Statistical and machine learning modelling of UK surface ozone

Lily Gouldsbrough, BSc

Lancaster Environment Centre

Lancaster University

A thesis submitted for the degree of

Doctor of Philosophy

October, 2023

Dedicated to my beloved husband and best friend, Nic. I am deeply grateful for your never-ending support, love and encouragement.

Statistical and machine learning modelling of UK surface ozone

Lily Gouldsbrough, BSc.

Lancaster Environment Centre, Lancaster University

A thesis submitted for the degree of *Doctor of Philosophy*. October, 2023.

Abstract

In addition to atmospheric observations, numerical models are crucial to understand the impacts of human activities on the environment, from attributing poor air quality to assessing climate change impacts. While process-based models, such as chemistry transport models (CTMs), are widely used, recent data science advances enable greater use of statistical and machine learning methods as alternatives to describe and predict atmospheric composition. State-of-the-art data science methods can be faster to run than CTMs and used at high temporal and spatial resolutions due to codebase efficiencies.

This thesis focuses on modelling UK surface ozone and its drivers (high levels of which are detrimental to human and plant health) through the development and novel application of sophisticated statistical and machine learning techniques. Motivated by possible adverse effect of climate change on ozone concentrations, a temperature-dependent Extreme Value Analysis is used to explore the probability, magnitude, and frequency of extreme ozone events over recent decades. For 2010–2019, it is found that the 1-year return level of daily maximum 8-h mean (MDA8) ozone exceeds the 'moderate' health threshold ($100\,\mu g/m^3$) at >90% of sites, but that the probability of extreme ozone events has markedly decreased since the 1980s.

A machine learning methodology to downscale and bias correct a CTM (EMEP4UK) ozone surface was developed and evaluated. Compared to the unadjusted CTM, the downscaled surface exhibits a lower bias in reproducing MDA8 ozone allowing more robust assessments of important policy metrics. Analysis of

the downscaled product (2014–2018) reveals on average 27% of the UK fails the government long-term objective for MDA8 ozone to not exceed 100 µg/m^3 more than 10 times per year, compared to 99% in the unadjusted CTM. A classification-based machine learning analysis into high-level ozone drivers was also performed and shows a robust relationship between ozone and temperature. The method is demonstrated to offer remarkable promise as a tool with which to forecast the presence of high-level ozone. Despite a UK focus, the data-driven methods developed and applied here are applicable to modelling ozone in other regions of the world where measurements exist

Acknowledgements

I would like to express my profound gratitude to my wonderful supervisors Emma Eastoe, Ryan Hossaini and Paul Young for their invaluable guidance and support during the course of my PhD. Their expertise, patience, and mentorship have played a pivotal role in the completion of this thesis.

I extend my thanks to Lancaster Environment Centre for their continuous support and for creating a welcoming research environment. Their commitment to a culture of academic excellence has played a significant role in my growth as a researcher.

Finally, I want to acknowledge the support and encouragement from my family and friends, and in particular, my parents. Your unwavering belief in my abilities, especially during moments of self-doubt, has led me to here. I love you both dearly.

Declaration

I declare that the work presented in this thesis 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 thesis does not exceed the maximum permitted word length of 80,000 words including appendices and footnotes, but excluding the bibliography. A rough estimate of the word count is: 40,000.

Lily Gouldsbrough

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Atmospheric science has significantly advanced our understanding of the world by providing insights into both the natural composition of the Earth's atmosphere and changes in composition brought about by human activities. Atmospheric models play a key part in evolving our understanding of physical and chemical processes, including assessing the impact of climate change on the atmosphere. They are also an essential operational tool for meteorological and air quality forecasting. Whilst there are many different branches of atmospheric modelling, the focus of this thesis is on models which help to explain the chemical properties and composition of the atmosphere. Early atmospheric chemistry transport models (CTMs) were two dimensional and utilised only the most fundamental equations to explain atmospheric behaviour. However, advances in computing capabilities and our comprehension of the complex physical processes that govern the Earth's atmosphere have led to three dimensional models of increasing complexity. Today's CTMs are highly sophisticated, and can simulate a range of atmospheric phenomena, including atmospheric chemistry from the surface to the mesosphere, large-scale weather patterns, and storms, which has made them invaluable for weather forecasting, climate research and environmental monitoring.

Whilst CTMs have been widely used for decades and will continue to play a vital part in atmospheric science, recent advances in statistical and machine

learning methods combined with the availability of large quantities of atmospheric observational data has led to the increased use of data science methods in atmospheric science. These data-driven methods, which are often far more flexible and faster to train than atmospheric CTMs (i.e., cost of model runs and code development), provide a complementary way in which to study the changing behaviour of the atmosphere. In recent years, advances in technology have led to an unprecedented amount of atmospheric data being collected, including remote observations (e.g. satellite), in situ observations (e.g. weather stations), and climate model output. This data provides an enormous amount of information about the state and behaviour of the Earth's atmosphere. However, the sheer volume of data poses a significant challenge for traditional data analysis techniques. Modern data science methods offer powerful tools for handling and analysing large and complex datasets, providing the ability to identify patterns that may not be visible through traditional modelling approaches. Additionally, these models support the fusion of a wide range of variables and sources of data, combining information and providing a more comprehensive understanding of atmospheric behaviour.

Air pollution is a critical aspect of atmospheric science, and its study is of great importance for both environmental and public health reasons. Acute and chronic exposure to air pollution has been linked to a wide range of health issues, including respiratory problems, heart disease, dementia, and contributes to premature death (Anderson et al., 2004; Díaz et al., 2018; Nuvolone et al., 2018). Additionally, air pollution can harm plant and animal life, damage ecosystems, and contribute to climate change (i.e., as key air pollutants, such as ozone and aerosol, are also climate forcers) (Sandermann Jr, 1996; Ainsworth et al., 2012; Karmakar et al., 2022). Therefore, understanding and mitigating the effects of air pollution is crucial for the well-being of both current and future generations.

This thesis presents a collection of data-driven models for surface level ozone in the UK, developed using data science methodology. The adoption of data science methods in air pollution modelling has increased in recent years, from

operational forecasting of pollutant concentrations (Kleinert et al., 2021) to creating high-resolution bias-corrected surfaces of various air pollutants for understanding health impacts (Gariazzo et al., 2020; Silibello et al., 2021; Bertrand et al., 2022). Despite the increasing use of data science methods in atmospheric modelling, there are still important challenges to be addressed. These include issues related to model complexity, modelling and predicting rare events, and the need to develop interpretable models that can be used for decision-making and to help understand atmospheric processes. This thesis explores the application and development of data science methods in modelling of surface level ozone, with a particular focus on elevated concentrations.

This introduction presents a discussion of surface level ozone, the impacts of high-level ozone episodes, and the state of surface level ozone in the UK, followed by a summary of the existing methods to model surface level ozone. A brief overview of the thesis contributions is presented along with an outline of the remaining thesis structure. The subsequent chapters are each self-contained and include separate introductions and literature reviews that expand upon the following literature review.

1.1 Surface level ozone

1.1.1 Overview

Surface level ozone is formed through a complex series of chemical reactions. The concentration of ozone at any given time or place is controlled by a variety of factors, including emissions of ozone precursors and meteorological conditions. Key precursors are nitrogen oxides (NO_X) and volatile organic compounds (VOCs) (Tan et al., 2018; Lu et al., 2019; Zhao et al., 2022). Sources of NO_X include emissions from vehicles (Ogur and Kariuki, 2014), industrial processes (Olivier et al., 1998), and natural sources such as lightening (Levine et al., 1984), whilst VOCs are commonly found in various products such as paints, solvents, cleaning agents,

and fuels, as well as being emitted from transportation and industrial processes (Kim et al., 2001; H. Wang et al., 2013; Halios et al., 2022). Both precursors and ozone itself can be transported over long distances by atmospheric circulation patterns (Q. Li et al., 2002), with concentrations in the UK largely influenced by hemispheric- and regional- scale effects (Jenkin, 2008), consequently ozone concentrations are related not just to local point sources of precursor pollutants but also atmospheric transport and synoptic scale weather. Other factors that influence ozone formation and dispersion include meteorological conditions, such as temperature, humidity, and wind speed, and the availability of sunlight, which is necessary for the photochemical reactions that produce ozone (Otero et al., 2016; Carro-Calvo et al., 2017; Noelia Otero, Sillmann, et al., 2018). In addition to dispersion in the atmosphere, ozone concentrations can also be depleted through dry deposition onto vegetation and buildings and other surfaces (Clifton et al., 2020). The interactions between these factors can result in complex and variable patterns of ozone concentrations, both spatially and temporally.

1.1.2 Impacts of high-level ozone

Concerns over high levels of surface level ozone arise from the range of negative impacts that such levels have on human health, vegetation, and ecosystems. One of the most significant impacts is its effect on human health. Ozone is an irritant to the respiratory system, and exposure to high concentrations can cause coughing, throat irritation, and shortness of breath. In individuals with pre-existing respiratory conditions such as asthma or chronic obstructive pulmonary disease, exposure to high levels of ozone can exacerbate symptoms and lead to hospitalisation or mortality (Nuvolone et al., 2018; J. Zhang et al., 2019).

High-level surface ozone can also have negative effects on vegetation. Ozone exposure can lead to a decrease in photosynthesis and plant growth, as well as an increase in leaf damage and premature leaf drop (Sandermann Jr, 1996; Karmakar et al., 2022). This can have significant impacts on crops, leading to decreased yields

and lower quality produce (Heagle, 1989; Avnery et al., 2011; Ghude et al., 2014). Reduced crop yields have a severe economic impact, with ozone flux-based estimates of wheat yield losses to be 4.56 billion Euro in Europe (specifically, regions in the European Monitoring and Evaluation Programme), equating to a mean yield loss of 13%. The greatest economic losses are found in important wheat growing areas in western and central Europe (Harmens et al., 2015).

Lastly while surface level ozone is most detrimental to human health and vegetation, it is also detrimental to climate change. Trees and other plants play a key role in absorbing carbon dioxide and mitigating the impacts of climate change, so damage to vegetation from high levels of ozone has the potential to exacerbate the effects of climate change by affecting carbon uptake (Ainsworth et al., 2012). Furthermore, ozone is a greenhouse gas in the upper troposhere, as it absorbs and emits infrared radiation in the atmosphere (Mohnen et al., 1993), and its tropospheric burden has increased by ∼44% between the year 1850 (pre-industrial) and 2005–2014 (Griffiths et al., 2021). Although the impact of ozone, a short-lived climate forcer, on the pre-industrial to present warming is relatively small compared to other well-mixed greenhouse gases such as methane and CO_2, reducing its concentrations could have a beneficial effect on the overall climate system.

1.1.3 State of UK ozone

1.1.3.1 Monitoring networks

The UK's first Clean Air Act was introduced in 1956 following the great London smog in 1952 which resulted in an estimated 3000–12,000 deaths (Bell et al., 2004). The National Survey, the world's first co-ordinated national air pollution monitoring network, was established in 1961 to monitor black smoke and sulphur dioxide at around 1200 sites in the UK. Subsequently, several pieces of legislation and additional monitoring networks were introduced to combat and measure air quality (Defra, 2023). A brief overview of these is now given.

Since the 1960s, the focus has been on monitoring pollutants generated from

vehicular emissions, including ozone, nitrogen dioxide, and fine particulate matter. The Enhanced Urban Network was established in 1992, consolidating all statutory and other urban monitoring into one comprehensive program. Over the next five years, more than 50 local authority sites, including 14 of the London Air Quality Monitoring Network sites, were integrated into one network. In 1998, the UK urban and rural automatic networks were combined to form the Automatic Urban and Rural Network (AURN), which, as of the end of 2021, comprises over 170 sites across the UK (Defra, 2023). The Scottish Air Quality Network (see https://scottishairquality.scot), run by the Scottish Environment Protection Agency, and the Welsh Air Quality Network (see https://airquality.gov.wales/), run by the Welsh government, are mostly incorporated into the AURN network. Together, these networks provide a comprehensive picture of air quality across the UK, allowing policymakers to identify areas of concern and target interventions to improve air quality.

The London Air Quality Network (LAQN, see https://londonair.org.uk/) was formed in 1993 and is a collaborative project between King's College London and the Greater London Authority. The aim of the network is to monitor and report air quality across the Greater London region. The LAQN is the largest urban monitoring network in the UK, and is composed of over 100 sites across the city which collect data on a range of air pollutants including nitrogen dioxide, particulate matter, and ozone. The data collected by the network is used to inform air quality policy, assess compliance with legal air quality limits, and provide information to the public about the levels of air pollution in their local area. The LAQN has played a vital role in improving air quality in London, helping to inform the introduction of measures such as the Ultra-Low Emission Zone (ULEZ) and the Low Emission Bus Zone.

1.1.3.2 Trends over time and spatial distribution

Peak concentrations of surface ozone in the UK have decreased since the 1990s (Jenkin, 2008; Diaz et al., 2020). This decrease can be attributed to a combination of factors, including the implementation of stricter air quality regulations, reductions in emissions from industrial processes and transportation, and changes in weather patterns (Jenkin, 2008). However, despite this overall decreasing trend, surface ozone levels in the UK still exceed national regulatory limits; for example, the UK's regulatory limit for 8-hr average daily maximum MDA8 ozone to not exceed $100\,\mu g/m^3$ is broken more than 10 times per year and in 2018, 84% of AURN monitoring stations broke this objective (Diaz et al., 2020).

Ozone concentrations exhibit a strong seasonal cycle, with levels generally peaking during the spring and summer in the northern mid-latitudes when temperatures and solar radiation are higher, conditions conducive to ozone formation (R. G. Derwent et al., 1998; Paul S Monks, 2000). Concentrations of ozone also vary throughout the day, following a diurnal cycle that typically peaks in the mid-afternoon and is lowest at night. During the daytime, high temperatures and strong solar radiation result in increased surface level ozone production. At night, a lack of sunlight and cooler temperatures lead to reduced production, which, combined with increased deposition and depletion, results in lower concentrations. The subsequent chapters in this thesis do not consider the diurnal cycle of ozone, and only MDA8 ozone data is used due to the noise present in hourly measurement data. Furthermore, the UK's health metrics for ozone are based on MDA8 ozone instead of hourly.

In addition to exhibiting temporal variability, surface ozone concentrations also display spatial heterogeneity across the UK, as shown in 1.1. Concentrations of surface level ozone tend to be highest in rural areas of the UK (AQEG, 2021), with the greatest number of high-level ozone events occurring in the south (Diaz et al., 2020). In contrast, urban and suburban areas typically experience lower concentrations, although local emissions from road transport can significantly

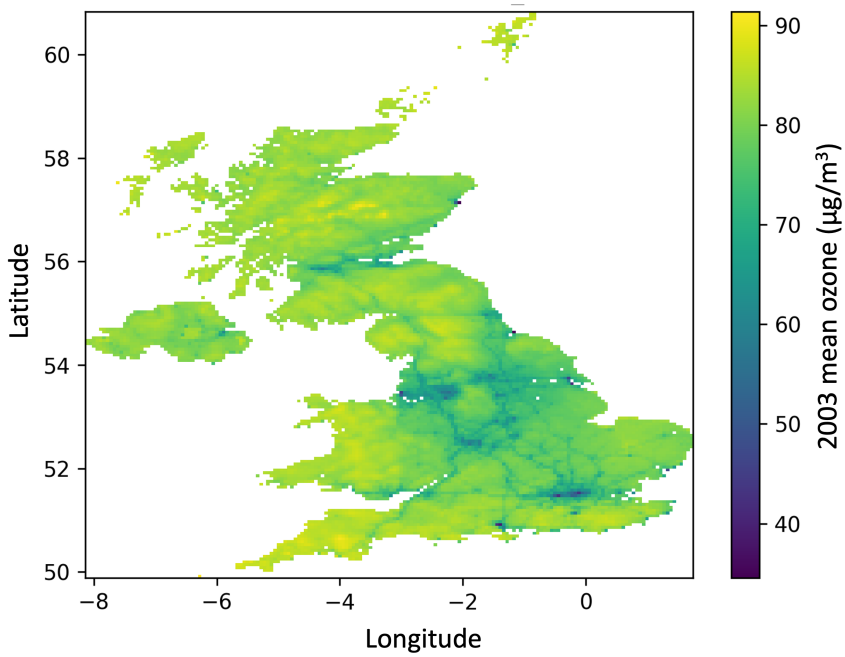

Figure 1.1: Yearly mean ozone for 2003 across the UK from EMEP4UK.

influence these levels. Hemispheric baseline concentrations of ozone play a large role in determining ozone levels in the western and northern regions of the UK, while local emissions from the UK and mainland Europe are important contributors to ozone levels in the eastern and southern parts of the UK (AQEG, 2021). This reflects the complex interaction of local-, regional- and hemispheric-scale factors that influence ozone concentrations, including long-range atmospheric transport patterns, local emissions sources, and meteorological conditions.

Whilst peak ozone levels have been declining in the UK (Jenkin, 2008; Diaz et al., 2020; Finch and Palmer, 2020), the acceleration of climate change leaves no room for complacency. Climate change may already be contributing to increasing concentrations of surface ozone in many regions, including the UK (Orru et al., 2013), due to higher temperatures and longer periods of sunlight resulting in more efficient ozone production. In addition, climate change can also alter the atmospheric circulation patterns that govern the transport of pollutants, potentially increasing the frequency and intensity of ozone episodes in certain regions.

1.2 Challenges of modelling surface level ozone

Although there have been many advances in our understanding of surface level ozone, many challenges remain, some of which form the core research questions for this thesis.

1.2.1 High-level ozone

While the general behaviour of surface level ozone is well understood, predicting occurrence, persistence and magnitude of extreme high-level ozone events can be challenging. These events are rare by definition and are often localised, meaning that there is limited observational data available with which to study them. Most process-based environmental models, including CTMs, are developed to reproduce average behaviour. However, in air quality monitoring, modelling and forecasting it is often the highest levels that are of most importance, as it is these that do the most damage. With their tendency to generalise to mean ozone behaviour, some traditional CTMs are known to underrepresent high-level ozone (e.g. Wilczak et al., 2009; C. Lin et al., 2017; Abdi-Oskouei et al., 2020). Further, many common statistical and machine learning modelling techniques are also designed to generalise to the mean behaviour of a distribution, and therefore fail to capture extremes (Velthoen et al., 2022).

1.2.2 High resolution ozone surfaces

High-resolution spatial surfaces are essential to evaluate population exposure and hence the risk from high-level ozone, not least because of the often localised nature of episodes. Whilst in situ measurement data has good accuracy, measurement sites are sparse and not spread evenly across the UK and it is often unknown how representative individual sites are of a region. CTMs, which generate gridded surfaces of ozone concentrations, provide a useful extension to measurement data. However, when compared to measured data, CTMs can exhibit biases

and often underestimate peak concentrations of high-level ozone events, due to the localized nature of these events and the coarse gridded surfaces produced by the models, along with imperfect model representation of various processes (e.g. emissions). Additionally, CTMs are computationally expensive and often rely on high performance computers to run. The data produced from these models is large, as they are replications of massively complex systems with vast numbers of variables across space and time, making the processing and visualisation of the data non-trivial. This also limits the resolution (important for exposure assessment) at which the models can feasibly be run due to the computational requirements needed. Further, CTM model outputs are not necessarily freely available, and developments to CTMs can be challenging due to new components needing to fit within existing model code and the lead time for their implementation.

1.2.3 Identifying drivers of high-level ozone

As previously mentioned, the processes surrounding the generation of very high-level ozone concentrations are not generally well understood. Identifying the drivers of such episodes is crucial to both improving this understanding and allowing us to forecast future episodes. Concentrations of ozone are influenced by a complex set of interacting and often non-linear relationships between numerous factors, including local and regional weather (e.g. temperature, wind speed, atmospheric circulation patterns), and emissions of NO_X and VOCs. Despite these challenges, studies have attempted to identify the most significant drivers of ozone concentrations, such as synoptic weather patterns that can transport ozone precursors over long distances or trap pollutants in specific regions (Pope et al., 2016), and source regions of transported ozone including Europe and hemispheric background (Romero-Alvarez et al., 2022). These studies have highlighted the need for a comprehensive and integrated approach to understanding the drivers of ozone concentrations, including the use of advanced data-driven modelling techniques and the collection of high-quality data across multiple variables and spatial locations and over a long-term

period.

1.2.4 Forecasting high-level ozone events

Forecasting high-level ozone concentrations is necessary to be able to alert the public of high-level ozone events, and particularly keeping those in high-risk groups safe during ozone episodes. In the UK, operational forecasting models include the Air Quality Unified Model (AQUM) developed by the UK Met Office and using the Unified Model dynamical core (Savage et al., 2013). Accurate local forecasts of pollutants from such models are limited by the spatial resolutions at which the model operates, chemical boundary conditions, and the reliability of the meteorological forecasts. For example, official pollution forecasts in the UK are made at a 12×12 km resolution by the Met Office (Savage et al., 2013). ML models present us with computationally efficient alternatives to produce forecasts based on measurement data, allowing for forecasts to be made at a sub-12×12 km resolution. Rather than forecasting ozone concentrations, we forecast the likelihood of high-level ozone events of most relevance to the public and health services, i.e., health threshold exceedance days (Neal et al., 2014). Further, finer resolution forecasts would be helpful in improving localised warnings of high-level ozone events.

