bounds of a 95% confidence for this average.

We combined the data from Figures 4.7 to 4.9 by summing the CO_2e for each GHG (Equation 4.2) to explore the reduction in global warming potential from sealing over soils. This is shown in Figure 4.10. The saving from CO_2e appears to be greatest in the summer and winter seasons, particularly in hottest July and August months. Also, regardless of the month, the vast majority of the CO_2e saving from sealing came from Cropland $(47.3 - 47.9\%)$ and Grassland $(49 - 49.6\%)$. As mentioned earlier, this is again partially due to these being the dominant land covers in the UK. For the other land covers, on this scale the change in CO_2e from sealing is not recognisable.

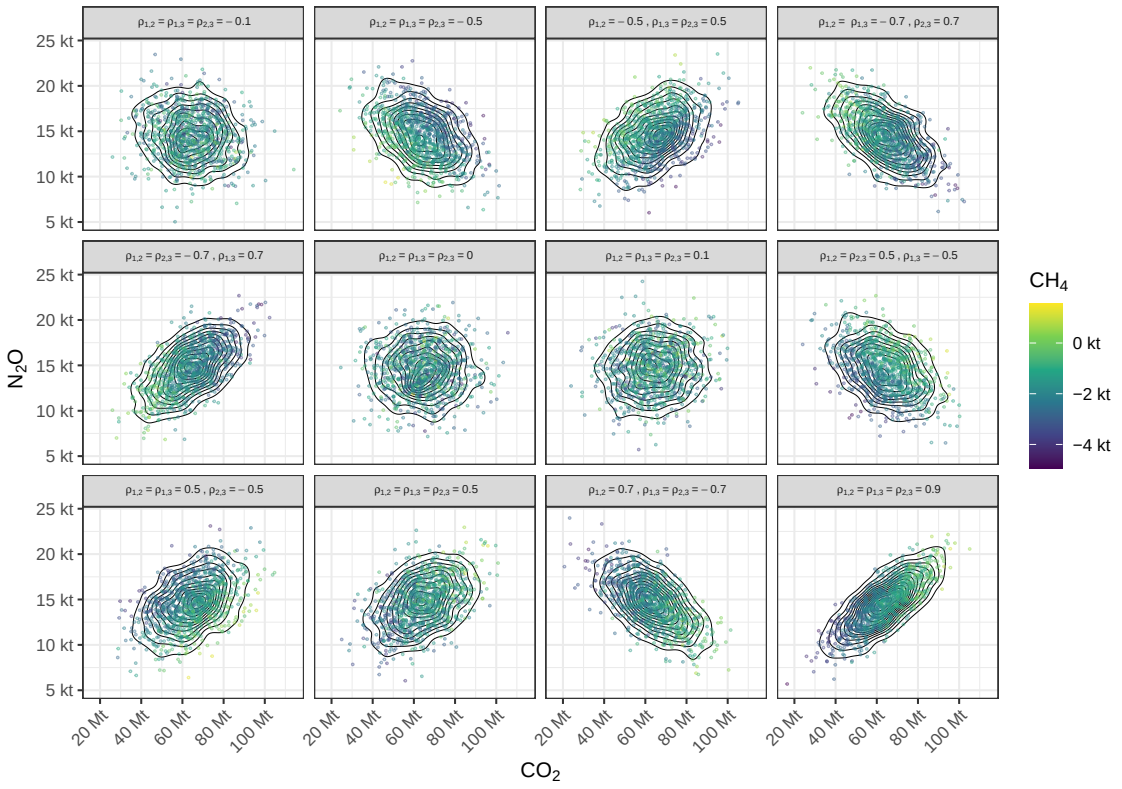

Figure 4.11: The annual saving in GHG soil fluxes from sealing given different correlation scenarios. The contours represent the 2D kernel density estimation of the saving in CO_2 and N_2O over the 1000 simulations. In these scenarios $\rho_{1,2} = \text{corr}(CO_2, CH_4)$, $\rho_{1,3} = \text{corr}(CO_2, N_2O)$ and $\rho_{2,3} = \text{corr}(CH_4, N_2O)$.

4.3 Correlation Scenarios

The different correlation scenarios in Figure 4.11 give us an idea of different ways the annual fluxes from sealing CO_2, CH_4 and N_2O may be distributed. We do expect the relationships shown here. For instance in scenario $\rho_{1,2} = \rho_{1,3} = \rho_{2,3}$ the GHGs have strong positive correlation so in simulations where one GHG is high, so are the others.

Figure 4.12 shows the difference in the 95% confidence intervals of the 2D kernel density estimation under different correlation scenarios. We generally see tighter confidence intervals for milder correlation settings. This is not a strict rule though,

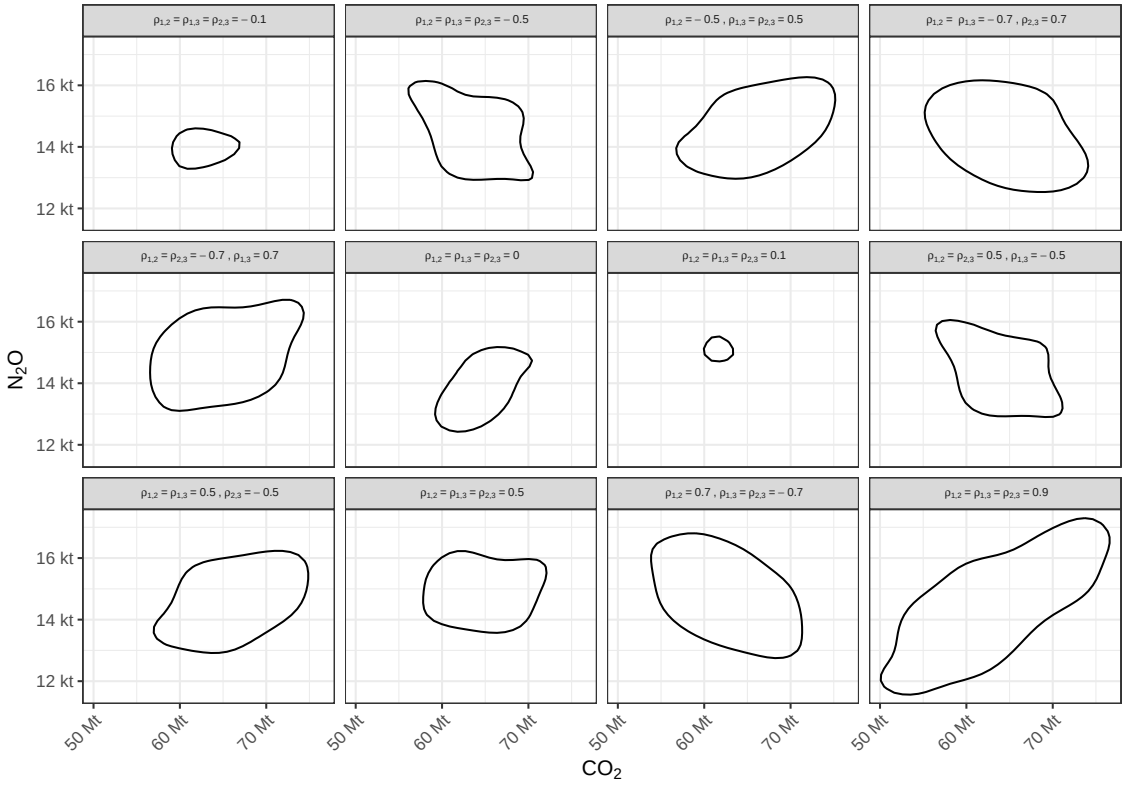

Figure 4.12: Contours showing the 95% confidence interval of the 2D kernel density estimates of CO_2 and N_2O emissions under different correlation scenarios. In these scenarios $\rho_{1,2} = \text{corr}(CO_2, CH_4)$, $\rho_{1,3} = \text{corr}(CO_2, N_2O)$ and $\rho_{2,3} = \text{corr}(CH_4, N_2O)$.

as the setting with 0 correlation between GHGs has a wider interval than some milder settings. Whilst variable in their area, all the confidence intervals still indicate towards the same results as before. There is, again, a substantial saving in soil emissions of CO_2 due to sealing and N_2O lesser but mostly positive savings.

Chapter 5

Discussion

Our results show a substantial saving in overall soil GHG fluxes due to soil sealing. The confidence intervals of our estimates were quite wide - partially owing to the manner in which we allocated the fluxes to the unsealed map. However, the lower bound of these wide intervals still represented a major aggregate saving in GHGs; the results strongly suggest that the sealing of soil causes a significant saving in CO_2e.

Our second major result is the enormous portion of this saving (90.3%) owing to CO_2 emissions from grassland and cropland. We attribute this to grassland and cropland both covering a high portion of the UK - 70% (Table 3.7), and emitting relatively high quantities of CO_2 (Table 3.1). Patterns from existing research on unsealed soil persisted in the difference between the unsealed and sealed flux we focused on (Hashimoto, 2012). This was true of the seasonal results where we saw more pronounced savings in the winter and summer.