1.3 Data science methods to model surface level ozone

1.3.1 Statistical models

Statistical models provide a stochastic representation of real-world systems. Unlike CTMs, they do not aim to create an exact mathematical replication of the physical system, rather they are used to identify patterns, trends and correlations in data, and can be used to make predictions or estimates based on the data.

Several studies have used statistical methods to model surface level ozone,

with the methods used being of varying complexity. Multiple linear regression (MLR) is a commonly used method to model the relationship between explanatory and response variables. The method involves fitting relationships between the predictors and the mean response, with the assumption that given the predictors, the responses are normally distributed about the mean. MLR has been widely used to model ozone concentrations in various regions, e.g. Malaysia (Ghazali et al., 2010; Ramli et al., 2010; Napi et al., 2020), India (Allu et al., 2020) and Kuwait (Abdul-Wahab et al., 2005). However, comparisons of MLR models with artificial neural networks, a type of machine learning model, have shown that neural networks perform better in modelling ozone concentrations (Bandyopadhyay and Chattopadhyay, 2007; Moustris et al., 2012).

Extreme value analysis (EVA) is a branch of statistics that focuses on the analysis of extreme or rare events. The method is used to model the behaviour of the tail of a probability distribution, and can be used describe the likelihood, frequency and magnitude of an extreme event. In atmospheric chemistry, EVA can be applied to the analysis of air pollution events, such as high levels of surface ozone or particulate matter, and can help to identify the risk of these events over space and time. EVA has been used to model ozone concentrations in limited regions of the UK (Eastoe, 2009; Eastoe and Tawn, 2009), India (Hazarika et al., 2019), California (Wilson et al., 2022), and across the US (Shen et al., 2016), and the approach is gaining prominence as a useful tool with which to characterise exposure risk.

Spatial statistical models are useful to create high-resolution surfaces of ozone, including at locations where there are no measurements. Several statistical methods exist to interpolate between measurement stations, including kriging (Adam et al., 2014), Bayesian inference (Zidek et al., 2000; Reich, Fuentes, et al., 2011) and Gaussian processes (Gelfand and Schliep, 2016). Statistical bias correction methods can also be used to correct CTM output; such output provides a useful alternative to interpolating point measurement data; however, CTM output is often biased compared to measurements due to model uncertainty and coarse spatial resolution.

Possible methods include linear regression (Onwukwe and Jackson, 2021), Bayesian Maximum Entropy (Reyes and Serre, 2014), Kalman filters (Ponomarev et al., 2021), and quantile mapping (Stähle, 2019). Biases due to the coarse resolution of CTMs can alternatively be addressed by downscaling the model outputs. Statistical downscaling methods include fitted empirical orthogonal functions (Alkuwari et al., 2013), principle fitted components (Alkuwari et al., 2013) and a model diagnostic and correction approach (Guillas et al., 2008).

A better understanding of the drivers behind ozone, particularly high-level ozone, is useful in informing future development of both data-driven and numerical models. One particularly statistical method that is useful in identifying these relationships is principal component analysis (PCA). PCA can be used to determine the most relevant features in modelling ozone concentrations, and can be used as a variable selector to reduce the dimensionality of a multivariate dataset (Abdul-Wahab et al., 2005).

Time series modelling is useful in forecasting future concentrations of ozone. Time-series models, such as ARIMA and regression, have previously been used to forecast ozone concentrations (Robeson and Steyn, 1990; K. Kumar et al., 2004; Dueñas et al., 2005; U. Kumar and De Ridder, 2010; Y.-R. Li et al., 2021). CTMs are commonly used to produce operational air quality forecasts (e.g. Savage et al., 2013), however, the forecasts from these models often benefit from statistical post-processing methods that incorporate recent observations to improve the accuracy of the forecasts (e.g. Neal et al., 2014).

1.3.2 Machine learning models

While statistics draws population inferences from a sample, machine learning (ML) finds generalizable predictive patterns (Bzdok et al., 2018). With the emergence of open-source ML software packages and the increasing availability of very large datasets, there has been a growing interest in using ML methods as an alternative to traditional statistical models.

ML methods have been widely used to create spatio-temporal surfaces of ozone by combining data from CTMs and measurements, e.g. extreme gradient boosting has been used to create a spatio-temporal ozone surface for a region in China (Hu et al., 2022), a Bayesian ensemble ML framework has been used to downscale the Community Multiscale Air Quality model (Ren et al., 2022), and a convolutional neural network has been used to downscaled GEOS Composition Forecast (Geiss et al., 2022).

With the help of explainer models which interpret the output of a ML model, ML can identify the drivers of ozone by analysing the interactions between multiple input features. Several explainer models exist, for instance SHapley Additive exPlanations is a model-agnostic method that provides a way to explain the output of any machine learning model by assigning importance scores to each feature in the input data (Lundberg and S.-I. Lee, 2017). Another popular explainer model is Local Interpretable Model-Agnostic Explanations that is used to identify the important features in a model and explain how they contribute to the model's predictions (Ribeiro et al., 2016).

ML has also emerged as a powerful tool for forecasting concentrations of surface level ozone. Methods can be based solely on measurement data or combined with information from CTMs and meteorological forecasts. Random forests and gradient boosted trees (ensembles of decision trees) are popular ML models as they can model non-linear relationships between many input features and have been used to forecast ozone in Australia (Jiang and Riley, 2015) and the US (Du et al., 2022). Artificial neural networks are a type of deep learning method that use layers of interconnected nodes or neurons in a structure inspired by the human brain. Artificial neural networks have been used to forecast daily maximum ozone in Greece (Chaloulakou et al., 2003), South Korea (Eslami et al., 2020) and Germany (Kleinert et al., 2021; Deng et al., 2022). ML ensembles are methods that combine multiple ML models to improve the accuracy and robustness of predictions. ML ensembles can improve overall accuracy of ozone forecasts (Gong and Ordieres-Meré, 2016), and have been

shown to perform better than the best performing model within the ensemble (Mallet et al., 2009).

1.4 Thesis contributions

The contributions of this thesis are motivated by the need to better understand the risk and drivers of high-level ozone in the UK, especially in a warming climate, and the need for improved tools with which to model ozone. The overarching objective of the thesis is to exploit the power of sophisticated data science methods to address gaps in our knowledge of high-level UK ozone. The aims of this thesis are as follows:

1. To quantify extreme ozone events evaluating their magnitude, frequency, and likelihood and examining how these events have changed over time.

2. To determine the variability of ozone in both space and time and identifying long-term trends that are relevant to policymaking.

3. To assess the association of high concentrations of ozone with various meteorological, spatial and temporal factors.

4. To produce data-driven forecasts for the occurrence of high concentrations of ozone.

Chapter 2 presents a comprehensive EVA of UK surface level ozone using a temperature-dependent extremes model. The magnitude, frequency, and likelihood of extreme ozone events are estimated at ozone measurement stations across the UK, including an analysis into how these have changed over time. Further, changes in temperature-dependent risk of high-level ozone are presented.

Chapter 3 develops a ML downscaling methodology to downscale a CTM ozone surface from a 5×5 km to 1×1 km resolution, using a gradient boosted tree. The downscaled surface is comprehensively evaluated and shown to better represent measurement ozone, particularly high-level ozone. An analysis is performed on the

downscaled surface, original CTM surface and measurement data, and the current state of UK-wide ozone is presented along with trends over time. Finally, the effects of three NO_X reduction scenarios on UK ozone are considered.

Chapter 4 presents a ML classification-based experimental analysis into the drivers of high-level ozone measurement data, using gradient boosted trees. The impact of synoptic weather, meteorological, spatial and temporal features on high-level ozone is identified. The ML classification method is used to forecast the presence or absence of high-level ozone events, exploring the potential efficacy of ML approaches for operational forecasts.

Chapter 5 concludes the thesis with a summary of all findings including discussion on the contributions to new knowledge of UK surface ozone and the new application and development of surface ozone modelling methods. Limitations of this work are also discussed alongside future work.

Chapter 2 is a peer reviewed and published paper, Chapter 3 is under review and Chapter 4 is in the process of being submitted. Consequently, more specific introductions and motivation are including in each chapter.

Chapter 2

A temperature dependent extreme value analysis of UK surface ozone, 1980–2019

Lily Gouldsbrough[1], Ryan Hossaini[1,2], Emma Eastoe[3], Paul J. Young[1,2]

[1]Lancaster Environment Centre, Lancaster University, Lancaster, UK

[2]Centre of Excellence in Environmental Data Science, Lancaster University, Lancaster, UK

[3]Department of Mathematics and Statistics, Lancaster University, Lancaster, UK

Corresponding author: Lily Gouldsbrough (l.gouldsbrough@lancaster.ac.uk)

The following work was published in Atmospheric Environment on 15th March 2022 (citation: A temperature dependent extreme value analysis of UK surface ozone, 1980–2019, Atmospheric Environment, Volume 273, 2022, 118975, ISSN 1352-2310, https://doi.org/10.1016/j.atmosenv.2022.118975). The authors contributions are listed on the following page.

Statement of contribution. Lily Gouldsbrough led the methodology, data curation and analysis and conceived the study alongside Ryan Hossaini, Emma Eastoe and Paul J. Young. Lily Gouldsbrough drafted the manuscript, with guidance from Ryan Hossaini, Emma Eastoe and Paul J. Young.

Abstract

Elevated surface ozone during heatwaves and recent hot summers raises concerns over the potential for climate change to exacerbate ozone air pollution in the UK. In this paper, we perform a robust statistical analysis of four decades worth of daily maximum 8-hour (MDA8) ozone measurements from the UK's Automated Urban and Rural Network. A temperature dependent extreme value model is developed to characterise the magnitude and frequency of extreme ozone events and to determine probabilities for ozone exceeding health thresholds, as defined in the UK's air quality index. Our model is found to describe the tails of the MDA8 ozone distributions well at all 119 monitoring sites considered. For the decade 2010–2019, we estimate that >90% of sites have a 1-year MDA8 ozone return level greater than the 'moderate' ozone threshold of $100\,\mu g/m^3$. We also find that 33% of sites are currently expected to breach the UK government's national air quality objective that MDA8 ozone should not exceed $100\,\mu g/m^3$ more than ten times per year. We estimate the present overall probability of MDA8 ozone exceeding $100\,\mu g/m^3$ on a given day to be between <0.1% and 5.4%, depending on site, with averages of 2.7% (rural) and 1% (urban background locations). Our analysis reveals a significant decline over time in the likelihood of the UK experiencing extreme ozone episodes, with 1-year return levels in the 1980s now roughly comparable to 10-year return levels in the present. Similarly, probabilities of MDA8 ozone exceeding $100\,\mu g/m^3$ have decreased by a factor of ∼2–6 since the 1980s in some locations. However, our results also highlight a strong positive temperature dependence to the risk of ozone exceedances. In consequence, increasingly hot summers due to climate change may offset some of these gains.

2.1 Introduction

Tropospheric ozone (O_3) is a short-lived secondary air pollutant and greenhouse gas. It is formed in the presence of sunlight through a complex set of chemical reactions involving precursor pollutants nitrogen oxides (NO_X = NO_2 + NO) and volatile organic compounds (VOCs). At ground level, ozone concentrations vary spatially and temporally on seasonal, interannual and decadal time-scales due to heterogeneity in sources and sinks, meteorological variability and trends in precursor emissions from natural and anthropogenic sources (O. R. Cooper et al., 2014; P. S. Monks et al., 2015; Pope et al., 2016). Strong evidence exists linking short-term ozone exposure to respiratory health issues, hospital admission and mortality (Ji et al., 2011; COMEAP, 2015; Nuvolone et al., 2018), and it has been estimated that, globally, exposure to elevated ozone caused an added 254,000 deaths and a loss of 4.1 million disability-adjusted life years (DALYs; number of years lost due to ill health, disability or early death) from chronic obstructive pulmonary disease in 2015 (Cohen et al., 2017). The damaging effects of ozone on crops and their yields are also well documented (e.g. Van Dingenen et al., 2009. Of special relevance to short-term exposure are episodes of elevated ozone concentrations. Such episodes involve only the highest values in the dataset and are sometimes referred to instead as extreme events. Understanding the drivers of such events has been a significant area of research in recent years, motivated in part by concern that climate change could exacerbate air pollution (e.g. Otero et al., 2016; Y. Zhang and Yuhang Wang, 2016; M. Lin et al., 2020). Here, we use extreme value analysis to investigate the magnitude and frequency of extreme ozone events and to determine trends over time.

In the UK, surface ozone is monitored routinely at more than a hundred locations nationwide, with records dating back to the 1970s at some sites. Despite the availability of this rich set of ozone data, only a few studies have investigated the occurrence, likelihood and spatiotemporal variability of high level ozone. Diaz et al. (2020) analysed UK ozone trends over the period 1992 to mid-2019 at selected sites.

They reported positive trends in annual mean ozone of 0.13 ppb/yr (0.5%/yr, p <0.001) and 0.20 ppb/yr (1.1%/yr, p <0.001) across 13 rural and 6 urban sites, respectively. Annual maximum ozone was shown to have decreased over the same period at a rate of 1.0 ppb/yr and 0.68 ppb/yr. Using all available monitoring data, Finch and Palmer (2020) reached similar conclusions for annual mean ozone levels. They found no significant trends in the annual maximum, but did find a reduction in the magnitude and occurrence of high levels of ozone at all but one site. These studies broadly corroborate and update earlier work on UK ozone trends (Jenkin, 2008; Munir et al., 2012).

In this study extreme value analysis (EVA) is used to investigate extreme ozone events in the UK. EVA provides a robust, flexible statistical method to model and analyse observations that are unusually large (or small), i.e., values in the tail of the sample distribution. It is a valuable tool for estimating both the magnitude and the probability of extreme events and has found use in a range of environmental applications, including extreme precipitation (e.g. Towler et al., 2020), temperatures (e.g. Leeson et al., 2018) and wind speeds (e.g. Hundecha et al., 2008). There are two main EVA methods: generalised extreme value (GEV) and peak-over-threshold (POT) models. GEV models are appropriate for block maxima (or minima) of time series data, such as annual maxima, whereas the POT model is applied to all peak values that exceed (or fall below) a pre-defined high (or low) threshold. Both models have stationary and non-stationary versions.

Rieder et al. (2013) applied a stationary POT model to summer ozone concentrations in the US. They found that the frequency and magnitude of high ozone events has, over time, significantly declined in response to air quality regulations aimed at lowering NO_X emissions. Along with others (Phalitnonkiat et al., 2016), their results highlight the usefulness of EVA in assessing the long-term efficacy of air quality interventions. Stationary POT models have also been used to assess air quality in areas such as Peninsular Malaysia (Masseran et al., 2015), Istanbul (Saygın et al., 2018), and Barcelona, where there was found to be

a pronounced seasonal effect of ozone levels exceeding European health threshold levels (Tobías and Scotto, 2005). However, environmental time-series often display non-stationarity (e.g. shifting mean state) and both GEV and POT models are routinely adapted to reflect this non-stationarity, e.g. the seasonal patterns and trends seen in ozone observations. A non-stationary GEV model, incorporating precursor pollutants and meteorological variables, applied to ozone episodes in Delhi found that the highest ozone concentrations are most likely to occur in the monsoon season and the lowest in the winter season (Hazarika et al., 2019). A few studies do apply EVA to UK air quality including non-stationarity but all are limited to analysis of a single site (Eastoe, 2009; Eastoe and Tawn, 2009; Gyarmati-Szabó et al., 2017).

Observational studies in many parts of the world report positive ozone-temperature correlations (e.g. Bloomer, Stehr, et al., 2009). From a process standpoint, efforts to understand the ozone-temperature relationship are still underway. However, the positive ozone-temperature association may reflect several underlying mechanisms (e.g. W. Sun et al., 2017; Romer et al., 2018; Porter and Heald, 2019): (1) the net effect of temperature-dependent reaction kinetics that govern production and loss of ozone and its precursors, (2) the association of temperature with air stagnation at the synoptic scale, and (3) the temperature-dependence of biogenic ozone precursor emissions. Further, dry and hot weather conditions can cause vegetation to become stressed, resulting in reductions in dry deposition leading to maintained high ozone concentrations (M. Lin et al., 2020). In the UK, elevated ozone concentrations occur in spring and summer, accompanying warmer temperatures. Pope et al. (2016) showed that summertime ozone was particularly elevated under anticyclonic conditions and when the UK is subject to a south-easterly weather flow type. While studies to understand the effects of temperature on surface ozone in parts of Europe have been conducted (e.g. Noelia Otero, Rust, et al., 2021), few UK-focused studies exist on temperature and extreme ozone. Finch and Palmer (2020) demonstrated that the likelihood of

UK ozone health threshold exceedances (>100 µg/m^3) increases with temperature: 50% of observed exceedances were found to occur above 18 °C for rural monitoring sites and 21 °C for urban monitoring sites.

In this study, a temperature dependent extreme value model is developed and applied to daily maximum 8-hour average (MDA8) ozone concentrations recorded at 119 measurement sites across the UK. The relationship between daily maximum temperature and MDA8 ozone is characterised for each site. Our objectives are to (1) determine the probabilities, magnitudes and frequencies of extreme ozone episodes across the UK; (2) examine how extreme ozone events have changed on decadal timescales; and (3) characterise how the extremal characteristics of ozone are influenced by temperature. Section 2.2 describes the measurements of surface ozone and temperature data used in the analysis. Section 2.3 provides a technical overview of our EVA approach. Results are presented in Section 2.4, including analysis of present-day ozone extremes (Section 2.4.1), trends over decades (2.4.2) and the influence of temperature (2.4.3). A summary and conclusions are given in Section 2.5.

2.2 UK surface ozone and temperature data

We calculated MDA8 ozone from hourly concentration data reported for 119 measurement sites of the UK's Automatic Urban and Rural Network (AURN): https://uk-air.defra.gov.uk. The network is a major component of the UK's air quality monitoring infrastructure. It provides near-continuous statutory nationwide monitoring of the ambient surface concentration of ozone and other important air pollutants (e.g. NO_X, particulate matter). AURN monitors ozone concentrations for six different site types: rural background, urban background, suburban background, urban industrial, suburban industrial and urban traffic. Most of the sites used in this paper (83% of the total) are rural or urban background. Table A.1 provides information on each site, with the geographical regions in which sites are located

shown in Figure A.1. We focus on the period 1980–2019, though the length of the ozone record varies across sites within this period. Seven monitoring sites have data records since the 1980s and 30 monitoring sites have data records since the 1990s, allowing for decadal changes to be examined at these locations.

The UK Department for Environment, Food and Rural Affairs (Defra) defines a Daily Air Quality Index (DAQI) to provide the public with health-related information based on the levels of various air pollutants. Our analysis uses the thresholds defined in the DAQI as relevant health threshold metrics to classify ozone concentrations exceeding 100 µg/m^3, 160 µg/m^3 and 240 µg/m^3 as 'moderate', 'high' and 'very high', respectively. These thresholds are based on the maximum of the running 8-hour mean ozone concentration (MDA8). In subsequent sections we consider how surface temperature influences the likelihood of exceeding these thresholds.

The daily maximum temperature data used for this analysis are from the HadUK-Grid 1 km gridded dataset (Office et al., 2020). This dataset is constructed based on UK station observations from the Met Office's extensive Integrated Data Archive System (MIDAS) and has been through rigorous quality control and evaluation (Hollis et al., 2019).

2.3 Extreme Value Analysis model

2.3.1 Ozone season definition

It is well established from extensive global monitoring programmes that surface ozone concentrations exhibit a marked seasonal cycle (Scheel et al., 1997; Paul S Monks, 2000). In the UK, ozone is generally elevated in boreal spring and summer and at a minimum in boreal autumn and winter. Using all available ozone data from 1980–2019, we define a site-specific 'ozone season' as the months where monthly mean MDA8 ozone is greater than the 40$^{\text{th}}$ percentile of all MDA8 ozone observations at the site. This provides an objective, data-driven means to extract only periods

with elevated ozone. We find variation in when each site experiences elevated levels of ozone, as shown for two example sites in Figure 2.1. A summary of each site's ozone season is given in Table A.1. Our ozone season definition captures 98% of days on which the MDA8 ozone concentrations exceed the moderate health threshold level of $100\,\mu g/m^3$ across all sites. Consequently, for the rest of the paper we consider only MDA8 ozone within each site's ozone season (typically March–September). Subsetting the data in this way focusses attention on periods most relevant to human health.

2.3.2 Extreme value model

Extreme value models are probabilistic models that are used to extrapolate beyond a sample of data to estimate values that are rarer than those observed in-sample. Extreme value models can be either stationary or non-stationary. A stationary model assumes that the extreme values have no trends or seasonality, whereas a non-stationary extremes model accounts for trends and/or seasonality. The classical approach to modelling extremes is based on block maxima (Fisher and Tippett, 1928), where a block is an amount of time. In this method, the blocks for the block maxima must be large (e.g. annual maxima) for the modelling assumptions to hold. Furthermore, if data other than block maxima are available, retaining only one value per block can be wasteful of the information contained in other large values. The POT method that we use instead specifies a model for all exceedances above a chosen high threshold u (Davison and Smith, 1990) utilising all extreme data and allowing blocks to contribute zero, one or multiple events.

2.3.2.1 Threshold choice

To separate the extreme MDA8 ozone observations from non-extreme values, we first defined a modelling threshold u. This threshold is inherent to the POT approach and is not related to any health threshold concentration of ozone. The choice of modelling threshold is a crucial step for two reasons. First, a threshold that is too

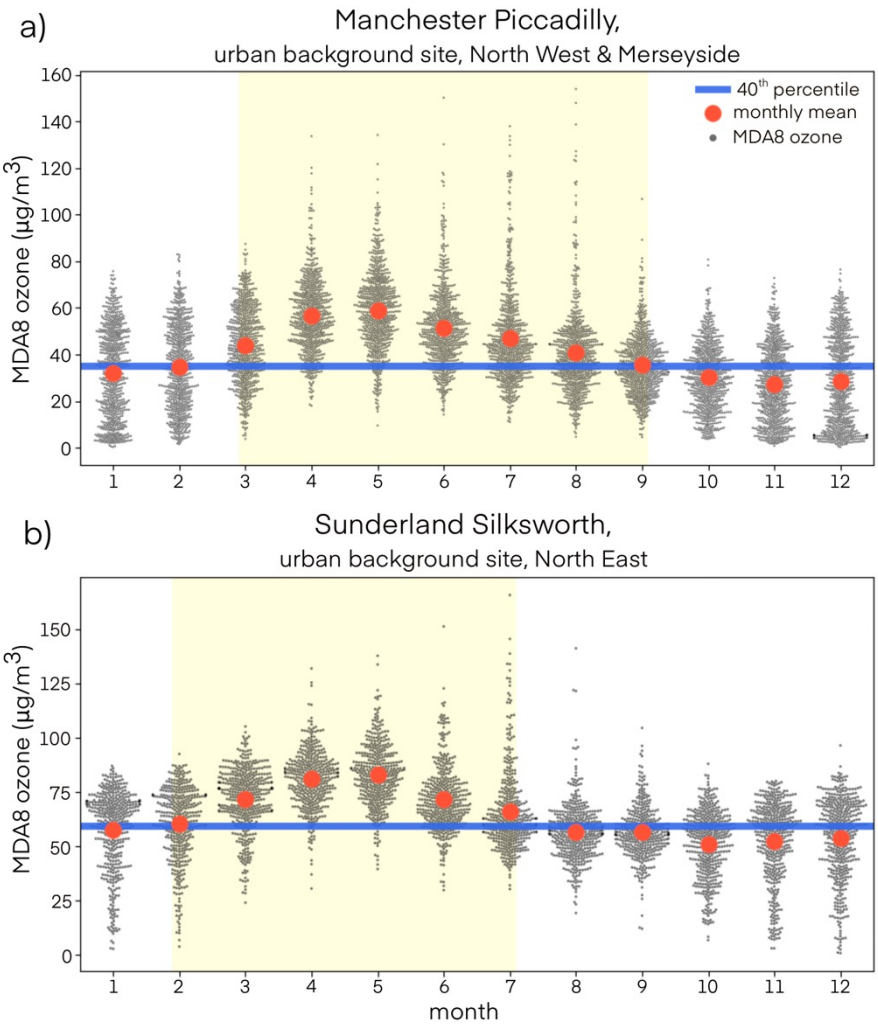

Figure 2.1: The ozone season period (highlighted in yellow) for two example sites: (a) Manchester Piccadilly and (b) Derry Rosemount. The grey swarm plots are the MDA8 ozone observations (µg/m^3) from 1980–2019 (where available). The blue line is the 40$^{\text{th}}$ percentile of all MDA8 ozone observations. The red dots are the monthly mean MDA8 ozone.

low leads to bias within the parameter estimates, caused by non-extreme data being included in the model. Second, a too high threshold will lead to high variance in the parameter estimates, because too few data will have been used to fit the model. A common compromise is to set a stationary threshold to be the 90^{th} percentile of the dataset.

Ozone is formed in the presence of sunlight, and particularly high concentrations are often reached during heatwaves or periods of prolonged warm temperatures with stagnant meteorological conditions (e.g. Vieno, Dore, et al., 2010; Pope et al., 2016; Varotsos et al., 2019). In this work, the relationship between ozone and temperature was evaluated at all sites considered. We find a non-linear relationship between MDA8 ozone and daily maximum temperature as shown for two example sites in A.2. Due to the seasonal behaviour of ozone, use of a stationary threshold (e.g. the fixed 90^{th} percentile of the data) would lead to modelling only the peak ozone concentrations that typically occur with higher temperatures. To mitigate this, we use a temperature dependent modelling threshold. This way, we model extreme values of ozone relative to the observed temperature, resulting in a lower (higher) extreme threshold for colder (warmer) temperatures.

To illustrate the above concepts, Figure 2.2 shows scatter plots of MDA8 ozone versus daily maximum temperature at three example sites. The fit of three different modelling thresholds were evaluated: a stationary threshold, a temperature dependent linear threshold, and a temperature dependent natural cubic spline threshold. Since the natural cubic spline provides the best approximation to the non-linear ozone-temperature relationship, we set the modelling threshold u, to be the 90th percentile of a natural cubic spline regression, i.e., the concentrations above the green line in Figure 2.2 are deemed to be 'extreme', with the threshold computed individually for each site. Previous studies have reported positive, non-linear relationships between ozone concentration and temperature (e.g. Bloomfield et al., 1996).

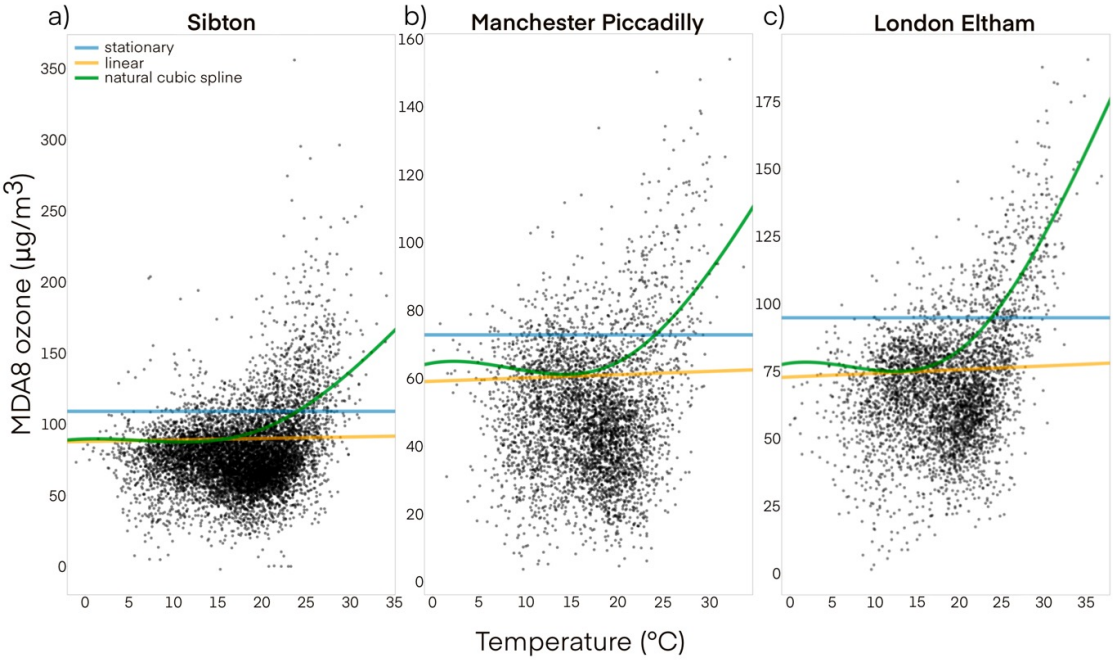

Figure 2.2: Examples of three different extreme modelling thresholds for MDA8 ozone for three different monitoring sites: (a) Sibton, (b) Manchester Piccadilly and (c) London Eltham. The coloured lines indicate different threshold definitions: blue is a stationary threshold, set at the 90$^{\text{th}}$ percentile of all MDA8 ozone observations; orange is a linear threshold, set at the 90$^{\text{th}}$ percentile of a linear regression between MDA8 ozone and daily maximum temperature; and green is a threshold that varies with temperature, consisting of the 90$^{\text{th}}$ percentile of a natural cubic spline regression between MDA8 ozone and daily maximum temperature. This latter threshold is used in the study.

2.3.2.2 Generalised Pareto distribution

The next step is to apply the generalised Pareto distribution to the MDA8 ozone observations that exceed the temperature dependent modelling threshold. The generalised Pareto distribution (GPD) was introduced by Pickands (1975) as a way of modelling tails of distributions, and was popularised by Davison and Smith (1990). It is defined by three parameters: σ, ξ and ϕ, which denote scale, shape and rate of exceedance, respectively. Suppose that the sequence of MDA8 ozone observations X_i is an independent and identically distributed sample. Then, for a large enough threshold u, the distribution function of the excesses above the threshold $y = (X - u)$, conditional on $X > u$, is approximately

$$H(y) = 1 - \left(1 + \frac{\xi y}{\sigma}\right)^{-1/\xi}, \quad y > 0 \tag{2.1}$$

Equation 2.1 denotes the generalised Pareto family of distributions. The next step in applying the GPD in the peaks-over threshold method is estimating the model parameters. We use maximum likelihood estimation to do this. This involves finding the maximum of a likelihood function, which describes the probability of the observations as a function of the GPD parameters.

2.3.2.3 Parameter estimation: scale and shape

To estimate σ and ξ we must check if there is any additional temperature effect on the k excesses above the temperature dependent threshold $y_1, \ldots, y_k$. If there is no temperature effect, we can use a stationary extremes model where the log-likelihood, the natural logarithm of the likelihood, is derived from equation 2.1 as

$$\ell(\sigma, \xi) = -k \log \sigma - (1 + 1/\xi) \sum_{i=1}^{k} \log\left(1 + \xi y_i/\sigma\right), \tag{2.2}$$

provided $\xi \neq 0$ and $(1 + \sigma^{-1}\xi y_i) > 0$ for $i = 1, \ldots k$; otherwise, $\ell(\sigma, \xi) = -\infty$. When $\xi = 0$ the log-likelihood is defined as

$$\ell(\sigma) = -k \log \sigma - \sigma^{-1} \sum_{i=1}^{k} y_i. \tag{2.3}$$

If there is a temperature effect, a non-stationary temperature dependent extremes model can be used. The log-likelihood for the temperature dependent extremes model can be found by replacing σ in equation 2.2 with $\sigma(T_i) = \exp(\sigma_0 + \sigma_1 T_i)$. The new temperature dependent log-likelihood can be written as

$$\ell(\sigma, \xi) = \sum_{i=1}^{k} \log \sigma\left(T_i\right) - (1 + 1/\xi) \sum_{i=1}^{k} \log\left(1 + \xi y_i / \sigma\left(T_i\right)\right), \tag{2.4}$$

where T_i are the daily maximum temperatures corresponding to y_i. Note the use of the exponential function in defining $\sigma(T_i)$ is to ensure that the scale parameter is always positive regardless of the temperature.

As the stationary model is a special case of the non-stationary model (when $\sigma_1 = 0$), the models are nested. To decide if temperature is significant in the scale parameter σ, a deviance test can be applied to the maximum likelihood estimation of the nested models, which is a goodness-of-fit measure. With the stationary and non-stationary models $M_0 \subset M_1$, the deviance test is defined as

$$D = 2\left[\ell_0\left(M_1\right) - \ell_0\left(M_0\right)\right]$$

where $\ell_0(M_0)$ and $\ell_1(M_1)$ are the maximised log-likelihoods of models M_0 and M_1, respectively. Figure A.3 shows the mean deviance test statistic for each region and each site type. Comparing site types, urban industrial sites, followed by rural background sites, have the highest residual effect from temperature in the scale parameter σ for the model. Urban traffic sites have the lowest residual effect from temperature. We find temperature to be significant in the scale parameter σ for all sites when compared to the χ^2 distribution at a 5% significance level. Therefore, the non-stationary temperature dependent extremes model was used for the rest of the analysis.

2.3.2.4 Parameter estimation: rate of exceedance

In addition to σ and ξ, the rate of exceedance parameter ϕ is needed to quantify the magnitude of a predicted extreme event. For the chosen model, the rate of

exceedance is conditional on the temperature, which can be written formally as $\phi(T_i) = Pr(X_i > u(T_i)|T_i = t_i)$, where T_i is the temperature on day t_i. We use a binary logistic regression to find the rate of threshold exceedance, conditional on temperature. A binary logistic regression is a model used when there are only two outcomes from a dependent variable i.e., a threshold exceedance or not. The resulting log-likelihood for the rate of exceedance $\phi(T_i)$ is

$$\ell\left(\phi\left(T_i\right)\right) = \sum_{i=1}^{k}\left[s_i \log\left(\frac{\exp\left(\beta_0 + \beta_1 T_i\right)}{1+\exp\left(\beta_0 + \beta_1 T_i\right)}\right) + (1-s_i)\log\left(\frac{1}{1+\exp\left(\beta_0 + \beta_1 T_i\right)}\right)\right]$$

where s_i is the indicator variable for a threshold exceedance.

2.3.3 Return levels and return periods

A return level is defined as the value that is expected to be equalled or exceeded on average once every interval of time t with a probability of $1/t$. The N-year return level when $\xi \neq 0$ is defined by

$$z_N = u_T + \frac{\sigma_T}{\xi}\left[(N n_y \phi_T)^{\xi} - 1\right] \tag{2.5}$$

where u_T is the threshold value and $\phi_T = PrX > u_T$, both for a specified temperature T, and n_y is the number of observations within a year.

A return period (R.P.) is the predicted length of time until an observance of a specified level occurs, and is defined as

$$R.P. = \sum_{i=1}^{n_y} \phi\left(T_i\right)\left[1 + \xi\left(\frac{U - u_{T_i}}{\sigma_{T_i}}\right)\right]^{1/\xi}, \tag{2.6}$$

where T_i is a temperature series for 1 season, and U is the specified level of ozone.

2.3.4 Model training, model fit and uncertainty estimates

We fit individual extreme value models for each monitoring site. For the present-day analysis (Section 2.4.1), we trained our model on data from 2010–2019. For the decadal trends analysis (Section 2.4.2), our model is trained on individual decades

of data (1980–1989, 1990–1999, 2000–2009, 2010–2019). Only sites with at least three years of data for the given decade were analysed. We evaluated the fit of our model using quantile-quantile plots, a graphical method to assess if two datasets come from populations with a common distribution. Due to the non-stationarity, we used a 1:1 mapping to transform the model-based percentile estimates onto a "quantile" scale, mapping to a common exponential scale to magnify the tail fit and aid comparison between sites. Figure A.4 shows examples of the 2000–2009 model fit for sites in the Greater London region. The peaks-over-threshold model provides a good fit to the tail of the MDA8 ozone dataset at each site within this region and the average R^2 result for every region in the UK was found to be >0.9 (not shown).

In addition to temperature being used in defining the modelling threshold, the parameters that define the GPD are also temperature dependent (Section 2.3.2). This allows us to explore the influence of temperature on the likelihood and magnitude of extreme ozone events. The decision to include temperature as the only covariate in our model was made to facilitate interpretation of the model outputs (following Shen et al., 2016). In future work, trends driven by other factors could be considered, including precursor pollutants, such as NO_X and VOCs, and other meteorological variables, such as wind speed, wind direction and relative humidity.

A benefit of our non-stationary model over the stationary model is that it allows extremal characteristics of the ozone data to be quantified either dependent or independent of the temperature. Such characteristics include return periods, return levels or the probability of ozone exceeding a pre-defined health threshold. Temperature dependent outputs describe risk for a given temperature and can be obtained directly from the trained GPD model. Non-temperature dependent outputs describe risk after accounting for the temperature distribution over the period for which we are making estimates. Non-temperature dependent outputs were obtained by simulating 1000 years of MDA8 ozone from the trained GPD model (Text A.3), then estimating the desired outputs from the empirical distribution of the simulated data. For example, the 100-year return level estimate would be the 10^{th} largest

value in the simulated 1000-year dataset. Note the 1000-year dataset is a tool used for predicting risk for the current time-period and should not be interpreted as a forecast of MDA8 ozone for the next 1000 years.

Confidence intervals (CI) for the estimates were obtained using a non-parametric bootstrap algorithm. We simulated 1000 data samples of the same size as the original dataset by sampling (with replacement) from the original dataset. We then retrained the GPD model on each bootstrapped data. Using the trained models, we can predict any temperature or non-temperature dependent output of interest using the same methods as applied to the original dataset. For each output (return level, etc), the 2.5% and 97.5% quantiles of the resulting 1000 bootstrap estimates give the bootstrapped 95% confidence interval which is then reported alongside our estimates as [lower bound, upper bound].

2.4 Results and discussion

We use the extreme value model both to estimate return levels for MDA8 ozone associated with pre-defined return periods (1-, 10- and 100-year) and to quantify the likelihood of an exceedance of each of the three UK health thresholds defined in Section 2.2. The latter is measured by (i) estimating the return period associated with the health threshold and (ii) comparison with estimated return levels associated with known return periods. We use these measures to illustrate both inter-decadal and temperature-related changes in the frequency and magnitude of extreme events as well as the usefulness of the current health thresholds in identifying such events for different decades and temperatures. We show that for some decades and temperatures, an exceedance of the moderate health threshold ($100\,\mu g/m^3$) as currently defined would not be considered a rare event.

2.4.1 Present-day analysis (2010–2019)

We first consider results from the model fitted to the most recent 'present-day' decade (2010–2019) of data for 87 monitoring sites across the UK. Figure 2.3 shows the 1-, 10- and 100-year return levels of MDA8 ozone estimated using the simulation method (see Section 2.3.4). The 1-year return level exceeds the moderate health threshold (100 µg/m^3) at 93% of sites. In contrast, the percentage of sites which have 1-year (10-year) return level that exceeds the high health threshold (160 µg/m^3) are 0% (15%). No sites have a 100-year return level that exceeds the very high health threshold (240 µg/m^3) implying that it is a far rarer event. This threshold does fall within the CI at nine sites suggesting that an exceedance of this level is plausible at these sites, but on a 100-year time scale. We further note a greater spread in CIs at these nine sites, with the highest (lowest) CI range being 129 µg/m^3 (71 µg/m^3) compared with 82 µg/m^3 (18 µg/m^3) for all other sites.

To consider the rarity of a health threshold exceedance more directly, Figure 2.4 shows the predicted exceedance count per year for the 100 µg/m^3 moderate health threshold. The highest numbers of predicted exceedance days occur in the East and South East regions. Two rural background sites (Weybourne in the East and Chilbolton Observatory in the South East) are predicted to exceed the moderate health threshold more than 25 times per year, whilst three rural background sites (High Muffles, Yarner Wood and Charlton Mackrell) are predicted to have over 20 exceedances per year. We also find that 29 out of 87 sites are estimated to breach the moderate health threshold over 10 times per year, thus failing one of the UK's national air quality objectives (Defra, 2020).

To further characterise risk, the probabilities of exceeding the health thresholds of the UK's air quality index are presented in Figure 2.5. To our best knowledge, this probabilistic analysis is the first of its kind for the UK. Based on the model fit to 2010–2019 data, we find the probability of MDA8 ozone exceeding 100 µg/m^3 on a given day within the ozone season to be between <0.1% and 5.4%. We find the cross-site average probability is greater for rural sites (0.027, i.e. 2.7%) vs urban

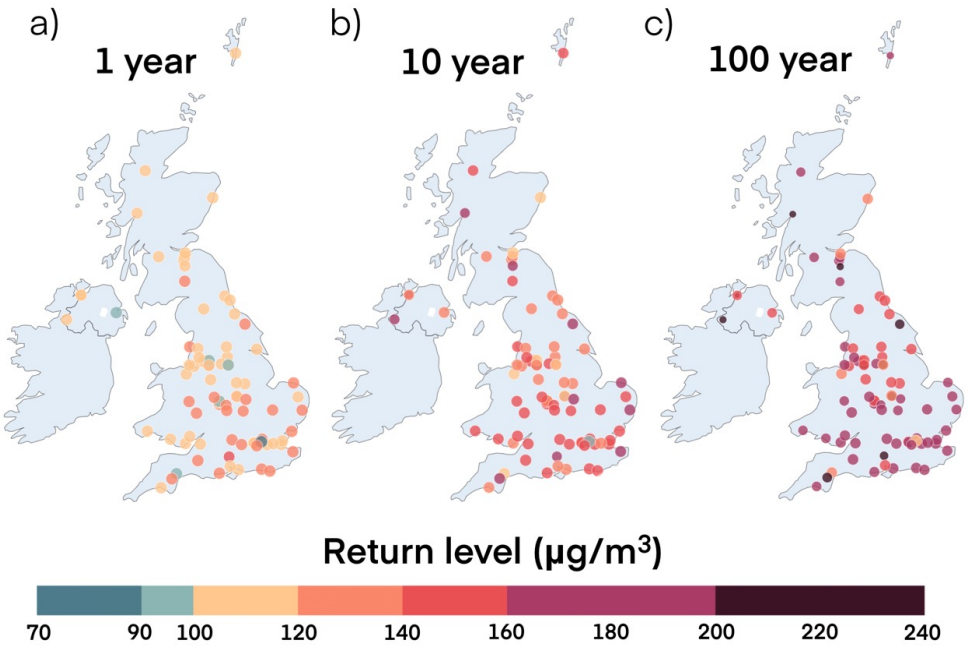

Figure 2.3: The (a) 1-, (b) 10-, and (c) 100-year return levels of MDA8 ozone ($\mu g/m^3$) for each AURN measurement site in the UK, found by using the simulation method described in Section 2.3.4. The size of the dot is proportional to the standard error of the estimate, with a highest standard error for the 100-year return level of $\sim 30\,\mu g/m^3$.