However, all of these results are from the perspective of soil GHG fluxes. To best inform efforts to mitigate climate change, our results must be combined with the effects of sealing on the additional ecosystem services of unsealed soil.

5.1 Additional side-effects of paving

Our results pave the way for further research on the importance of additional ecosystem services of unsealed soil. It is vital that the effect of sealing on these services is accounted for. Ecosystem services that could be incorporated include flood regulation, the urban heat effect, CO_2 sequestration from plants and the destruction of animal habitats from sealing. We now give some examples of existing research into these services.

Qiu et al. (2024) measures the global impact of paving over carbon rich ecosystems. They report that the effect of expanding urban areas on biomass causes emissions in the range of 113.7 ± 20.2 Mt CO_2 yr^{-1}. Also from urban area expansion, they report an upper estimates of 274.6 ± 50.2 Mt CO_2 yr^{-1} in emissions from soil organic carbon losses. If we scale these ranges for the portion of the world's surface taken up by the UK (0.49%), and to cover all of the expansion of urban areas between 1993-2018, we can get crude estimates to compare our UK-only results against. In those 25 years alone, our estimates indicate that carbon losses from biomass and topsoil due to expanding sealed surfaces has caused emissions of 48.1 ± 8.7 Mt CO_2 yr^{-1}. We stress that this estimate covers - and is derived from - data observed over only the past 25 years. Further research could better determine the reliability of this range, perhaps by considering our GHG emission saving result from the perspective of land cover expansion. Regardless, we have a sign that the saving we calculated - 68.3 ± 24.8 Mt CO_2e - could well be cancelled out by the emissions from paving over carbon rich ecosystems because its confidence interval overlaps substantially with our estimate (48.1 ± 8.7 Mt CO_2 yr^{-1}). Settling the ambiguity here is a natural path for further research.

Xi et al. (2016) estimated that 43% of the CO_2 produced from cement production between 1930 and 2013 was offset by concrete sequestering CO_2 throughout its lifetime. However, this did not account for the emissions from the production of materials that create cement. Thus, the overall offset from the cement supply chain would be lower than 43%. The albedo effect and associated urban heat island

phenomenon will have an additional impact on the temperature of urban areas. If we were to calculate the emissions from the full life cycle of materials used for sealing, would this negate the saving in GHG emissions that our results show? Moreover, how does this change depending on the material soil is paved with?

The albedo effect of a surface is its capacity to 'reflect radiation back to the sky' (Building Services Engineers, 2022). Since, urban areas are often composed of darker materials they absorb more radiation, raising the surface temperature. Ouyang et al. (2022) estimate the global warming effect of the albedo effect following future urbanisation to be 0.001°C. This is for the next 100 years. The warming effect in cities could be significantly higher. Akbari, Matthews, and Seto (2012) discuss the long-term effect of using lighter coloured surfaces in urban areas. They estimate that a small albedo increase in urban areas (0.1) could lead to a long-term saving of 100–110 Gt CO_2. (Saher, Stephen, and Ahmad, 2021; Markvart and Castañer, 2003) show that different land covers, can have significantly different albedos. Yu and Lu (2014) found that the albedo of different materials was significant in the life-cycle assessment of pavements. The albedo alone caused a 19.1% increase in CO_2e from asphalt against a 9.2% decrease in CO_2e from concrete.

The overall environmental and social impacts of paving over soil are not limited to the net GHG emissions resulting from paving. For instance, the presence and capacity of unsealed soil soaks up water during a flood, reducing flood risk. The expansion of urban settlements of all sizes into flood zones is highlighted by Rentschler et al. (2023). This susceptibility to flooding should also be considered by policymakers when expanding urban areas. Further, in making way for land to be sealed, it is inevitable that animal habitats will be destroyed. This too, disrupts environmental systems but could be overlooked if we focus exclusively on urbanization from the perspective of GHG emissions. Finally, the well-being of humans could be at odds with the expansion of sealed areas. Policymakers should try to conceptualise this opportunity cost in their plans - will the net utility from urban environments exceed that of the natural environments that preceded them?

5.2 Future research

The results for CO_2 in this dissertation provide a foundation that could be built upon by integrating existing soil emission databases (Jian et al., 2021; Bond-Lamberty, Christianson, et al., 2020) so that the fluxes we assign can be better calibrated to the region of interest. In our case this could be done, for instance, by weighting CO_2 fluxes by their distance from the UK. The bound of typical, temperate CO_2 fluxes that we used (Oertel et al., 2016) is likely wider than such a bound for the UK. Accordingly, we anticipate that assigning calibrated fluxes would produce results that fall within the confidence intervals we generated. However, they would allow for the generation of stronger point estimates and narrower confidence intervals.