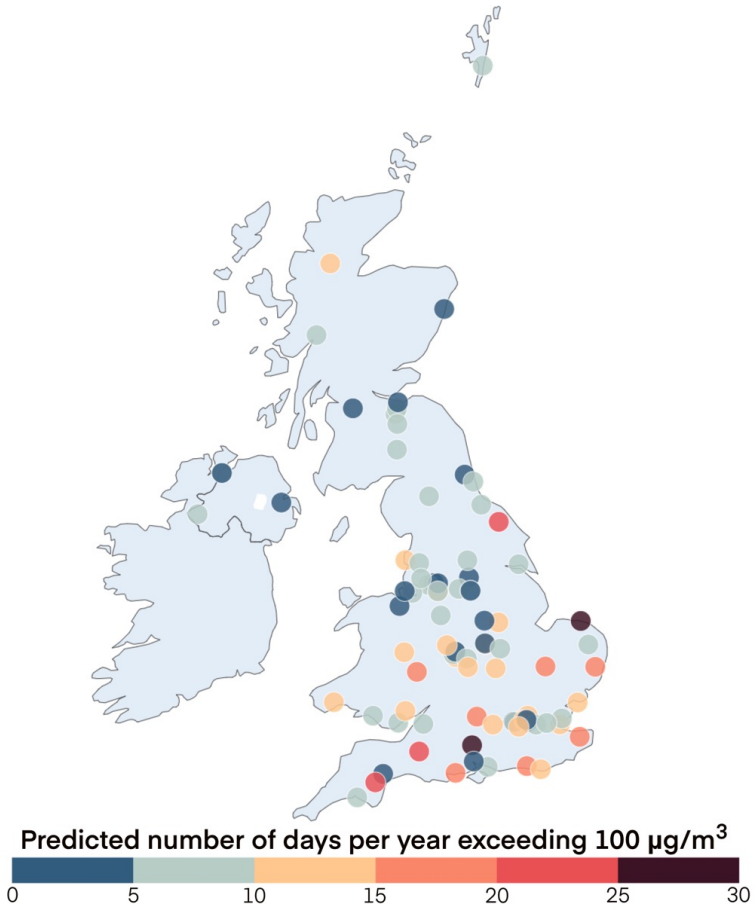

Figure 2.4: The number of predicted exceedance days per year for the moderate health threshold level of 100 µg/m^3 for each UK measurement site used in the study, found using the simulation method described in Section 2.3.4. Based on analysis of the decade 2010–2019.

sites (0.016 or 1.6%). This is consistent with the 'urban decrement', i.e. that higher ozone levels are generally found at rural over urban sites (e.g. Munir et al., 2012; P. S. Monks et al., 2015) due to the titration of ozone from higher NO_X levels in urban regions. There is, however, substantial cross-site variation in the exceedance probabilities. The variability across the sites is 1.5 times greater for rural sites than urban sites, with standard errors for the estimated group means of 0.0021 and 0.0014, respectively. The mean probability of an exceedance decreases for both site types at higher health thresholds (Figure 2.5) when compared with the moderate health threshold: 75 (100) times lower at rural (urban) sites for the high health threshold, and ∼7,000 (∼12,000) times lower at rural (urban) sites for the very high health threshold.

Some sites have a negative shape parameter in the GPD model, which implies an upper limit to the range of plausible ozone values. To demonstrate the effect of this, Figure 2.5 (see annotated numbers) shows the estimated proportion of sites that the model permits to exceed each health threshold. All rural and urban sites are found to be able to exceed the moderate health threshold. More rural background (96% / 25%) than urban background sites (86% / 10%) can conceivably exceed the high / very high health thresholds.

2.4.2 Trends over decades

Figure 2.6 shows the percentage change in estimated 1-, 10- and 100-year return levels for the 2010s compared to earlier decades. All seven sites online since the 1980s have seen a significant decrease in 1- and 10-year return levels and six have seen a significant decrease in 100-year return levels. Decreases in the 1-year return level between the 2010s and 1980s range from -16% to -44%, equivalent to absolute decreases of 31–82 µg/m^3 (Table A.2). At these sites, the 1-year return level of MDA8 ozone in the 1980s is comparable to the 10-year return level in the 2010s. That is, what was once an ozone concentration expected to be equalled or exceeded at least once in 1 year is now expected to be equalled or exceeded only once in

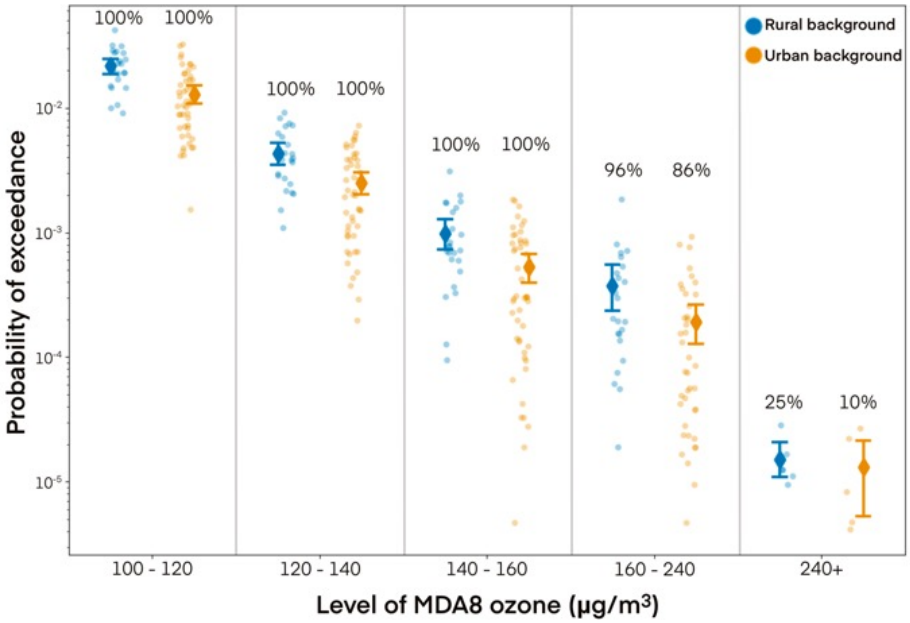

Figure 2.5: The non-zero probabilities of exceeding various levels of MDA8 ozone, found using the simulation method described in Section 2.3.4. The background dots are the individual measurement sites, grouped into rural background (blue) and urban background (orange) site types. The diamonds are the mean estimates for each group, with the 95% confidence intervals of the mean estimate also shown by the error bars. The percentage numbers above each group show the proportion of sites which can reach the given level. Based on analysis of the decade 2010–2019.

10 years. Qualitatively, similar reductions have been reported for the Eastern US using EVA, attributed to reductions in NO_X emissions Rieder et al., 2013. For the 32 UK sites monitoring since the 1990s, 69%, 75% and 75% of sites have seen a significant decrease in their 1-, 10- and 100- year return levels, respectively. For the 68 sites monitoring since the 2000s, 66%, 59% and 46% of sites have seen a significant decrease in their 1-, 10- and 100- year return levels, respectively.

To compare decadal changes in the frequency of predicted health threshold exceedances, the changes in return periods corresponding to ozone health thresholds are shown in Figure A.5. For the seven sites monitoring since the 1980s, significant increases in the return period are seen in five sites for the moderate health threshold, seven sites for the high health threshold and six for the very high health threshold level. For the 32 (68) sites monitoring since the 1990s (2000s), 59% (66%), 75% (54%), 91% (78%) of sites have seen a significant increase in return period for the moderate, high and very high health threshold level, respectively. This signifies that high ozone events are occurring less often. Across all health thresholds and decades, the only significant decrease in return period was -0.5% [-0.1%, -0.75%] (moderate health threshold) at Manchester South, a suburban industrial site in the North West and Merseyside region. We do not overinterpret this single site finding, though note that Finch and Palmer (2020) reported for this particular site (1) a relatively strong positive annual mean 8-hr ozone trend (1999–2019) of 0.74 $\mu g/m^3 yr^{-1}$ compared to the all-site mean (0.41 $\mu g/m^3 yr^{-1}$), along with a positive NO_2 trend, in contrast to negative trend at all other sites.

Extending the analysis in Section 2.4.1, Figure 2.7 shows the change over decades in the probabilities of exceeding various levels of MDA8 ozone from 100 to 240 $\mu g/m^3$ for all monitoring sites online within each decade. The mean estimates for each decade show a general reduction in the probabilities of exceedances for all level groupings of ozone since the 1980s. The proportion of sites in the 2010s that can theoretically reach levels of ozone of 160 $\mu g/m^3$ (i.e., where the probability of an exceedance is greater than zero) has fallen to 86% compared to 100% for the 1980s.

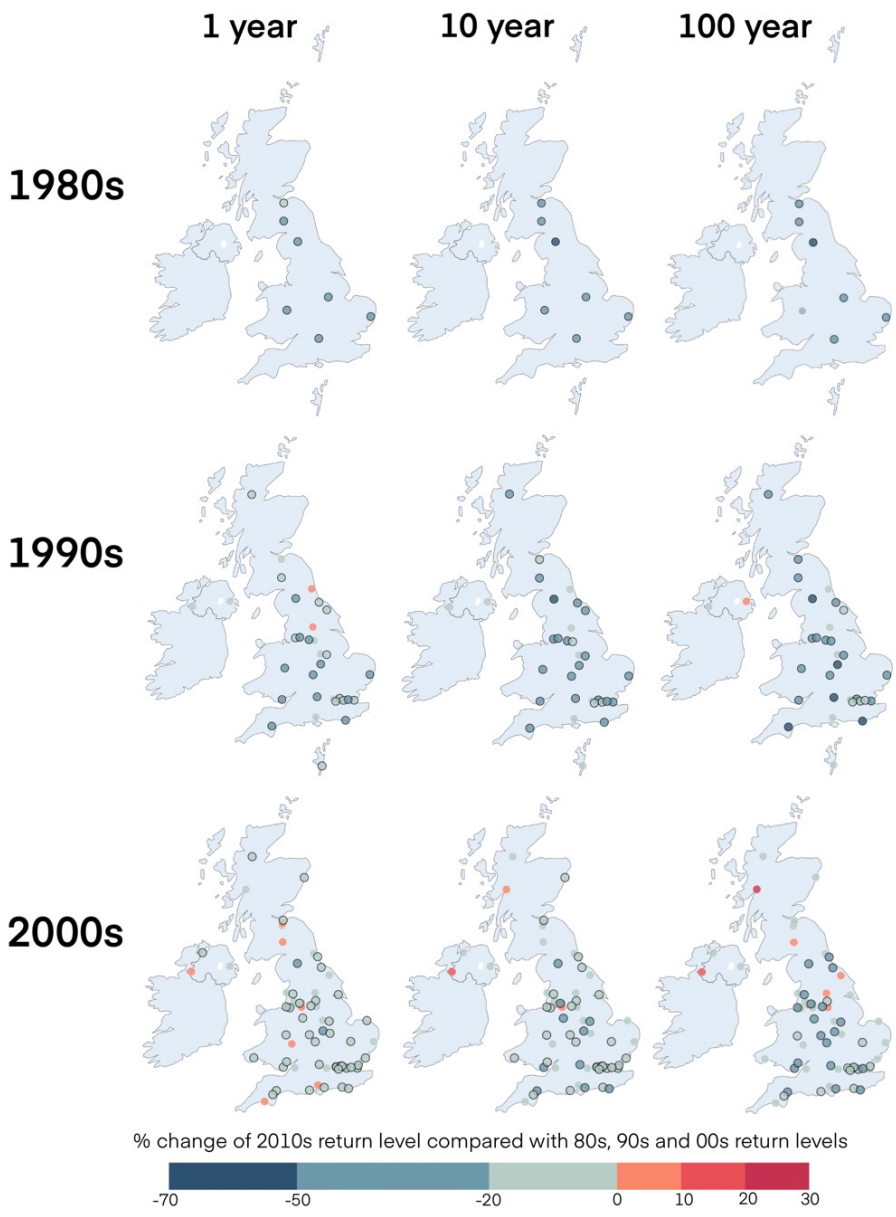

Figure 2.6: The change (%) in the 1-, 10- and 100-year return levels in the 2010s compared with those in the (top) 1980s, (middle) 1990s and (bottom) 2000s for each UK measurement site, found by using the simulation method described in Section 2.3.4. A decrease (or increase) means that a site's return level is lower (or higher) in the 2010s that it was in the compared decade. Sites are outlined in black if the change is significant ($p < 0.05$).

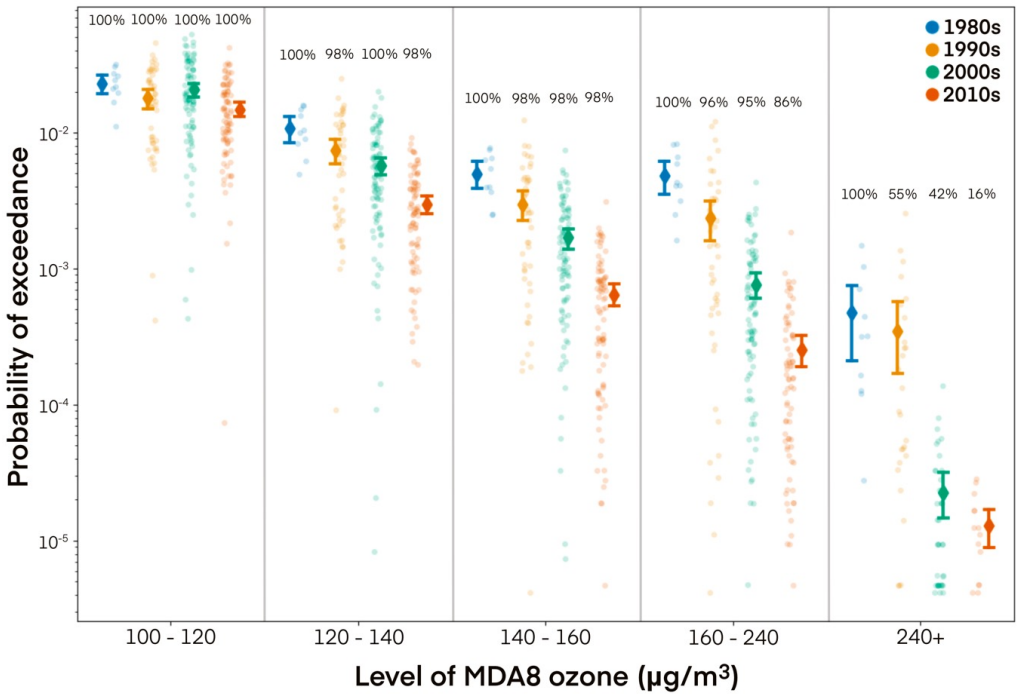

Figure 2.7: The probabilities of exceeding various levels of MDA8 ozone (µg/m^3), grouped by decade. The dots in the background are the individual sites. The diamonds are the all-site mean of each grouping, with the 95% confidence intervals of the mean estimate also shown. The percentage numbers above each group show the proportion of sites which can reach the given level (so where the probability is greater than 0).

For the very high health threshold (240 µg/m^3), again 100% of sites had potential to exceed that level in the 1980s, falling to 55% in the 1990s, 42% in the 2000s and just 16% in the 2010s.

When interpreting the results shown in Figure 2.7 it is important to bear in mind that the number of AURN monitoring sites has increased over time, as shown in Figure A.6. This also highlights a change in the ratio of different site types; specifically, many more urban background sites came online in the 1990s. Consequently, a potential sampling bias may exist due to the different number of operational sites and site types in each of the decades. A more robust assessment of long-term changes is obtained using only the seven individual sites that have been

online since the 1980s for which we find the probability of exceeding 100 µg/m^3 has declined by a factor of 2–6 between the 1980s and present (Table A.3).

Taken together, the above findings provide a novel characterisation of the changing occurrence of extreme ozone events in the UK. Our results broadly corroborate previous studies that have reported a general reduction over time in extreme levels of ozone, albeit those studies employed simpler methods (e.g. Diaz et al., 2020). Using AURN data at multiple sites, Finch and Palmer (2020) fitted a gamma probability distribution to describe changes in UK surface ozone between 1999 and 2019. They found that surface ozone distributions have reduced in skewness over time, determining a reduction in the probability of high ozone events. As noted in Section 2.3.4, the GDP employed in this study provides a very good fit to the tails of the ozone distributions and our EVA approach benefits from its ability to quantify the reduction in the probability of high ozone events.

We anticipate that the reductions in extreme ozone discussed above are due to reduced emissions of ozone precursor pollutants, NO_X, and VOCs, over our study period. Between 1970 and 2019, annual UK emissions of NO_X and non-methane VOCs have fallen by 71% and 66%, respectively (Defra, 2021a; Defra, 2021b). Decreases in precursor pollutants have been associated with a reduction of high level ozone in the United States (Gégo et al., 2007; Bloomer, Vinnikov, et al., 2010; Butler et al., 2011; Owen R. Cooper et al., 2012) and parts of Europe (e.g. Fleming et al., 2018). However, we note also that UK annual average ozone concentrations are observed to be increasing at almost all environment types (Diaz et al., 2020; Finch and Palmer, 2020), despite strong NO_X decreases. Given the complexities of photochemical ozone production and loss, along with the confounding factor of a changing hemispheric baseline ozone concentration (Richard G. Derwent et al., 2018), it is beyond the scope of this work to attribute the ozone changes. However, we note that the observed behaviour (reductions in peak ozone but an increasing mean) is possible in complex photochemical environments characterised by VOC-limited conditions in some areas (e.g. Y. Li et al., 2013; Simon et al., 2015).

2.4.3 The influence of temperature

Due to climate change and the increasing risk of hotter than average summers (Christidis et al., 2015; Fleming et al., 2018; Chapman et al., 2019; Hanlon et al., 2021), it is important to understand the relationship between extreme ozone episodes and extreme temperature. Outside of one study (Finch and Palmer, 2020), which demonstrated that temperature is an important driver of the number of ozone exceedance events in the UK, there is a paucity of UK-focused analysis in the recent literature.

Figure 2.8 characterises some of the important aspects of the extreme ozone/temperature association in our model. Figure 2.8a provides evidence of the positive temperature dependence in the risk of a moderate ozone health threshold exceedance in the 2010s. At the 90^{th} percentile of temperature (see Figure A.7 for the absolute temperature values computed for each site), the probability of exceeding $100\,\mu g/m^3$ ranges from 0.51% to 11.85% for sites where $100\,\mu g/m^3$ is still considered extreme. Increasing the temperature percentile further increases the risk of an exceedance. However, recall from Section 2.3 (and Figure 2.2) that (1) for an ozone concentration to be classified as 'extreme' it must fall above the modelling threshold used in our extreme value model, and (2) that this threshold is computed based on temperature (days with higher temperatures have a higher concentration threshold above which ozone is 'extreme'). For this reason, towards more extreme temperatures (90^{th} percentile and above), a moderate health threshold level of $100\,\mu g/m^3$ is no longer considered to be an extreme concentration at some sites (see also Figure 2.8b). The Midlands, South East, East and South, where the 90^{th} percentile of temperature in the ozone season is mostly >20 °C (Figure A.7), are the regions where this is most evident. However, in the absence of a denser monitoring network, a conclusive analysis of spatial variability across the whole UK is not possible.