If databases like COSORE and SRDB grow, using more sophisticated models to predict the flux at different locations might become more reasonable. Such models would allow new variables to be added into consideration more easily than our approach, and give researchers in the area of soil GHG fluxes generally a much wider selection of options to draw from. Furthermore, as databases composed of site data grow, researchers should keep in mind the potential for the oversampling of extreme results. Sites are sometimes chosen to measure some extreme pattern in an ecosystem that environmental scientists view with concern . Accordingly, measurements may have a tendency to be more extreme and it might be advisable to adjust measurements like this in future to smooth out the extreme measurements. In addition, calibrating GHG flux data to only a small number of UK sites might represent the UK as a whole worse than using a larger number of sites including those from the UK. All of these challenges with using site data could be alleviated by new techniques for measuring soil GHG fluxes, that can be taken at a larger scale. Already, there have been related efforts to this from Adachi et al. (2017). The establishment of additional sites at land covers and locations that are currently lacking in databases would be extremely helpful. For global studies, this would be particularly helpful filling in the gap in data in the southern hemisphere. Sites at new land covers in a range of different climates would enable future studies to

consider a more intricate combination of land covers than current research.

When land covers overlap within a $1km^2$ raster cell, the sealed area in these cells is likely covering what would be either grassland or cropland. This is simply because these are the easier land covers to pave over. However, we assigned regions of wetland in these cells the same imperviousness value, likely introducing a degree of error. However, since the majority of the land covers are not isolated but exist in significant clumps we anticipate that this error will not substantially undermine our results. This limitation could be overcome by using a computer capable of dealing with higher resolution imperviousness data, though, to make best use of this, a land cover map of similarly high resolution would be required. Furthermore, our imperviousness data does not pickup narrow ($< 10m$) patches of sealed soil in otherwise natural areas - like minor roads or pavements. We adjusted our results for these roads but not pavements. Whilst we do not anticipate that such features would dramatically alter results, higher resolution imperviousness data would capture these, thus improving the accuracy of derivative research.

When higher resolution data becomes available, the ability to utilise such a large dataset will be more important. Already, the 10m resolution data we used was too fine for large-scale computations. For the effect of such data to be maximised, an equivalently high resolution land cover map is also required. Assigning site-scale GHG flux data to higher resolution maps could have adverse effects on estimates. Calibrating GHG flux data from larger databases could partially alleviate this. Hashimoto (2012) found that the mismatch in scale between their coarse resolution global simulation based on site-scale data did partly contribute to lower variation in simulated values than the observed data. Whether such a result would carry through to our UK-only results with a much less coarse resolution is something worth considering in future research. Oertel et al. (2016) warns that up-scaling site measurements can lead to improbable results. Hence, focusing on the range of possible results is important. Care will also be required when comparing these new results with historical results. Specifically, we will likely see the impact of

sealing increase because smaller patches of impervious land will be captured at higher resolutions.

In each of our simulations the fluxes assigned to each grid cell of the different land covers was the same. In reality, we expect fluxes in the middle of region of, for instance, a forest to correspond to the site measurements from forests used to generate the simulated fluxes. In contrast, the fluxes at the boundary between a forest and a field of crops could be something in between the site measurements from forestland and cropland. We expect that aggregated over the UK, this spatial dependency could be averaged out and produce estimates not too different from ours. Assigning the same GHG flux to all, for example, grassland in each simulation will results in particularly high variance. Therefore, our confidence intervals will be wider than if assigned each squared kilometer in the UK a newly simulated flux. However, as Hashimoto (2012) found, there is a tendency to underestimate the variance of simulated fluxes. We want to avoid this as much as possible because the risk of underestimating the width of confidence intervals is higher in this research area. Accounting for extreme results gives us more confidence that the overall picture we present is the actual effect of sealing in reality.

Our findings also make way for further studies into the additional factors influencing GHG emissions that we outlined earlier. In particular Oertel et al. (2016) consider it particularly important to incorporate soil moisture. However, there is a risk associated with using too many climate-based factors like temperature and moisture because of potentially high levels of correlation between them. In particular, variables like these could be describing some hidden variable. A modelling approach could provide a route around this problem through the use of an interaction term. Regardless of approach, drawing on the expertise of existing researchers and scientists, can help discern these kinds of relationships and is vital for future research in this space.