Figure 2.8c shows the percentage of sites in the 2010s which have a non-zero probability of exceeding the various ozone health threshold levels for given site-specific temperature percentiles. Due to the limiting behaviour of the ozone extremes

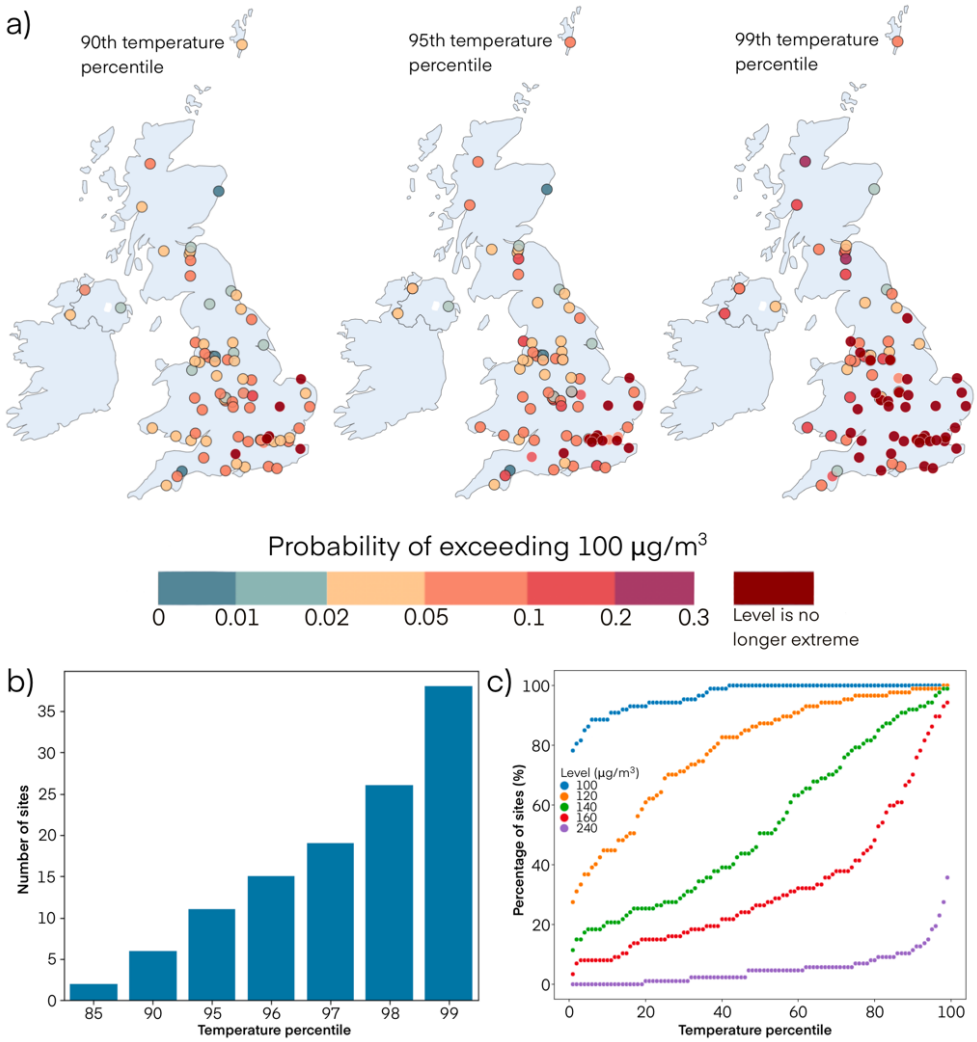

Figure 2.8: (a) The probabilities of exceeding the moderate health threshold MDA8 ozone level of 100 µg/m^3 for each site's 90$^{\text{th}}$ (left), 95$^{\text{th}}$ (middle) and 99$^{\text{th}}$ (right) temperature percentile for each UK monitoring station. Outlined circles indicate sites confidence in estimate of 100 µg/m^3 being an extreme level or not ($p \leq 0.05$). (b) The number of sites where 100 µg/m^3 is no longer considered extreme by the model for the given site-specific temperature percentile, i.e., the number of sites where 100 µg/m^3 is below the extreme modelling threshold for the given temperature percentile. (c) The percentage of sites which have a non-zero probability of exceeding the levels of 100, 120, 140, 160 and 240 µg/m^3 for the given site-specific temperature percentile, ranging from the 1st to 99$^{\text{th}}$. Based on analysis of the decade 2010–2019.

at some sites, not all sites are able to reach high levels of ozone. This is particularly evident for the very high 240 µg/m^3 level, regardless of temperature, with a low percentage of sites able to exceed this level. For the 1st temperature percentile, 78% of sites can exceed a level of 100 µg/m^3, while just 28%, 11%, 3% and 0% of sites can exceed levels of 120, 140, 160 and 240 µg/m^3, respectively, showing a large disproportion in the risk of exceeding these levels at lower temperatures. An exceedance of the levels 100 and 120 µg/m^3 is possible for all sites for temperatures above the 42nd and 98th temperature percentiles, respectively. By the 99th temperature percentile (i.e., the most extreme temperature), 99%, 93% and 33% of sites can exceed levels of 140, 160 and 240 µg/m^3. However, it is important to emphasise that while this analysis demonstrates the positive link between the possibility of an extreme ozone event and extreme temperature, the absolute overall probability (when accounting for the observed temperature distribution) of exceeding the moderate ozone threshold (100 µg/m^3) was, on average, 2.7% for rural sites and 1.6% for urban sites in the decade 2010–2019. Exceedance probabilities for higher thresholds are substantially lower (Section 2.4.1). Interestingly, Figure 2.8c also highlights the considerable difference in the temperature dependent possibility of exceeding 100 µg/m^3 compared with all greater levels.

Our final analysis compares temperature dependent health threshold exceedance risks over decades. Figure 2.9 shows the mean probability of exceeding the high and very high ozone health threshold levels for sites that have been online since the 1990s and continue to monitor in the 2010s. These 30 sites comprise 15 rural background and 15 urban background locations, and the analysis is split by site type. The moderate health threshold is excluded from this analysis as it is not extreme for all sites at higher temperature percentiles (see above). Significant reductions in the probability of exceeding the high ozone health threshold are apparent between the 1990s and 2010s, with the most striking reductions occurring at extreme temperatures. For instance, at the 99th temperature percentile, the mean rural background probability of exceeding the high health threshold level is 0.097

[0.061, 0.132], 0.029 [0.015, 0.043] and 0.006 [0.003, 0.009] for the 1990s, 2000s and 2010s respectively, (Figure 2.9a); the comparable results for urban background sites (Figure 2.9c) are 0.032 [0.011, 0.053], 0.013 [0.004, 0.022] and 0.002 [0, 0.004]. Large reductions are also apparent for the very high health threshold level.

The above analysis shows that from the 1990s to the 2010s, the mean probability of exceeding the high health threshold level decreased by a factor of ∼16 at both rural and urban sites for the most extreme temperatures (99^{th} temperature percentile). These differences cannot be attributed to changes in temperature over time, since for these decades and these sites, the temperature distributions are similar (Figure A.8). We surmise that a reduction in the probability of extreme ozone across the decades, and the sensitivity of such to temperature, to be due to reductions in precursor pollutants of ozone, as outlined in Section 2.4.2. Other recent analysis of the so-called ozone 'climate penalty factor' also shows the sensitivity of ozone to temperature has decreased at a large number of sites in Germany, likely in response to reductions in NO_X and VOCs (Noelia Otero, Rust, et al., 2021). Overall, our findings on the temperature dependence of extreme UK ozone extend and substantiate previous studies (J. D. Lee et al., 2006; Finch and Palmer, 2020), namely that the occurrence of high level ozone has a strong temperature dependence.

2.5 Conclusions

Characterisation of extreme ozone events, along with their association with temperature, is important to better understand the risks of air pollution to human health and how air pollution is managed, in a changing climate. Here, we have presented a thorough, statistically grounded analysis of extreme surface ozone in the UK using extreme value analysis. Our analysis considered daily maximum 8-hr running mean ozone (MDA8), obtained from hourly ozone measurements, at >100 monitoring sites of the UK's AURN network, along with gridded daily maximum temperature data from the UK Met Office. We focused analysis on the 'ozone season', defined for

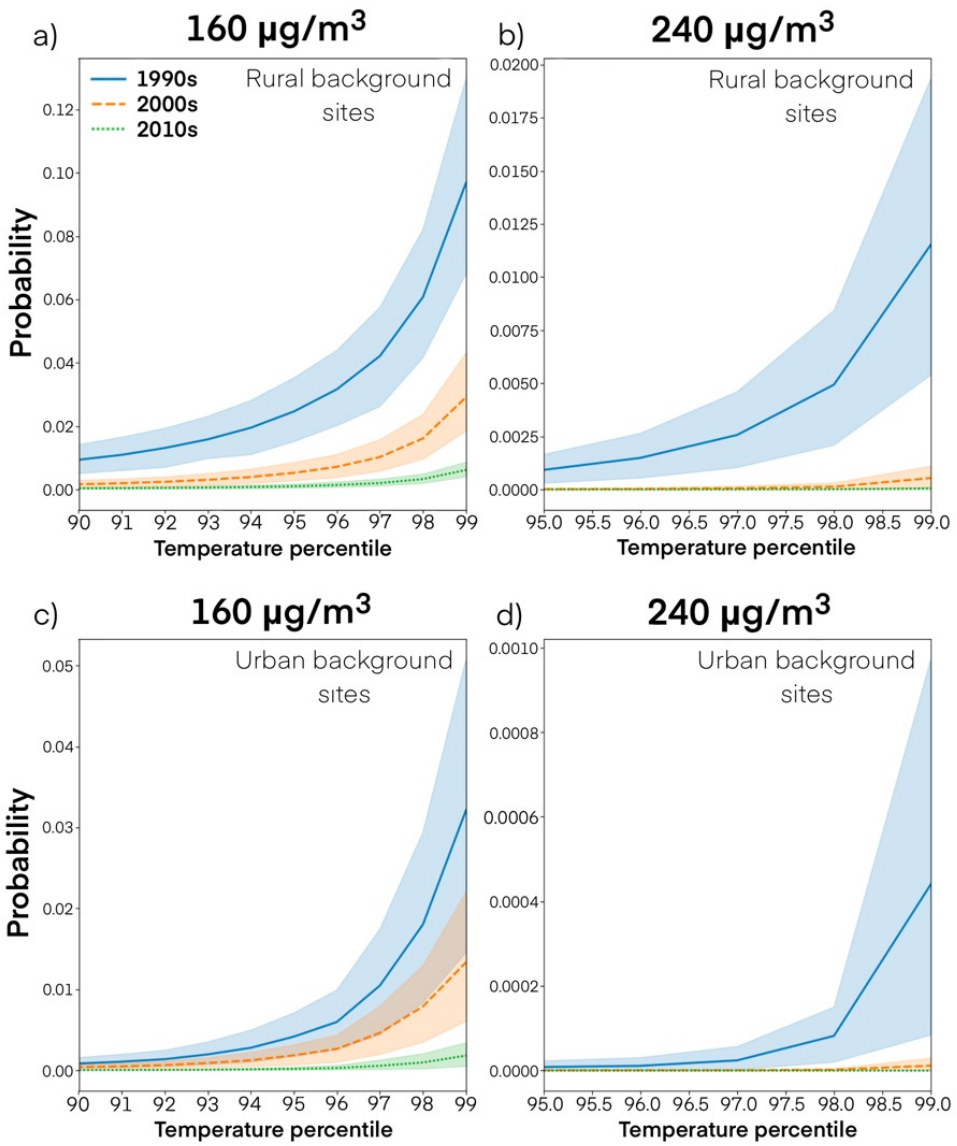

Figure 2.9: The mean probability of exceeding the high (160 µg/m^3; panels (a) and (c)) and very high (240 µg/m^3; panels (b) and (d)) health threshold levels of MDA8 ozone for rural (top) and urban background (bottom) sites, dependent on temperature and compared across different decades, denoted by colour: blue for the 1990s, orange for the 2000s and green for the 2010s. The analysis considers 15 rural sites and 15 urban sites that have been monitoring ozone from the 1990s through to the 2010s. The 95% confidence intervals of the mean estimate are shown as shaded bands for each decade.

individual sites (roughly March–September), consisting of the months containing virtually all occurrences of MDA8 $> 240\,\mu g/m^3$, i.e., the period most relevant for short-term exposure to extreme ozone and human health.

From our analysis of data from the most recent decade (2010–2019), we find that at 93% of UK monitoring sites, MDA8 ozone is predicted to exceed a 'moderate' health threshold level ($100\,\mu g/m^3$) at least once in one year. The overall probability of a $100\,\mu g/m^3$ exceedance on a given day is estimated to be between <0.1% and 5.4%, depending on site, with an average of 2.7% at rural sites and 1.6% at urban sites. Exceedances of the 'high' ($160\,\mu g/m^3$) and 'very high' ($240\,\mu g/m^3$) thresholds, defined by the UK's air quality index, are plausible but extremely rare events at the sites considered and are thus expected to be less relevant for exposure assessment.

Our results show statistically significant changes in the extremal characteristics of ozone over the last three decades. Considering sites with the longest records only, 1-year return levels of MDA8 ozone in the 1980s are comparable to 10-year return levels seen in the 2010s. Thus, the magnitude of extreme ozone events has decreased over time, likely reflecting the large reductions in ozone precursor emissions in this period. At these sites, we estimate the probability of MDA8 ozone exceeding $100\,\mu g/m^3$ on a given day has decreased by a factor of 2–6 between the 1980s and present. Across all monitoring sites, return periods for the moderate health threshold have increased at 88% of sites since the 1990s and 90% of sites since the 2000s. This means that most sites are expected to reach or exceed $100\,\mu g/m^3$ less often now than they did in the 1990s and 2000s. Only 16% of sites can theoretically reach the very high health threshold level in the 2010s, compared with 100% in the 1980s, 55% in the 1990s and 42% in the 2000s, thus at most sites the high end of the plausible range of extreme ozone levels has decreased significantly.

As elevated temperature is associated with extreme ozone concentrations in Europe (e.g. Otero et al., 2016), temperature is incorporated in the threshold and scale parameter of our extreme value model. This enables us to model the effect from temperature on ozone extremes. We find that higher temperatures increase

the risk of a moderate health threshold exceedance across the UK. However, we also find that above the 90th temperature percentile, an exceedance of 100 µg/m^3 is not considered an 'extreme' event at a number of sites. This is important as health threshold exceedances should be rare events. The temperature-dependent health threshold exceedance risk has changed over the decades. Where there has been monitoring since the 1990s, we find that rural background sites have the greatest absolute decrease in the risk of a high health threshold exceedance for the 99th temperature percentile, highlighting a dynamic relationship between extreme ozone and extreme temperature over our study period.

In conclusion, in the absence of continuing decreases in ozone precursor emissions, expected increases in the frequency of extreme temperatures (Lowe et al., 2018; Hanlon et al., 2021) may offset gains made in the reduction of extreme ozone over time. Further work is now required to assess the response of UK extreme ozone concentrations to the combined influence of anticipated future reductions in ozone precursors, both locally and at the continental-scale, alongside rising temperature.

2.6 Acknowledgements

LG acknowledges the UK Engineering and Physical Research Council (EPSRC) for a PhD studentship. RH was supported by a UK Natural Environment Research Council (NERC) Independent Research Fellowship (NE/N014375/1). EE and PJY were supported by the EPSRC-funded Data Science of the Natural Environment project (EP/R01860X/1).

Chapter 3

A machine learning approach to downscale EMEP4UK: analysis of UK ozone variability and trends

Lily Gouldsbrough[1], Ryan Hossaini[1,2], Emma Eastoe[3], Paul J. Young[1,4], Massimo Vieno[5]

[1]Lancaster Environment Centre, Lancaster University, Lancaster, UK
[2]Centre of Excellence in Environmental Data Science, Lancaster University, Lancaster, UK
[3]Department of Mathematics and Statistics, Lancaster University, Lancaster, UK
[4]JBA Risk Management Limited, Broughton Park, Skipton, BD23 3FD, UK
[5]UK Centre for Ecology & Hydrology, Bush Estate, Penicuik, Midlothian EH26 0QB, UK

Corresponding authors: Lily Gouldsbrough (lilygouldsbrough@outlook.com) and Ryan Hossaini (r.hossaini@lancaster.ac.uk)

The following work was submitted to Atmospheric, Chemistry and Physics (ACP) on 31st March 2023. The authors contributions are listed are the next page.

Statement of contribution. Lily Gouldsbrough led the methodology, data curation and analysis and conceived the study alongside Ryan Hossaini, Emma Eastoe. Lily Gouldsbrough drafted the manuscript, with guidance from Ryan Hossaini, Emma Eastoe, Paul J. Young and Massimo Vieno.

Abstract

High-resolution modelling of surface ozone is an essential step in the quantification of the impacts on health and ecosystems from historic and future concentrations. It also provides a principled way in which to extend analysis beyond measurement locations. Often, such modelling uses relatively coarse resolution chemistry transport models (CTMs), which exhibit biases when compared to measurements. EMEP4UK is a CTM that is used extensively to inform UK air quality policy, including the effects on ozone from mitigation of its precursors. Our evaluation of EMEP4UK for the years 2001–2018 finds a high bias in reproducing daily maximum 8-hr average ozone (MDA8), due in part to the coarse spatial resolution. We present a machine learning downscaling methodology to downscale EMEP4UK ozone output from a 5×5 km to 1×1 km resolution using a gradient boosted tree. By addressing the high bias present in EMEP4UK, the downscaled surface better represents the measured data, with a 128% improvement in R^2 and 37% reduction in RMSE. Our analysis of the downscaled surface shows a decreasing trend in annual and March–August mean MDA8 ozone for all regions of the UK between 2001–2018, differing from increasing measurement trends in some regions. We find the proportion of the UK which fails the government objective to have at most 10 exceedances of 100 µg/m^3 per annum is 27% (2014–2018 average), compared to 99% from the unadjusted EMEP4UK model. A statistically significant trend in this proportion of -2.19%/year is found from the downscaled surface only, highlighting the importance of bias correction in the assessment of policy metrics. Finally, we use the downscaling approach to examine the sensitivity of UK surface ozone to reductions in UK terrestrial NO_X (i.e., NO + NO_2) emissions on a 1×1 km surface. Moderate NO_X emission reductions with respect to present day (20% or 40%) increase both average and high-level ozone concentrations in large portions of the UK, whereas larger NO_X reductions (80%) cause a similarly wide-spread decrease in high-level ozone. In all three scenarios, very urban areas (i.e., major cities) are the most affected by increasing concentrations of ozone, emphasising the broader air quality challenges of NO_X control.

3.1 Introduction

Ground level ozone is a harmful air pollutant that causes respiratory health issues, hospitalisation, and in severe cases, mortality (Ji et al., 2011; COMEAP, 2015; Nuvolone et al., 2018), with an estimated 254,000 deaths globally in 2015 due to elevated ozone exposure (Cohen et al., 2017). Ozone is a secondary pollutant formed via chemical reactions of precursor pollutants – nitrogen oxides (NO and NO_2, known as NO_X) and volatile organic compounds (VOCs) – in the presence of sunlight. Because of the harmful consequences of increased ozone levels, air quality standards of varying degrees of stringency have been set in many countries for both ozone and its precursors. Compliance with, and the efficacy of, these standards is primarily assessed through analysis of surface ozone measurements. Whilst useful, the empirical analysis of ozone measurements is limited in scope by the density and location of monitoring sites and length of data records (e.g. Lang, 2020). Mathematical models, either process-based air quality models or statistical, are therefore useful in further improving our understanding of ozone formation, long-term trends, and spatial variability, at spatial and temporal scales that measurements alone cannot match. However, it is well established that ozone concentrations vary spatially, seasonally and temporally due to meteorological conditions and precursor availability and reactivity (O. R. Cooper et al., 2014; Pope et al., 2016), making it a particularly challenging pollutant to model.

Recent statistical analysis of UK surface ozone has centred on data from measurement stations (Diaz et al., 2020; Gouldsbrough et al., 2022). However, these stations are not spread uniformly across the UK, leaving substantial portions of the country unmonitored (Finch and Palmer, 2020). In consequence, robust estimates of regional variability in ozone concentration and trends are unavailable, though it is known that both vary significantly across sites and site type. Taking the UK as a whole, a notable finding of the above studies is that the probability of the occurrence of extreme ozone concentrations has reduced in recent decades, but reported trends are generally not statistically significant (e.g. Finch and Palmer, 2020). Chemistry

transport models (CTMs), numerical models that simulate the various processes that affect pollutant concentrations (emissions, chemistry transport, deposition, etc.) are routinely used to produce spatial surfaces of ozone, and other air pollutants, with a far greater spatial coverage than can be achieved by the monitoring stations. Still, these are run on a grid, often at a coarse resolution with respect to what is optimal for exposure assessment, and thus may not capture local-scale behaviour. High resolution multi-annual CTM simulations are also computationally expensive and, like all process models, subject to a degree of bias (e.g. Z. Liu et al., 2022). An alternative to high-resolution process modelling is to develop a statistical model to interpolate the ozone measurements (Wong et al., 2004; Hooyberghs et al., 2006). However, given the complexity of the processes that underpin ozone production, downscaling approaches that use measured data, where it is available, to remove local-scale bias from the numerical model output are attractive.