Chapter 6

Conclusion

We have now discussed the limitations of our findings alongside the potential paths for further research that they open. We also underlined different ways that the results of research in this area could change over time. We found that, when considering only soil GHG fluxes, sealed soil in the UK causes a substantial saving in GHG emissions. The vast majority of this saving comes from CO_2, specifically in grassland and cropland. Seasonally, the greatest overall savings come from the winter and summer months. Efforts to mitigate the impact of paving on climate change must incorporate these results alongside the additional side effects of paving.

Appendix A

Appendices

Project Specification

Project title

Pave the World: The soil ecosystem under threat by expanding cities and towns.

Project background

Soil is full of microorganisms that both store and omit greenhouse gases (GHGs). When urban areas expand and soil is paved over, how does this impact the soil atmosphere? The project serves to uncover what this impact is.

Project aims

The goal is to evaluate how much soil sealing alters regional soil-atmosphere GHG fluxes by assigning a range of gas flux rates to land cover / use maps (UK, Europe, global) with and without soil sealing and establish the difference through a sensitivity analysis.

Project data

Source of data include:

- Data on soil moisture, texture, pH, UK: https://mapapps2.bgs.ac.uk/ukso/home.html
- Imperviousness data for EU (this includes the UK): https://land.copernicus.eu/en/products/resolution-layer-imperviousness
- Land cover data for Europe: https://land.copernicus.eu/en/products/corine-land-cover.
- Satellite derived land cover maps - UK land cover from 1990: https://www.ceh.ac.uk/data/uk land-cover-maps

Project deliverables

- A literature review of soil net GHG fluxes, with a focus on the interrelationship between physico-chemical soil parameters (pH, soil moisture, organic material) and fluxes.
- The linking of existing UKCEH held and other datasets on soil moisture, pH, organic matter, structure, and texture with soil function to develop a model for expected soil function over the UK, European and global landscape.
- A careful interpretation and analysis of time series of a variety of existing data sources to establish how much land has been sealed over time

Also, alongside the dissertation a GitHub repository that allows for calculating the emissions of CO2, CH4 and NO2 in the UK (or perhaps a wider area), with and without paved over land. This includes allowing, for example, the GHG flux rates used to be easily altered. Making the results reproducible is of high interest.

Candidate Techniques

(which techniques you might use to tackle the problem) Initially I need to obtain and pre-process the data to ensure the problem is presented in a credible way. This includes:

- Spatial projection
- Resampling/aggregation
- Cubic
- Bilinear
- Nearest neighbour
- Combining land cover maps to create one which can represent the different gas flux rates across land covers

Plan overview

To start:

- Get to know the literature
- Focus on just the UK – and therefore GHG flux data relating to a temperate climate.
- Use simple forms of the data first
- Think of soil generally – don't look too far into classes of soil just yet.
- Find key numbers relating to GHG flux rates Use greenhouse gas flux rates from existing literature to calculate assign yearly fluxes in CO2, CH4 and NO2 for land covers including:
  - Grassland

- Pastures
- Natural grassland
- Wetlands
- Marshes
- Peat bogs
- Bare soil
- Forests
- Coniferous
- Broad-leaved
- Landfill sites
- Cropland (some crossover with grassland here)
- Urban - assume these sealed areas have 0 GHG flux.

Once we have these GHG fluxes assigned to a composite map (based on both the CEH 2018 LCM and CORINE 2018 Land Cover map) we can begin to consider the relative significance of paving over soil on the atmosphere compared to other sources. Further, we can see how this significance changes as we alter the GHG fluxes (i.e. within some 95After this first round of measurements, we can try to adjust them further by now considering the properties of the soil in different locations, seasons, ect.

Larger challenges include:

- Considering Europe – and perhaps the world
- The level of complexity of the sensitivity analysis

Timeline

These are the tasks we planned to complete on the week beginning with the date listed.

10th June

- Create plan
- Begin background reading for GHG fluxes
- Organise project directory
- Download imperviousness data
- Read imperviousness data papers

17th June

- Begin background reading for land cover maps
- Match land uses to GHG flux levels
- Pre-process data
- Visualise imperviousness data against land cover for a few different examples

24th June

- Create GHG flux conversion tables for classes
- Look into methods used in preprocessing more deeply:
  - Desampling
  - Spatial projections
  - Interpolation
- How to incorporate uncertainty
- Start writing about the methodology for the above.

1st July

- Assign GHG fluxes to a UK land cover map

8th July

- Get a first draft of methodology section done
- Begin writing up background reading in literature review section

15th July

- Continue with methodology section and start to write results section.
- Consider results under different maps: e.g., assigning broad ranges or more specifically assigned values.