Data-driven downscaling methods are used to model complex physical systems by combining information from ground observations or satellites with information from process-based numerical models. Previous work to downscale numerical model surface ozone concentrations includes dynamical, statistical and machine learning downscaling. Dynamical downscaling uses high resolution regional simulations to extrapolate the effects of large scale processes to local scales, and has been applied in the US (Trail et al., 2013; J. Sun et al., 2015; Nolte et al., 2021) and Belgium (Lauwaet et al., 2013). Statistical downscaling of surface ozone has been performed using regression (Guillas et al., 2008; Bravo et al., 2016; Gauthier-Manuel et al., 2022), fitted empirical orthogonal functions (Alkuwari et al., 2013), and a spectral method (Reich, Chang, et al., 2014). Machine learning (ML) models have also been used to downscale surface ozone. For example, a Bayesian ensemble machine learning model that integrates 13 learning algorithms has been used to create a census tract level daily maximum 8-hr average ozone (MDA8) ozone surface for the US, demonstrated for 2011 (Ren et al., 2022). Global downscaled ozone surfaces have also been created using ML models, including a Bayesian neural network model to

create a 10×10 km ozone surface for 1990–2019 (H. Sun et al., 2022), and a random forest model to create a 0.1°×0.1°average ozone surface for 2010–2014 (Betancourt et al., 2022). To our knowledge, no studies have applied downscaling specifically to UK ozone. In this paper, we develop and evaluate a novel ML-based methodology for downscaling the EMEP4UK CTM from a 5×5 km to 1×1 km resolution. The CTM is developed as the UK regional application of the European Monitoring and Evaluation Program (EMEP) and has been widely used to study UK air quality and to inform policy decisions (e.g. Vieno, Heal, et al., 2016). Previous EMEP4UK evaluation shows the model generally performs well at reproducing observations of a range of pollutants, though notably ozone exhibits a non-negligible positive bias at almost all sites. The bias is larger at urban background locations, possibly reflecting a dilution of local NO_X emissions on the 5×5 km grid and, thus, insufficient NO_X titration (C. Lin et al., 2017). We apply the downscaled model to study UK surface ozone over an 18-year period (2001–2018), focusing on (1) quantifying regional variability in ozone, trends, and policy-related metrics; (2) comparing conclusions drawn from the downscaled surface relative to those from the unadjusted CTM and measurements alone; and (3) exploring the sensitivity of UK ozone concentrations to reductions in NO_X. Of the many ML tools available, we use a gradient boosted tree (GBT) (Friedman, 2001). Our choice is primarily driven by the fact that the GBT model can learn non-linear relationships from highly dimensional datasets. Thus, the GBT model allows us to use several measurement stations and covariates in the downscaling model, and thereby model ozone as the product of multiple highly non-linear and interacting systems. Moreover, GBTs have been used to downscaled numerical model surface ozone for China (Hu et al., 2022; R. Liu et al., 2020), and were found useful in predicting surface level ozone during wildfires in California (Watson et al., 2019).

This chapter is structured as follows. Section 3.2 outlines the data used in the analysis and the features included with the ML model. Section 3.3 describes the downscaling methodology, including an evaluation of the downscaled surface. In

Section 3.4 we analyse the downscale surface, focusing on regional ozone variability across the UK in recent years (Section 3.4.1), on longer-term trends and interannual variability (3.4.2), and on UK ozone-NO_X sensitivity (3.4.3). Across these results we compare the conclusions drawn from the new downscaled surface to those from the unadjusted EMEP4UK output and measurements alone. Finally, in Section 3.5 we present our conclusions and some brief recommendations for future research.

3.2 Data

Accurate downscaling requires not just the measured and modelled ozone data, but also information on variables which affect the net production or transport of ozone at the Earth's surface. We make use of information on the meteorology, climate and geophysical characteristics at a given location. For consistency with ML terminology, we refer to the individual variables as 'input features', a concise summary of which can be found in Table 3.1. All datasets cover the period 2001–2018 (inclusive).

3.2.1 Modelled ozone from EMEP4UK

EMEP4UK is a UK focused version of the EMEP MSC-W model (https://www.emep.int), an Eulerian CTM used to assess concentrations and deposition of various air pollutants across Europe (Simpson et al., 2012). Various studies related to UK ozone have been performed using EMEP4UK: quantifying the burden of heat and ozone on mortality (Doherty et al., 2009), modelling ozone during the 2003 heatwave (Vieno, Dore, et al., 2010), modelling the effect of climate change on ozone health impacts (Vardoulakis and Heaviside, 2012), quantifying the socioeconomic and urban-rural differentials in exposure (Milojevic et al., 2017), and modelling air pollution exposure in relation to workplace mobility (Liška, 2021). The EMEP4UK model is also used to inform policy decisions concerning air quality: the extent to which UK source abatement measures can mitigate UK particulate matter concentrations (AQEG et al., 2013; Carnell et al., 2019); the impact of reductions

in UK anthropogenic emissions on various pollutants (Vieno, Heal, et al., 2016); the impacts of climate change and mitigation options for agriculture, forestry, land use and waste sectors (SRUC, 2017); the effect of changes in vegetation coverage on air pollution (EIDC, 2021); and to quantify the spatial variation in average ozone across the UK, including the calculation of population-weighted ozone exposure during workdays, long term exposure, and the implication of a 2030 emissions scenario on surface ozone concentrations (AQEG, 2021).

Previous evaluation of the EMEP4UK model quantified its performance in the reproduction of the 10-year mean measured ozone concentrations for 2001–2010. Considering 17 rural and 30 urban sites, R^2 values of 0.21 (0.81 when removing one erroneous rural site) and 0.73, respectively, were obtained (C. Lin et al., 2017). A positive model bias for ozone at urban background sites is due to the dilution of urban NO_X emissions at the model 5×5 km resolution, meaning that EMEP4UK insufficiently captures the urban NO_X titration of ozone. In our preliminary investigations, we found that the original EMEP4UK output fails to capture the behaviour of MDA8 ozone in the larger sample of 198 measurement stations used in this analysis, with a cross-year mean R^2 of 0.32, and cross-year mean RMSE of 19.4 µg/m^3. Therefore, we determined it was necessary to develop a methodology with which to downscale the EMEP4UK model output to a higher resolution spatial grid and to address the above bias.

The dataset we wished to downscale is a 5×5 km gridded dataset of hourly surface ozone concentrations from EMEP4UK covering the period 2001–2018. To produce this ozone field, the offline CTM was run with meteorology from the Weather Research and Forecast (WRF) model version 3.7.1 (C. Skamarock et al., 2008) between 2001–2017 and WRF4.1 (W. C. Skamarock et al., 2019) for 2018. The WRF simulation in this work assimilates data from the numerical weather prediction model meteorological reanalysis of the US National Center for Environmental Prediction (NCEP) / National Center for Atmospheric Research (NCAR) Global Forecast System (GFS) (Environmental Prediction, 2000). MDA8 ozone concentrations were

calculated from the original hourly model output and linear interpolation was used to convert the 5×5 km field to a 1×1 km grid over the entire study period (2001–2018).

In addition to the output from the main transient model run described above, a series of additional model experiments were performed for the year 2018 to explore the sensitivity of UK ozone to NO_X reductions. These included a 2018 reference run with terrestrial UK NO_2 emissions of 743 Gg/yr and three otherwise identical runs with NO_2 emissions reduced by 20% (594 Gg/yr NO_2), 40% (446 Gg/yr), and 80% (148 Gg/yr). These reductions do not correspond to any specific future scenarios and are designed solely as a sensitivity analysis on the basis of the ongoing long-term decline in UK NO_X emissions. These emissions fell by 76% between 1970 (2920 Gg/yr) and 2020 (702 Gg/yr) (Defra, 2021a), with the expectation of further reductions in the future.

3.2.2 Meteorological variables

Surface ozone levels are strongly influenced by local and synoptic weather conditions (Pope et al., 2016) and meteorological variables are thus common input features in ML studies of ozone. The WRF model is a weather prediction system designed for atmospheric forecasting (Grell et al., 2005). For this study, WRF version 3.7.1 meteorological and terrain variables for the years 2001–2018 were collected on the same 5×5 km grid as the EMEP4UK model, and then also linearly interpolated to 1×1 km. Previous work found daily maximum temperature, relative humidity, thermal surface radiation and wind speed to be important drivers of MDA8 ozone in Europe (Otero et al., 2016). Thus, we include these (and other similarly relevant) meteorological variables in our ML model (Table 3.1).

3.2.3 Distance variables

Due to the interaction between NO_X and ozone, distance to the nearest road is a key explanatory variable (Granier and Brasseur, 2003). Distances to five road types from major to minor roads (Meijer et al., 2018) were calculated at each ozone

measurement station for the calibration of the model, and on a 1×1 km grid for the predicted downscaled surface. Similarly, distance to coast from a shapefile of the UK coastline (Natural Earth: https://www.naturalearthdata.com) was used to account for the increase of ozone concentrations in coastal areas (Entwistle et al., 1997). Note that we do not include NO_X itself, or any other precursor pollutants, directly in the model. The primary reason for this is that the EMEP4UK output for such chemicals is likely to be biased, and we wish to avoid this bias propagating into the downscaled ozone field. Instead, the distance to road variable acts as a proxy for NO_X concentration, which is reasonable given the importance of road transport NO_X emissions in the UK. To lessen the bias of the downscaling model towards the relatively dense measurement network in London, an indicator variable was included to delineate between inside London and outside London, with London defined using a bounding box from 0.489°W–0.236°E and 51.28–51.686°N.

3.2.4 Ozone monitoring network data

Surface ozone measurements for the years 2001–2018 were obtained from the Automatic Urban and Rural Network (AURN: https://uk-air.defra.gov.uk, 108 sites), Kings College London network (KCL: https://www.londonair.org.uk, 68 sites), Air Quality England network (AQE: https://www.airqualityengland.co.uk, 12 sites), Welsh Air Quality Network (WAQN: https://airquality.gov.wales, 9 sites) and Scottish Air Quality Network (SAQN: https://www.scottishairquality.scot, 1 site). These measurements are essential for the calibration and evaluation of the downscaling model. MDA8 ozone concentrations were calculated at each of the 198 measurement sites; for locations, see Figure B.1. There are differences in the observation period across the sites, but all sites had a minimum of 3 years data.

In our subsequent analysis, we explored regional variations in ozone concentration and trends. We considered 12 UK regions (see Figure B.2), the spatial definitions of which are taken from the Level 1 Nomenclature of Territorial Units for Statistics (National Statistics, 2018). Estimation of site-wise trends, including

the magnitude and significance, may be sensitive to the chosen statistical technique. Approaches used in previous studies include use of the non-parametric Theil-Sen method applied to deseasonalised monthly mean ozone time series (AQEG, 2021) and least square fits to annual mean data (Finch and Palmer, 2020). In this study, we calculated trends using ordinary least squares on the yearly averages (and seasonal averages, 90^{th} percentiles, 10^{th} percentiles) of MDA8 ozone. All trends are calculated using all available data from each region.

3.3 Downscaling methodology

Our goal is to produce a gridded downscaled surface ozone product which better represents the stochastic behaviour of the measurements than the original EMEP4UK output alone. The downscaling approach consists of five steps. First, the 5×5 km gridded model ozone surface is linearly interpolated to a 1×1 km resolution. Second, a matched dataset of modelled ozone is selected from the 1×1 km EMEP4UK surface by selecting, for each measurement station, the nearest grid cell. Third, a machine learning model is used to perform bias correction on the modelled ozone data. Fourth, the performance of the bias correction is evaluated at the measurement locations, by a comparison of the predicted ozone with the observed measurements. Steps three and four are iterated until no further improvements in the predictive capability of the model can be seen. Finally, the trained machine learning model is used to predict MDA8 surface ozone on a 1×1 km resolution grid for the UK. Figure 3.1 shows an example of the resulting downscaled surface. This surface consists of 234,187 cells, compared to the 10,941 cells of the original EMEP4UK surface. The increased resolution in the downscaled surface leads to greater local-level detail, resulting in improved inference on the probabilistic behaviour of the ozone surface which we demonstrate in our subsequent analysis. Further details on the specific machine learning model, and how it was tuned, are as follows.

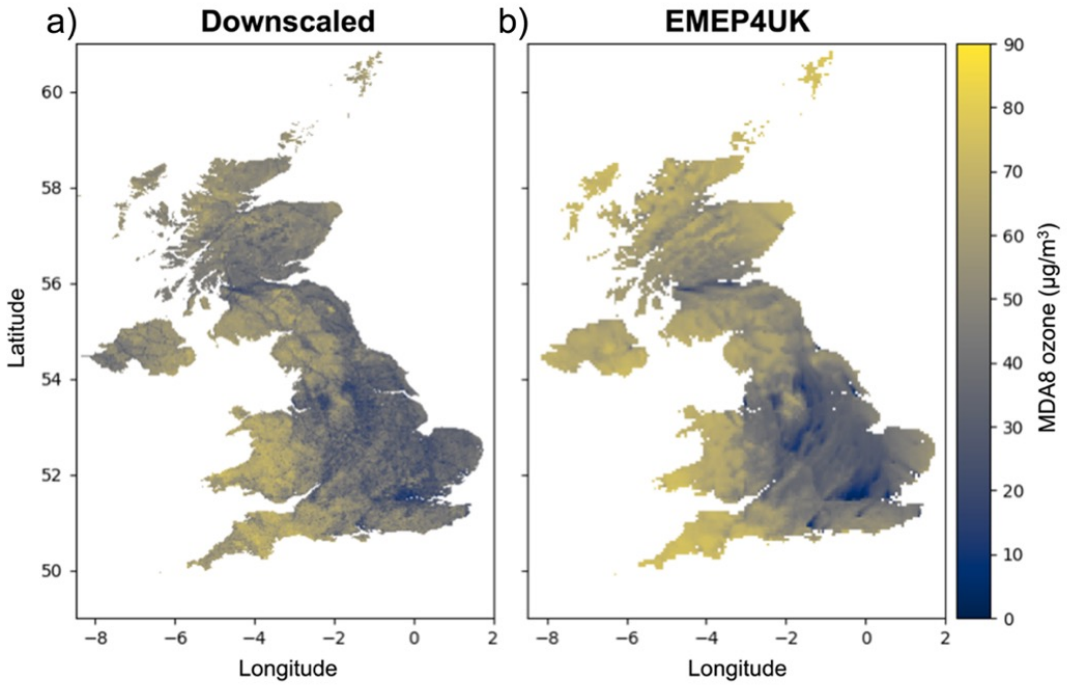

Figure 3.1: (a) An example of the downscaled MDA8 (µg/m^3) surface (1×1 km resolution) and (b) original EMEP4UK surface (right, 5×5 km) for 01-01-2008.

Table 3.1: Input features to the ML model.

Input features	Source	Abbreviation in source	Resolution
EMEP4UK surface ozone	EMEP4UK	O3	1×1 km
Latitude	Measurements	–	–
Longitude	Measurements	–	–
Daily maximum 2m temperature	WRF	T2	1×1 km
Daily minimum 2m temperature	WRF	T2	1×1 km
Daily mean surface pressure	WRF	PSFC	1×1 km
Downward short-wave flux at ground surface	WRF	SWDOWN	1×1 km
Daily mean planet boundary layer height	WRF	PBLH	1×1 km
Daily mean surface vapor	WRF	QVAPOR	1×1 km
Daily mean x component of wind	WRF	U10	1×1 km
Daily mean y component of wind	WRF	V10	1×1 km
Terrain height	WRF	HGT	1×1 km
Distance to highways	GRIP	1	Vector
Distance to primary roads	GRIP	2	Vector
Distance to secondary roads	GRIP	3	Vector
Distance to tertiary roads	GRIP	4	Vector
Distance to local roads	GRIP	5	Vector
Distance to coast	Natural Earth	–	Vector
Year (as integer)	EMEP4UK	–	–
Month (as integer)	EMEP4UK	–	–
Date (as integer)	EMEP4UK	–	–
London (or not)	Bounding box	–	–

3.3.1 Machine learning model

A gradient boosting tree (GBT) is an iterative, supervised, machine learning model, consisting of a parameterised ensemble of decision trees (Friedman, 2001). These decision trees are trained sequentially; each additional tree minimises the prediction error from the previous tree using gradient descent. As a supervised learning algorithm, training the model requires both a training dataset and a predefined objective function, the latter consisting of a loss function and a regularisation term. The training dataset is the subset of the full set of data from which the best fitting model is found. The loss function and regularisation term quantify the quality of model fit given the complexity of the model. The error for each decision tree is calculated from the loss function (Friedman, 2001), which measures how well the model predicts the training data, whilst the regularisation term penalises against model complexity to prevent overfitting. The fitting algorithm ends when either a predetermined number of trees have been fitted, the loss function falls below a predetermined threshold, or the addition of more trees provides no significant improvement to the model fit. The latter criterion is determined by an external validation dataset. The final model is then the summation over the entire ensemble.

Several characteristics of GBTs make them suitable for downscaling: they can capture non-linear relationships between variables far more effectively than competitor approaches (e.g. statistical regression models) and they are both computationally efficient and scalable, i.e., are suitable for large datasets (Chen and Guestrin, 2016). The specific GBT implementation used for this analysis is XGBoost, a highly optimised Python package (Chen and Guestrin, 2016). Since the measurement ozone data is long tailed, we chose a gamma regression for objective and evaluation functions of the GBT model; gamma regression is suited to modelling continuous, non-negative and long tailed data. Like many ML models, the fitting process used to train the GBTs prioritises the fit of the mean behaviour at the expense of characterising the tails (i.e., largest and smallest observations). To reduce this inequality and reduce the mean bias, the tails of the measurement data –

data above (and below) high (and low) concentrations of ozone – were oversampled. Resampling is a common approach to rebalance the distribution of training data for a ML model when the goal is to forecast rare values of the target variable (Torgo et al., 2015).

Lastly, as with any machine learning analysis, we require a balance between the fit of the model on the training data and the ability to apply the model to unseen data, i.e., making sure that the model is not overfit or underfit to the training data. To get satisfactory results, the hyperparameters of the XGBoost model needed substantial tuning. Initial hyperparameters were found using Hyperopt, a Bayesian optimisation package (Bergstra et al., 2013). Subsequent fine tuning was performed manually until no further improvement could be found in the cross-validation (CV) tests.

3.3.2 Evaluation of downscaling

3.3.2.1 Predicting ozone at measurement locations

Evaluation of the downscaling model used CV to assess the prediction of MDA8 ozone across multiple measurement locations. CV requires the model to be trained on a random subsample of the whole dataset, and the resulting model then used to predict the remaining, and previously unseen, data. The first CV test is to split the data into two random samples, selected from the entire dataset: 70% of the data is used to train the model and the remaining 30% is used for evaluation. Table 3.2 shows the annual R^2 and RMSE for the predictions combined across all sites. We found a good agreement between predicted ozone and measurement ozone, with a cross-year mean R^2 of 0.80 and RMSE of 10.61 µg/m^3 for 2001–2018, and no evidence of substantial between-year variation in performance. The second CV test assesses predictive performance at locations which are completely excluded from the model training. To do this, 10-fold CV was applied, with the measurement stations randomly split into 10 groups and each group used to evaluate the model trained on the remaining 9 groups. Again, we found a good agreement between predicted and

measurement ozone, with a cross-year mean R^2 of 0.70, as seen in Table 3.2.

Table 3.2: R^2 and RMSE ($\mu g/m^3$) results of predicted MDA8 ozone vs MDA8 measurements for the two cross-validation tests: 70/30 train/test split and 10-fold CV.

	70/30 train/test split		**10-fold CV**	
Year	**R^2**	**RMSE**	**R^2**	**RMSE**
2001	0.82	11.33	0.71	14.09
2002	0.79	11.36	0.70	13.72
2003	0.82	12.88	0.74	15.52
2004	0.79	11.42	0.69	13.82
2005	0.77	11.50	0.66	13.98
2006	0.84	11.66	0.74	14.72
2007	0.82	10.09	0.72	12.52
2008	0.83	10.56	0.74	12.89
2009	0.81	10.23	0.71	12.50
2010	0.79	10.43	0.68	12.87
2011	0.79	10.43	0.67	13.06
2012	0.79	10.18	0.68	12.61
2013	0.81	10.09	0.71	12.52
2014	0.77	10.22	0.65	12.56
2015	0.77	9.66	0.65	12.05
2016	0.80	9.88	0.73	11.63
2017	0.81	9.29	0.69	11.80
2018	0.83	9.81	0.73	12.42
Mean	**0.80**	**10.61**	**0.70**	**13.07**

3.3.2.2 Downscaled surface vs measurements

Having verified the accuracy of the ML downscaling model, we created the downscaled surface by training the ML model on all measurement data. To compare

the downscaled surface to both measurements and the unadjusted EMEP4UK surface, the data from the cell nearest to each measurement station was extracted. Due to the different grid specifications, the matched cells were not concentric, but they are very close.

Table 3.3 shows the R^2 for each year for the downscaled surface and original EMEP4UK surface when compared to the measurement data. The cross-year mean R^2 of the downscaled surface is 0.73, 128% higher than the equivalent for the unadjusted EMEP4UK, at 0.32. Notably, the extremely low R^2 values for 2014–2016 in the latter are significantly improved in the downscaled surface, e.g. 0.06 vs 0.73 in 2016. Notable improvement is also seen in 2003, the best performing year for EMEP4UK, with a downscaled R^2 of 0.81 compared with 0.61 for unadjusted EMEP4UK. There is also a 37% reduction in the cross-year mean RMSE for the downscaled surface compared with the unadjusted product: 12.26 µg/m^3 vs 19.40 µg/m^3.

Further differences in the capabilities of the two gridded products to represent surface ozone measurements are shown in Figure 3.2. A considerable reduction in noise can be seen in the scatter density plots of the downscaled surface vs measurements (panel a), which has a stronger linear signal and less scatter in comparison to the equivalent plot for EMEP4UK (panel b). This indicates that the unadjusted EMEP4UK surface is a less accurate representation of the measurement data. The bias in the stochastic behaviour of the unadjusted EMEP4UK output is evident when comparing the percentiles of the measurement data to the percentiles of the (i) unadjusted and (ii) downscaled EMEP4UK (panel c); the percentiles for the unadjusted output are consistently higher than those of the measurements, whilst those of the downscaled data are almost identical to those of the measurements. Similarly, we see a considerable shift in the density of the EMEP4UK percentiles compared to that of the measurement data (panel d); again, there is a far smaller discrepancy between the densities of the measurements and the downscaled data.