22nd July

- Write discussion/conclusion section

29th July

- Finish up sections
- Look over landcover matches again

5th August

- Introduction
- Abstract

12th August

- Finishing up first draft/editing

19th August

- Continue editing

26th August

- Proof read and edit

References

Abdalla, M. et al. (2009). "Nitrous oxide fluxes and denitrification sensitivity to temperature in Irish pasture soils". In: *Soil Use and Management* 25.4, pp. 376–388. DOI: https://doi.org/10.1111/j.1475-2743.2009.00237.x. eprint: https://bsssjournals.onlinelibrary.wiley.com/doi/pdf/10.1111/j.1475-2743.2009.00237.x. URL: https://bsssjournals.onlinelibrary.wiley.com/doi/abs/10.1111/j.1475-2743.2009.00237.x.

Adachi, Minaco et al. (2017). "Estimation of global soil respiration by accounting for land-use changes derived from remote sensing data". In: *Journal of Environmental Management* 200, pp. 97–104. ISSN: 0301-4797. DOI: https://doi.org/10.1016/j.jenvman.2017.05.076. URL: https://www.sciencedirect.com/science/article/pii/S0301479717305613.

Agency, European Environment (n.d.). *Administrative boundaries of EEA39 (NUTS 2016, aligned with Corine Land Cover 2012), Apr. 2018*. URL: https://sdi.eea.europa.eu/catalogue/srv/api/records/8526ff78-b000-42e1-8360-a2fb3a51e4ac.

Akbari, Hashem, H Damon Matthews, and Donny Seto (2012). "The long-term effect of increasing the albedo of urban areas". In: *Environmental Research Letters* 7.2, p. 024004.

Bernardo, Guillermo de et al. (2013). "Compact Querieable Representations of Raster Data". In: *String Processing and Information Retrieval*. Ed. by Oren Kurland, Moshe Lewenstein, and Ely Porat. Cham: Springer International Publishing, pp. 96–108. ISBN: 978-3-319-02432-5.

Bond-Lamberty, Ben, Danielle S. Christianson, et al. (2020). "COSORE: A community database for continuous soil respiration and other soil-atmosphere flux data". In: *Global Change Biology* submitted.

Bond-Lamberty, Ben and Allison Thomson (2010). "Temperature-associated increases in the global soil respiration record". In: *Nature* 464.7288, pp. 579–582.

Bontemps, S. et al. (Dec. 2013). "Consistent Global Land Cover Maps For Climate Modelling Communities: Current Achievements Of The ESA' Land Cover CCI". In: *ESA Living Planet Symposium*. Vol. 722. ESA Special Publication, p. 62.

Brander, Matthew and Gary Davis (2012). "Greenhouse gases, CO2, CO2e, and carbon: What do all these terms mean". In: *Econometrica, White Papers*.

Building Services Engineers, Chartered Institution of (2022). *Research insight : urban albedo: developing a canyon albedo calculator*. eng. London: The Chartered Institution of Building Services Engineers CIBSE. ISBN: 1-5231-5013-0.

Büttner, György et al. (2021). *Copernicus Land Monitoring Service-CORINE Land Cover. User Manual*. Tech. rep. Copernicus Publications.

Congalton, Russell G (2005). "Thematic and positional accuracy assessment of digital remotely sensed data". In: *Proceedings of the seventh annual forest inventory and analysis symposium*. Vol. 1. US Department of Agriculture General Technical Report WO-77, Washington, DC, p. 51.

Crippa, M. et al. (2020). *Fossil CO2 emissions of all world countries - 2020 Report, EUR 30358 EN, Publications Office of the European Union, Luxembourg, 2020, , JRC121460*. DOI: 10.2760/143674.

Dalal, Ram C and Diane E Allen (2008). "Greenhouse gas fluxes from natural ecosystems". In: *Australian Journal of Botany* 56.5, pp. 369–407.

Dijkstra, Feike A et al. (2012). "Effects of elevated carbon dioxide and increased temperature on methane and nitrous oxide fluxes: evidence from field experiments". In: *Frontiers in Ecology and the Environment* 10.10, pp. 520–527.

European Environment Agency (2020a). *CORINE Land Cover 2018 (raster 100 m), Europe, 6-yearly - version 2020_20u1, May 2020*. DOI: 10.2909/960998c1-1870-

4e82-8051-6485205ebbac. URL: https://doi.org/10.2909/960998c1-1870-4e82-8051-6485205ebbac.

European Environment Agency (2020b). *Imperviousness Density 2018 (raster 10 m), Europe, 3-yearly, Aug. 2020*. DOI: 10.2909/3bf542bd-eebd-4d73-b53c-a0243f2ed862. URL: https://doi.org/10.2909/3bf542bd-eebd-4d73-b53c-a0243f2ed862.

Fick, Stephen E and Robert J Hijmans (2017). "WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas". In: *International journal of climatology* 37.12, pp. 4302–4315.

Han, Dianyuan (2013). "Comparison of commonly used image interpolation methods". In: *Conference of the 2nd International Conference on Computer Science and Electronics Engineering (ICCSEE 2013)*. Atlantis Press, pp. 1556–1559.

Harris, Ian et al. (2020). "Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset". In: *Scientific data* 7.1, p. 109.

Hashimoto, Shoji (Aug. 2012). "A New Estimation of Global Soil Greenhouse Gas Fluxes Using a Simple Data-Oriented Model". In: *PLOS ONE* 7.8, pp. 1–8. DOI: 10.1371/journal.pone.0041962. URL: https://doi.org/10.1371/journal.pone.0041962.

Hijmans, Robert J. (2024). *terra: Spatial Data Analysis*. R package version 1.7-78. URL: https://CRAN.R-project.org/package=terra.

Hijmans, Robert J. et al. (2024). *geodata: Download Geographic Data*. R package version 0.6-2. URL: https://CRAN.R-project.org/package=geodata.

Hinze, Jonas, Axel Albrecht, and Hans-Gerhard Michiels (2023). "Climate-Adapted Potential Vegetation—A European Multiclass Model Estimating the Future Potential of Natural Vegetation". In: *Forests* 14.2. ISSN: 1999-4907. DOI: 10.3390/f14020239. URL: https://www.mdpi.com/1999-4907/14/2/239.

IPCC (2021). *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel*

on Climate Change. Vol. In Press. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press. DOI: 10.1017/9781009157896.

Jian, J. et al. (2021). *A Global Database of Soil Respiration Data, Version 5.0*. en. DOI: 10.3334/ORNLDAAC/1827. URL: https://daac.ornl.gov/cgi-bin/dsviewer.pl?ds_id=1827.

Kaya, Yoichi (1989). "Impact of carbon dioxide emission control on GNP growth: interpretation of proposed scenarios". In: *Intergovernmental panel on climate change/response strategies working group, May*.

Keys, R. (1981). "Cubic convolution interpolation for digital image processing". In: *IEEE Transactions on Acoustics, Speech, and Signal Processing* 29.6, pp. 1153–1160. DOI: 10.1109/TASSP.1981.1163711.

Kumari, Aradhna et al. (2023). "Soil microbes: a natural solution for mitigating the impact of climate change". In: *Environmental Monitoring and Assessment* 195.12. Cited by: 3. DOI: 10.1007/s10661-023-11988-y. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85176043846&doi=10.1007%2fs10661-023-11988-y&partnerID=40&md5=0fd492d966672fcbd6d1e0a4c13c06e3.

Maier, Martin et al. (2022). "Introduction of a guideline for measurements of greenhouse gas fluxes from soils using non-steady-state chambers". In: *Journal of Plant Nutrition and Soil Science* 185.4, pp. 447–461.

Mark Padgham et al. (June 2017). "osmdata". In: *Journal of Open Source Software* 2.14, p. 305. DOI: 10.21105/joss.00305. URL: https://joss.theoj.org/papers/10.21105/joss.00305.

Markvart, Tom and Luis Castañer (2003). *Practical handbook of photovoltaics: fundamentals and applications*. Elsevier.

Meyer, N., G. Welp, and W. Amelung (2018). "The Temperature Sensitivity (Q10) of Soil Respiration: Controlling Factors and Spatial Prediction at Regional Scale Based on Environmental Soil Classes". In: *Global Biogeochemical Cycles* 32.2, pp. 306–323. DOI: https://doi.org/10.1002/2017GB005644. eprint: https:

//agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2017GB005644. URL: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2017GB005644.

Morton, R.D. et al. (2022). *Land Cover Map 2018 (1km summary rasters, GB and N. Ireland)*. DOI: 10.5285/9b68ee52-8a95-41eb-8ef1-8d29e2570b00. URL: https://doi.org/10.5285/9b68ee52-8a95-41eb-8ef1-8d29e2570b00.

Mühleisen, Hannes and Mark Raasveldt (2024). *duckdb: DBI Package for the DuckDB Database Management System*. R package version 1.0.0-2. URL: https://CRAN.R-project.org/package=duckdb.