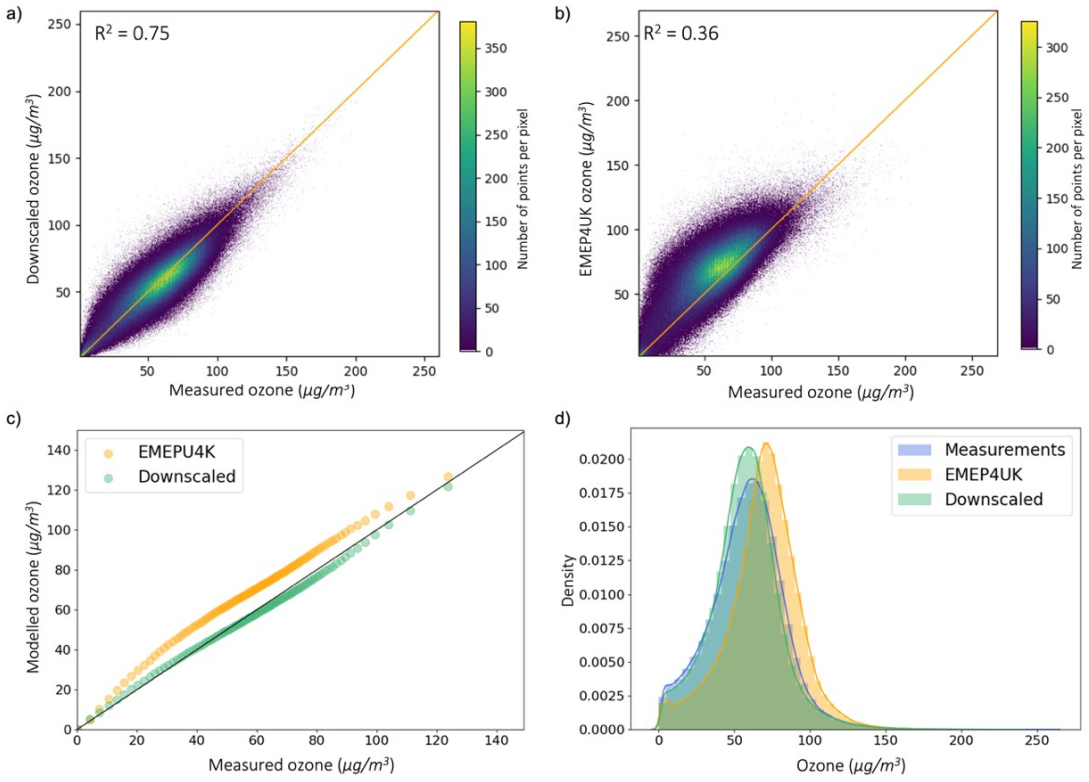

Figure 3.2: a) scatter density plot of downscaled surface vs measurement ozone; b) scatter density plot of EMEP4UK surface vs measurement ozone; c) percentile-percentile plot comparing ordered percentiles of downscaled vs measurement ozone (green) and EMEP4UK vs measurement ozone (orange); and d) density plots of measurement ozone (blue), corresponding downscaled surface ozone (green) and corresponding EMEP4UK surface ozone (orange).

Table 3.3: R^2 and RMSE ($\mu g/m^3$) results of predicted MDA8 ozone vs MDA8 measurements for the two cross-validation tests: 70/30 train/test split and 10-fold CV.

Year	**Downscaled R^2**	**Downscaled RMSE**	**EMEP4UK R^2**	**EMEP4UK RMSE**
2001	0.79	12.10	0.45	19.52
2002	0.77	12.05	0.30	20.91
2003	0.81	13.10	0.61	19.05
2004	0.76	12.25	0.47	18.01
2005	0.73	12.54	0.36	19.22
2006	0.80	13.13	0.54	19.81
2007	0.76	11.64	0.41	18.08
2008	0.78	11.97	0.44	19.07
2009	0.74	12.05	0.31	19.45
2010	0.70	12.65	0.29	19.40
2011	0.69	12.62	0.29	19.17
2012	0.71	12.02	0.29	18.88
2013	0.72	12.24	0.28	19.58
2014	0.65	12.73	0.08	20.54
2015	0.64	12.20	0.07	19.66
2016	0.73	11.46	0.06	21.52
2017	0.70	11.70	0.14	19.73
2018	0.74	12.15	0.45	17.62
Mean	**0.73**	**12.26**	**0.32**	**19.40**

3.3.3 Feature importance

Complex ensemble models, of which GBTs are an example, can be difficult to interpret. We make use of SHapley Additive exPlanations (SHAP) to quantify the importance of the input features to the trained GBT, and hence their importance to the predictive process, and to show that these features are consistent with what

would be expected given our understanding of the generating mechanisms of surface ozone. The Shapley values (Lundberg and S.-I. Lee, 2017) are one measure of feature importance that have been used previously to understand the relationship between input features and ozone in ML studies (e.g. Z. Liu et al., 2022. It is important to note that SHAP values cannot be interpreted as correlation coefficients; positive SHAP values can co-occur with either high (red) or low (blue) values of a feature, and similarly for negative SHAP values. This means that high values of a given feature may result in low or high ozone levels.

Figure 3.3 shows the feature importance (as SHAP values) for the final GBT model trained on all data, where negative SHAP values result in lower predictions and positive ones to higher predictions. Unsurprisingly, EMEP4UK ozone is the most important feature in predicting the measured ozone, followed by daily maximum 2m temperature, date (as an integer), month, and distance to road type 1 (i.e., motorways). Lower concentrations of EMEP4UK ozone have a greater impact on the GBT model output than high values, signifying that lower EMEP4UK ozone concentrations better represent the behaviour of measurement ozone than higher concentrations. Daily maximum 2m temperature is the most important meteorological feature, reflecting the well-established observed temperature/MDA8 relationship (e.g. Gouldsbrough et al., 2022) that is likely underpinned by several processes (W. Sun et al., 2017; Romer et al., 2018; Porter and Heald, 2019). Lower temperatures decreased the model prediction while higher temperatures increase the prediction. The high importance of the temporal features (date and month) indicates that seasonality and long-term trends of measurement ozone are not wholly captured in EMEP4UK. Road type 1 is the most important of the road types, and the fifth most important feature overall, reflecting the strong link between vehicles, NO_X, and ozone. Type 1 roads typically have a higher traffic volume and considerably higher NO_X concentrations than background locations, due to higher driving speeds and numbers of heavy good vehicles Mann, 1997.

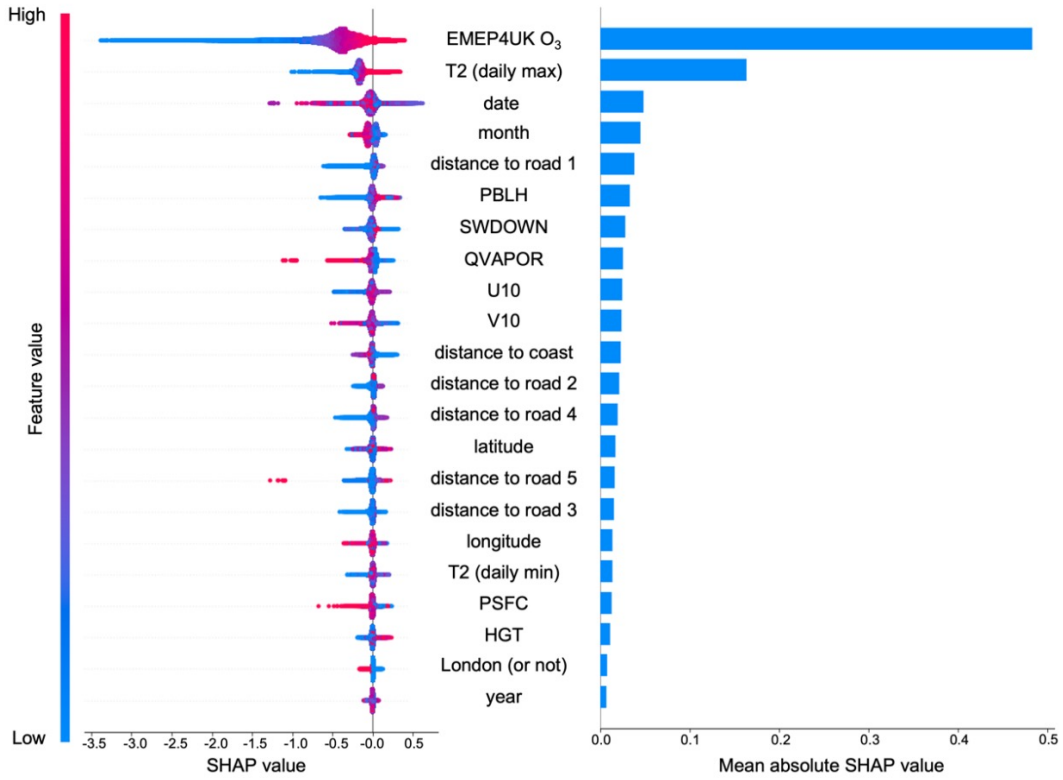

Figure 3.3: Feature importance in the GBT model ordered from most important (top), to least important (bottom). Local SHAP values (left) show the model impact of each feature based on feature value. Negative SHAP value results in a lower ozone prediction, and positive SHAP value results in a higher ozone prediction. Mean absolute SHAP values (right) show the overall impact of each feature on model output.

3.4 Results - analysis of the downscaled surface

We perform three analyses of our downscaled ozone surface: recent years (2014–2018), time trends (2001–2018), and heatwave years (2003, 2006 and 2018). In all cases, we compare the behaviour of four characteristics – annual mean, March–August mean, and the annual 10^{th} and 90^{th} percentiles – across the measurement stations and the EMEP4UK and downscaled surfaces.

3.4.1 Recent years analysis (2014–2018)

We examine the years 2014–2018 as these are the most recent years in the dataset. The five-year period accommodates the interannual variability in ozone concentrations, resulting in a broader overview of ozone behaviour. In the UK, elevated ozone mostly occurs in spring/summer (March–August) during anticyclonic conditions when slow moving air masses from mainland Europe contribute to increased accumulation of precursor emissions and increased rates of photochemical ozone production AQEG, 2021. Figure 3.4 shows the 2014–2018 annual (i.e., all months) and March–August (only) mean MDA8 ozone for each region and each data product. Recall that the measurement means are based on a limited and varying number of monitoring sites within each region (Figure B.1). Regional means, both annual and March–August, are consistently higher for the original EMEP4UK surface compared to the downscaled surface and measurements reflecting a high bias in the unadjusted CTM. See also Tables B.1 and B.2 which contain a summary of the data plotted in Figure 3.4. The all-region annual mean MDA8 from the downscaled surface ($\sim$62 µg/m^3) and measurements ($\sim$61 µg/m^3) are in close agreement, while the original EMEP4UK surface ($\sim$76 µg/m^3) is significantly larger (Table B.1). A similar pattern of agreement is found for the March–August means (Table B.2). One region where the downscaled surface and measurements differ considerably is London. Here, the annual mean MDA8 for the downscaled surface and measurements are 57 µg/m^3 and 49 µg/m^3, respectively (and 68 µg/m^3 versus

58 µg/m^3 for the March–August mean). The high proportion of urban measurement sites in London (i.e., sampling more NO_X titration), in contrast to the more varied site type sampling of the gridded downscaled surface, likely contributes to the higher MDA8 ozone in the downscaled surface.

Figure 3.5 shows the 2014–2018 regional average 90th and 10th percentiles of MDA8 ozone. The EMEP4UK 90th and 10th percentiles are higher for all regions compared to the downscaled surface and measurements, further demonstrating the high ozone bias present in the EMEP4UK surface. Southeast England has the highest regional 90th percentile MDA8 ozone concentration in both the downscaled (88 µg/m^3) and EMEP4UK (99 µg/m^3) surfaces, while Southwest England has the highest 90th percentile in the measurements at 90 µg/m^3. The most noticeable difference between the three datasets is the 90th percentile estimate for Scotland: 79 µg/m^3 for the downscaled surface, 95 µg/m^3 for the EMEP4UK surface, and 86 µg/m^3 for the measurements. Wales has the highest 10th percentile MDA8 ozone concentration in the EMEP4UK and downscaled surfaces, 62 µg/m^3 and 50 µg/m^3 respectively, while having only the third highest 10th percentile in the measurements, at 40 µg/m^3. The highest regional 10th percentile in the measurement data is 45 µg/m^3 for Scotland. The 10th percentiles for Scotland in the EMEP4UK and downscaled surfaces are higher still: 61 µg/m^3 and 48 µg/m^3, respectively. The relatively high 10th percentile in Scotland is likely due to the low regional NO_X emissions, as ozone in Northern Scotland reflects hemispheric background concentrations instead of the photochemical generated concentrations (Entwistle et al., 1997). The inter-region variation in 90th percentile is 12 µg/m^3 in the downscaled surface, whereas the inter-region variation in the 10th percentile is considerably higher at 21 µg/m^3, due to the particularly low 10th percentile ozone concentration in London of 29 µg/m^3. See Tables B.3 and B.4 which contain a summary of the data plotted in Figure 3.5.

Reports from Defra highlight that very few UK regions currently meet the UK government long-term ozone objective of MDA8 to not exceed 100 µg/m^3 more than

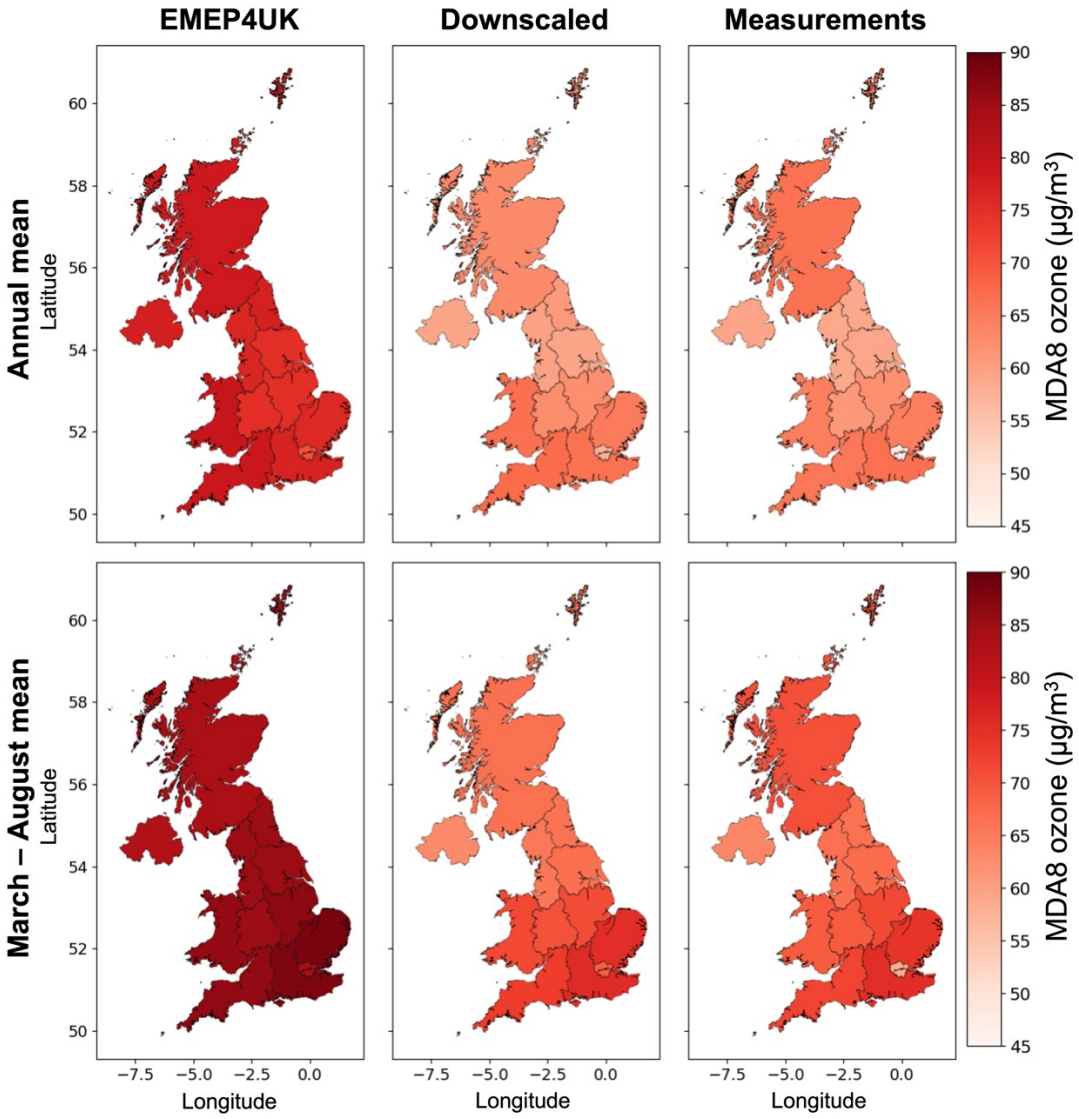

Figure 3.4: Comparing regional averages of EMEP4UK (left column), downscaled (middle column), and measurement (right column) MDA8 ozone (µg/m^3) for 2014–2018 annual mean (top row) and March–August mean (bottom row).

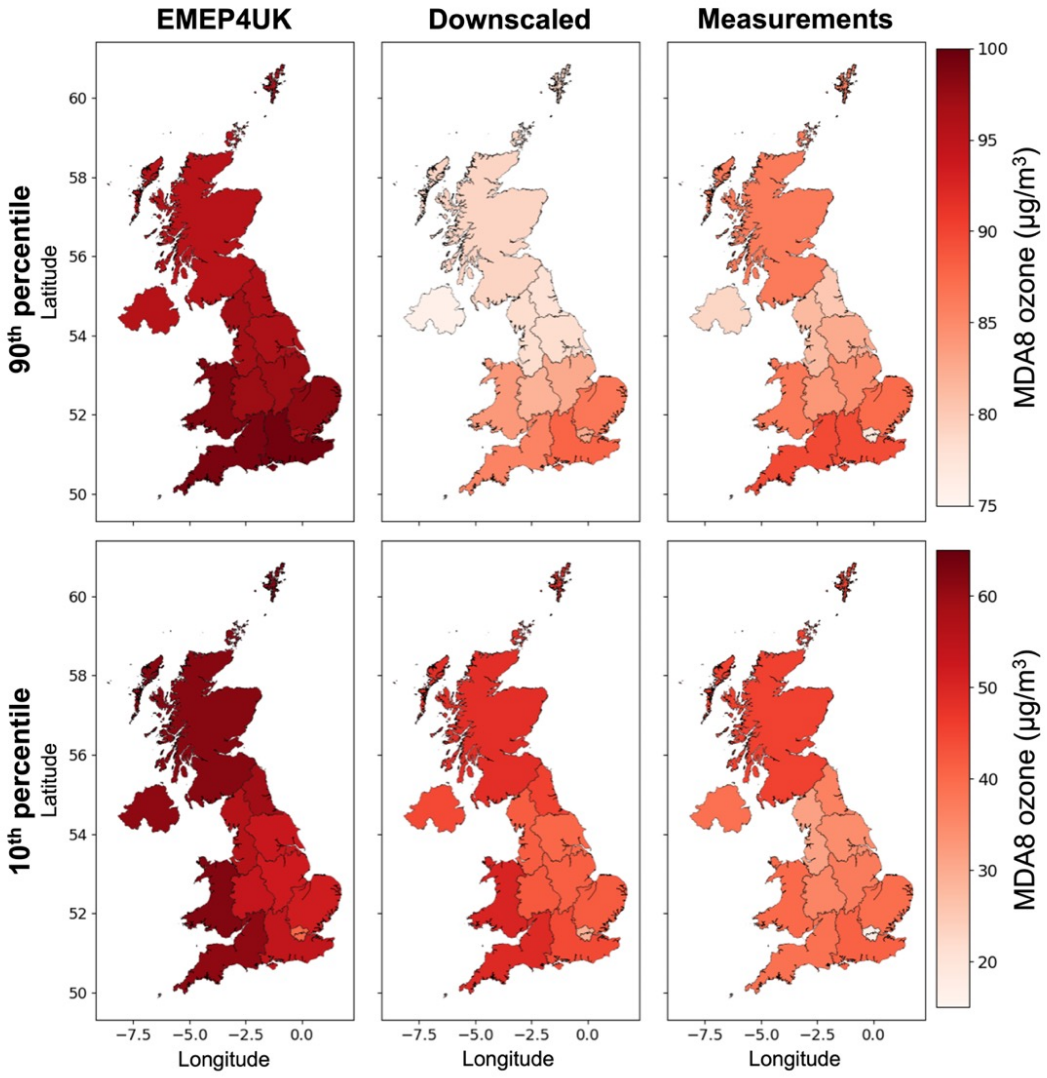

Figure 3.5: Comparing regional averages of EMEP4UK (left column), downscaled (middle column), and measurement (right column) MDA8 ozone (µg/m^3) for 2014–2018 90th percentile (top row) and 10th percentile (bottom row).