Mundim, Kleber C. et al. (2020). "Temperature coefficient (Q10) and its applications in biological systems: Beyond the Arrhenius theory". In: *Ecological Modelling* 431, p. 109127. ISSN: 0304-3800. DOI: https://doi.org/10.1016/j.ecolmodel.2020.109127. URL: https://www.sciencedirect.com/science/article/pii/S030438002030199X.

Oertel, Cornelius et al. (2016). "Greenhouse gas emissions from soils—A review". In: *Geochemistry* 76.3, pp. 327–352. ISSN: 0009-2819. DOI: https://doi.org/10.1016/j.chemer.2016.04.002. URL: https://www.sciencedirect.com/science/article/pii/S0009281916300551.

Ouyang, Zutao et al. (2022). "Albedo changes caused by future urbanization contribute to global warming". In: *Nature communications* 13.1, p. 3800.

Parker, J. Anthony, Robert V. Kenyon, and Donald E. Troxel (1983). "Comparison of Interpolating Methods for Image Resampling". In: *IEEE Transactions on Medical Imaging* 2.1, pp. 31–39. DOI: 10.1109/TMI.1983.4307610.

Pebesma, Edzer (2018). "Simple Features for R: Standardized Support for Spatial Vector Data". In: *The R Journal* 10.1, pp. 439–446. DOI: 10.32614/RJ-2018-009. URL: https://doi.org/10.32614/RJ-2018-009.

Qiu, Linghua et al. (2024). "Substantial terrestrial carbon emissions from global expansion of impervious surface area". In: *Nature Communications* 15.1, p. 6456.

R Core Team (2024). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: https://www.R-project.org/.

Rankin, Tracy et al. (2023). "Partitioning autotrophic and heterotrophic respiration in an ombrotrophic bog". In: *Frontiers in Earth Science* 11, p. 1263418.

Rentschler, Jun et al. (2023). "Global evidence of rapid urban growth in flood zones since 1985". In: *Nature* 622.7981, pp. 87–92.

Saher, Rubab, Haroon Stephen, and Sajjad Ahmad (2021). "Effect of land use change on summertime surface temperature, albedo, and evapotranspiration in Las Vegas Valley". In: *Urban Climate* 39, p. 100966. ISSN: 2212-0955. DOI: https://doi.org/10.1016/j.uclim.2021.100966. URL: https://www.sciencedirect.com/science/article/pii/S2212095521001966.

Story, Michael and Russell G Congalton (1986). "Accuracy assessment: a user's perspective". In: *Photogrammetric Engineering and remote sensing* 52.3, pp. 397–399.

UK Department for Transport (2022). *Road lengths in Great Britain: 2021*. https://www.gov.uk/government/statistics/road-lengths-in-great-britain-2021/road-lengths-in-great-britain-2021. [Accessed 12-08-2024].

UK Department for Transport and Ministry of Housing, Communities & Local Government (Mar. 2007). *Manual for Streets*. Tech. rep. UK Government. URL: https://www.gov.uk/government/publications/manual-for-streets.

Venables, W. N. and B. D. Ripley (2002). *Modern Applied Statistics with S*. Fourth. ISBN 0-387-95457-0. New York: Springer. URL: https://www.stats.ox.ac.uk/pub/MASS4/.

Wickham, Hadley (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. ISBN: 978-3-319-24277-4. URL: https://ggplot2.tidyverse.org.

Xi, Fengming et al. (2016). "Substantial global carbon uptake by cement carbonation". In: *Nature Geoscience* 9.12, pp. 880–883. DOI: 10.1038/ngeo2840. URL: https://doi.org/10.1038/ngeo2840.

Yan, Weiming et al. (2022). "Response of soil greenhouse gas fluxes to warming: A global meta-analysis of field studies". In: *Geoderma* 419, p. 115865. ISSN: 0016-7061. DOI: https://doi.org/10.1016/j.geoderma.2022.115865. URL: https://www.sciencedirect.com/science/article/pii/S0016706122001720.

Young, Ben et al. (2021). "LCIA Formatter". In: *Journal of Open Source Software* 6.66, p. 3392. DOI: 10.21105/joss.03392. URL: https://doi.org/10.21105/joss.03392.

Yu, Bin and Qing Lu (2014). "Estimation of albedo effect in pavement life cycle assessment". In: *Journal of Cleaner Production* 64, pp. 306–309. ISSN: 0959-6526. DOI: https://doi.org/10.1016/j.jclepro.2013.07.034. URL: https://www.sciencedirect.com/science/article/pii/S0959652613004952.