10 times in a year. Similarly, the EU's long-term objective (no MDA8 exceedances of 120 µg/m^3) is routinely breached in most areas. A summary of the guidelines against which the UK reports is given in Table 1 of Defra (2021c). For the above assessments, which are based on measurements from the AURN network, if one location (monitoring site) within a region is in breach of the objective, that region is deemed non-compliant. For example, in 2021 no regions outside of Scotland met the EU objective (Defra, 2022). Our downscaled surface provides an additional perspective on adherence that is not possible to obtain from a relatively sparse monitoring network alone or from a CTM with significant bias. Figure 3.6 shows the number of days in a year exceeding 100 µg/m^3 averaged over 2014–2018 for both the downscaled and unadjusted EMEP4UK datasets, with yellow cells highlighting areas where 100 µg/m^3 is exceeded less than 10 times per year and therefore passing the government objective. We find that 27% of the downscaled UK surface exceeds the government objective, compared to 99% from EMEP4UK. This underpins the importance of bias correction when using process models to examine policy metrics and air quality exposure indicators. At least one downscaled cell in all the 12 UK regions was found to have more than 10 days with MDA8 greater than 100 µg/m^3 averaged over 2014–2018, however the regions in the southeast of the UK have the greatest proportion of failing cells, with 86% and 88% of the East and South East regions failing.

3.4.2 Trends over time

As mentioned earlier, the use of measurement data only is limited by varying, and in some cases very short, measurement periods and this can prove problematic in air quality trend analysis (Lang, 2020). The gridded downscaled and EMEP4UK datasets facilitate the estimation of trends for all regions, regardless of the density of the measurement network and/or the completeness of the measurement records. We illustrate this in the subsequent analysis by quantifying regional trends in ozone concentrations, comparing, as before, measurement, downscaled and EMEP4UK

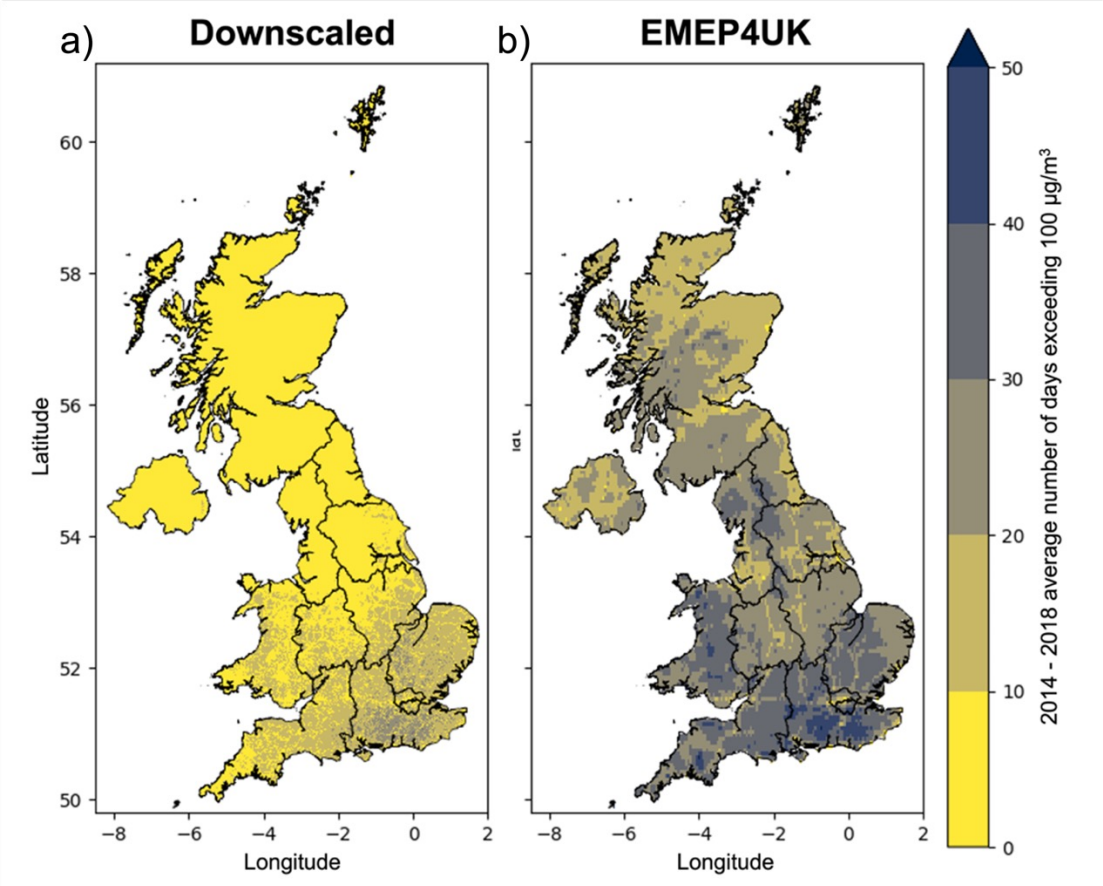

Figure 3.6: 2014–2018 average number of days per year where each cell exceeds a level of 100 µg/m^3 in a) downscaled ozone surface and b) EMEP4UK ozone surface.

estimates. We do not quantify a single UK-wide trend since the behaviour of ozone, and consequently also the observed long-term trends, differ considerably across the UK.

A benefit of using a gridded ozone surface is greater spatial coverage; this enables estimation of regional trends in regions where measurement stations are sparse. Figure 3.7 shows the annual and March–August regional mean MDA8 ozone trends for 2001–2018, for the three datasets. These trends are also given in Tables B.5 and B.6, respectively. All regions have a decreasing trend in annual mean MDA8 ozone in the downscaled surface; however, no trends are statistically significant. Southeast England has the greatest trend at -0.26 [-0.56, 0.04] µg/m^3, followed by Southwest England and Wales, at -0.25 [-0.55, 0.04] µg/m^3 and -0.25 [-0.50, 0.00] µg/m^3, respectively. In comparison, the measurements and EMEP4UK surfaces both have a combination of increasing and decreasing regional trends (few of which are statistically significant). In the measurements, Yorkshire and The Humber has the greatest increasing trend in annual mean ozone concentrations, at 0.33 [0.02, 0.64] µg/m^3 per year, followed by the West Midlands and Northwest England at 0.28 [-0.01, 0.56] µg/m^3 and 0.29 [-0.02, 0.59] µg/m^3 per year, respectively. EMEP4UK is the only dataset for which London has a significant increasing trend in annual mean ozone, at 0.43 [0.20, 0.66] µg/m^3, with non-significant decreasing trends of -0.20 [-0.48, 0.09] µg/m^3 and -0.23 [-0.50, 0.05] µg/m^3 per year in, respectively, the measurements and downscaled surface.

In contrast to the annual mean case discussed above, most March–August mean ozone trends are statistically significant in the downscaled and EMEP4UK surfaces (Table B.6). When comparing the March–August mean trends we also see a greater similarity between the downscaled and EMEP4UK surfaces (in terms of the sign of the trend), except for London where the trend is again positive in the EMEP4UK surface (0.17 [-0.06, 0.39] µg/m^3 per year), and negative in the downscaled surface (-0.34 [-0.77, 0.10] µg/m^3 per year). While all regions have a decreasing March–August mean trend in the downscaled surface, the largest reductions are seen in the south

of the UK: Southeast England and Southwest England at -0.58 [-1.02, -0.15] µg/m^3 and -0.52 [-0.95, -0.09] µg/m^3 per year, respectively. Note, the MDA8 ozone data that underpin the above trend analysis is shown in Figures B.3–B.5 and Figures B.6–B.8 for the annual mean and March–August mean, respectively. These figures demonstrate the regional interannual variation in MDA8 ozone concentrations, which is missed by only considering the trends.

Figure 3.8 shows the regional trends of 90$^{\text{th}}$ and 10$^{\text{th}}$ percentile ozone concentrations for the period 2001–2018; estimates and confidence intervals are also given in Tables B.7 and B.8. 90$^{\text{th}}$ percentile ozone is decreasing for all regions and all datasets. Looking at the downscaled surface, the downward trend is significant in half of the regions considered, with the greatest changes in 90$^{\text{th}}$ percentile ozone in the south of the UK, particularly for South East (England) with a trend of -0.74 [-1.35, -0.12] µg/m^3 per year. In comparison, regions with the greatest change in 90$^{\text{th}}$ percentile ozone in the EMEP4UK surface and measurements are East of England at -0.75 [-1.09, -0.40] µg/m^3, and North East (England) at -0.59 [-0.95, -0.22] µg/m^3 per year. 10$^{\text{th}}$ percentile ozone is increasing for most regions in the downscaled and EMEP4UK surface, and for all regions in the measurements. Northern Ireland, Wales, and Scotland have a slightly decreasing 10$^{\text{th}}$ percentile trends in the downscaled surface, at -0.01, -0.01 and 0.06 µg/m^3 per year, respectively, though none of these trends are statistically significant. We find a greater increase in 10$^{\text{th}}$ percentile ozone for London in the downscaled surface than in the measurements, at 1.19 [0.75, 1.62] µg/m^3 and 0.17 [-0.04, 0.37] µg/m^3 per year, respectively. We suspect this is again due to urban site type bias in the measurements, compared with the more varied sampling in the gridded downscaled surface. The regional yearly 90$^{\text{th}}$ percentiles in the downscaled, EMEP4UK and measurement datasets are shown in Figures B.9–B.11, respectively. The equivalent figures for the 10$^{\text{th}}$ percentiles are shown in Figures B.12–B.14.

The UK government has a long-term objective that MDA8 ozone should not exceed a level of 100 µg/m^3 more than ten times per year. Earlier we showed

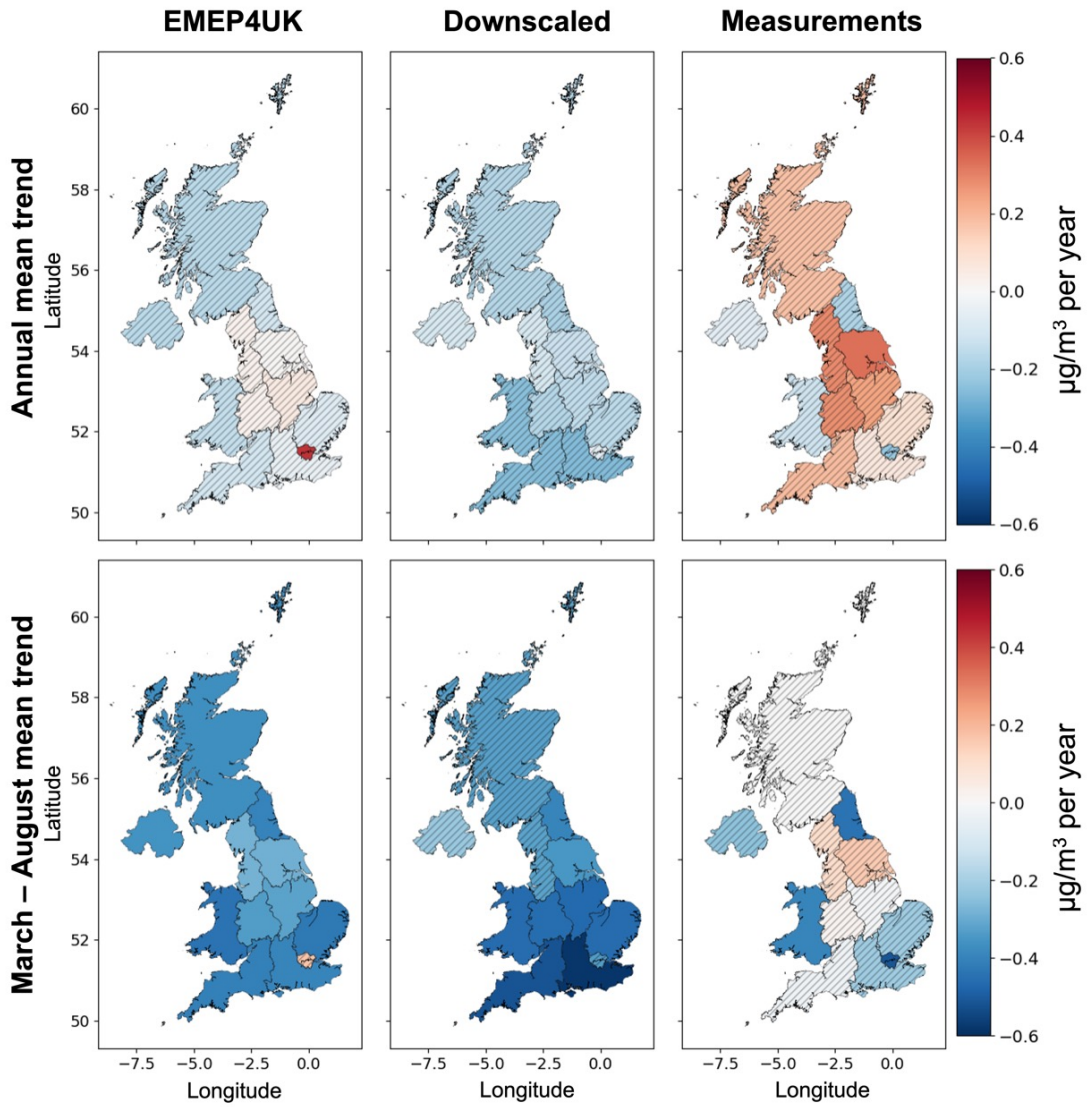

Figure 3.7: Annual mean and March–August mean trends for each region, for each dataset. Regions with insignificant trends are hatched.

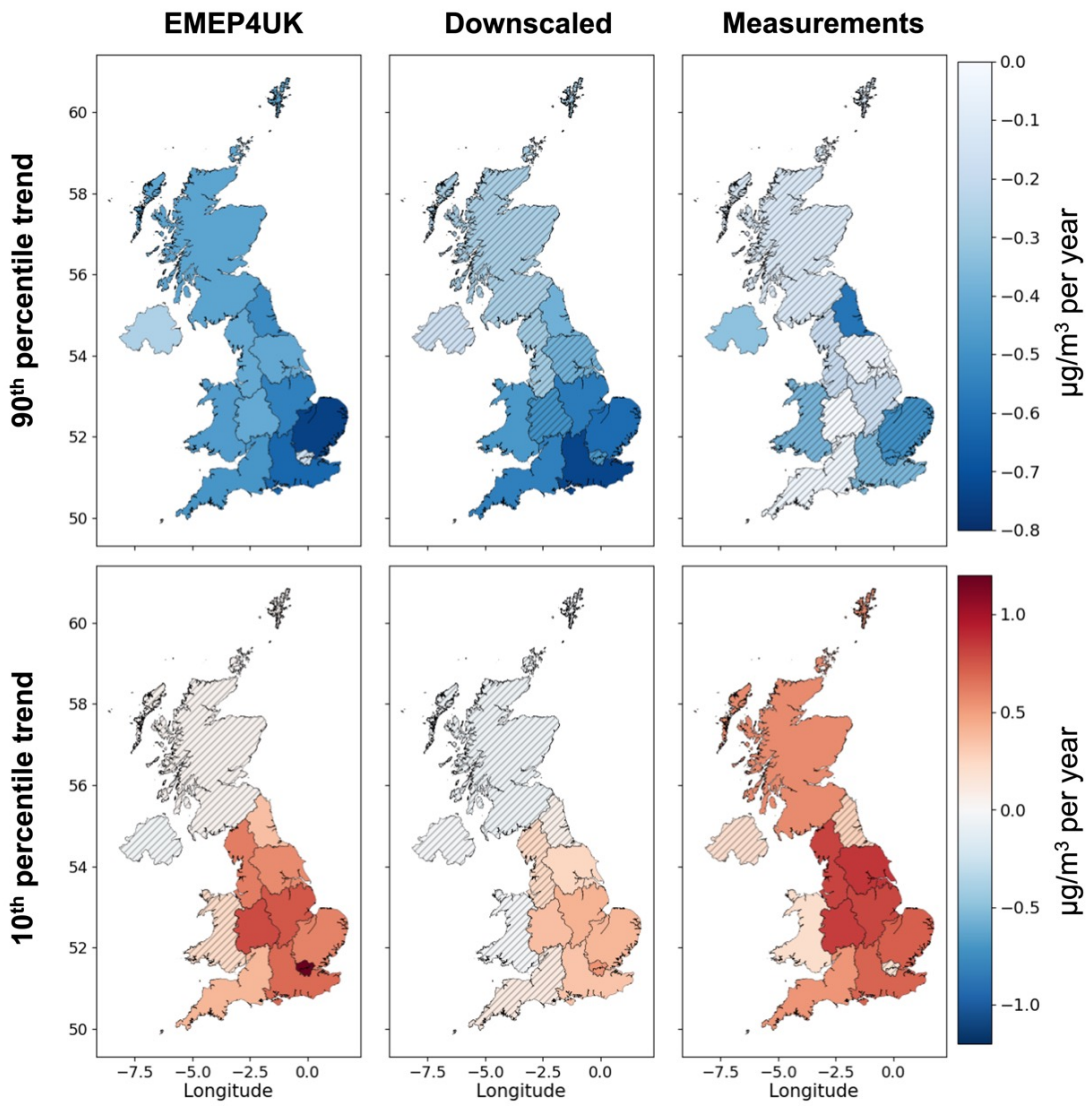

Figure 3.8: 90$^{\text{th}}$ percentile and 10$^{\text{th}}$ percentile trends for each region, for each dataset. Regions with insignificant trends are hatched.

that conclusions regarding the degree to which this objective is being met differ substantially between the downscaled and original EMEP4UK model (Figure 3.6). As the downscaled surface, measurement data and EMEP4UK surface all have a different number of cells/stations, we look at the time trend in the percentage of sites or grid cells which fail to meet this, rather than the trend in absolute number. A decreasing time trend in the percentage of the UK failing to meet the objective is seen for all three datasets. However, the only statistically significant trend is for the downscaled surface, at -2.19 [-4.32, -0.07] % per year. The EMEP4UK trend is less steep, -0.60 [-1.62, 0.43] % per year, and the measurements trend lies in between at -1.73 [-3.78, 0.32] % per year.

Recalling that a core aim for our downscaling methodology was to better represent the tail behaviour of measurement ozone, we now consider the specific years 2003, 2006 and 2018 (hereafter "heatwave years") which were significantly warmer than average (see later), and when UK ozone levels were elevated (Diaz et al., 2020). Figure 3.9 shows the number of days that exceed a level of 100 µg/m^3 for each heatwave year, along with the corresponding yearly mean temperature. Yellow cells highlight areas where 100 µg/m^3 is exceeded less than 10 times per year and therefore passing the government objective. In the EMEP4UK surface, almost all the UK (more than 99%) is exceeding the government objective in the heatwave years. In the downscaled surface, 88%, 87% and 53% of the UK is failing the government objective in 2003, 2006 and 2018, respectively. These percentages are substantially higher than the 2014–2018 average percentage of 27%, demonstrating the more frequent occurrence of exceedance days above 100 µg/m^3 in heatwave years. Both the EMEP4UK and downscaled surfaces show a change over time in the amount of the UK exceeding the government objective in heatwave years, with the highest number of exceedances in 2003, and lowest in 2018. The areas with the highest number of exceedances are correlated with the temperature maps, with more exceedance occurring where the yearly mean temperatures are higher, consistent with the well documented link between MDA8 ozone and temperature in

the literature.

3.4.3 Analysis of NO_X scenarios

A major application of CTMs is to aid understanding of pollutant behaviour under future emissions or climate change scenarios. In this section we explore the effect of reductions in UK NO_X emissions on ozone under 2018 meteorological conditions. We compare four downscaled surfaces: (1) a 2018 base run, (2) a run as base but with a 20% reduction in UK terrestrial NO_X emissions, (3) a 40% reduction, and (4) an 80% reduction (see also Section 3.2.1). The year 2018 is of specific interest as it was the seventh warmest year in the UK since 1884 (Kendon et al., 2019) with a mean temperature that was 0.6 °C above the 1981–2010 average.

Since 2018 was, climatologically, an atypical year for the UK, we first compared two 2018 EMEP4UK base run downscaled surfaces. The first was obtained from the ML downscaling model trained on the 2001–2018 EMEP4UK data used in Sections 3.3 and 3.4, and the second from the same downscaling model but trained only on the 2018 base run data. We found that the two sets of predicted surfaces performed almost identically in capturing the behaviour of the 2018 measurement data (both have an R^2 of 0.74 and RMSE of $\sim$12.15 µg/m^3). These findings support our decision to use the downscaling model trained on the 2001–2018 data (see Section 3.2.1) to downscale not just the base run of 2018, but also the ozone surfaces under the three NO_X scenario runs. Implicit in downscaling the scenarios in this way is the assumption that the associations of the input features and surface level ozone remain the same, even as NO_X levels decrease. We acknowledge this as a potential limitation but note that similar assumptions are endemic throughout the downscaling literature.

Figure 3.10 shows the point wise differences in downscaled annual mean, March–August mean, and 90th percentile ozone for the three NO_X scenarios compared to the equivalent statistics for the 2018 base run. Both 20% and 40% NO_X reductions result in most of the UK seeing increased annual mean ozone

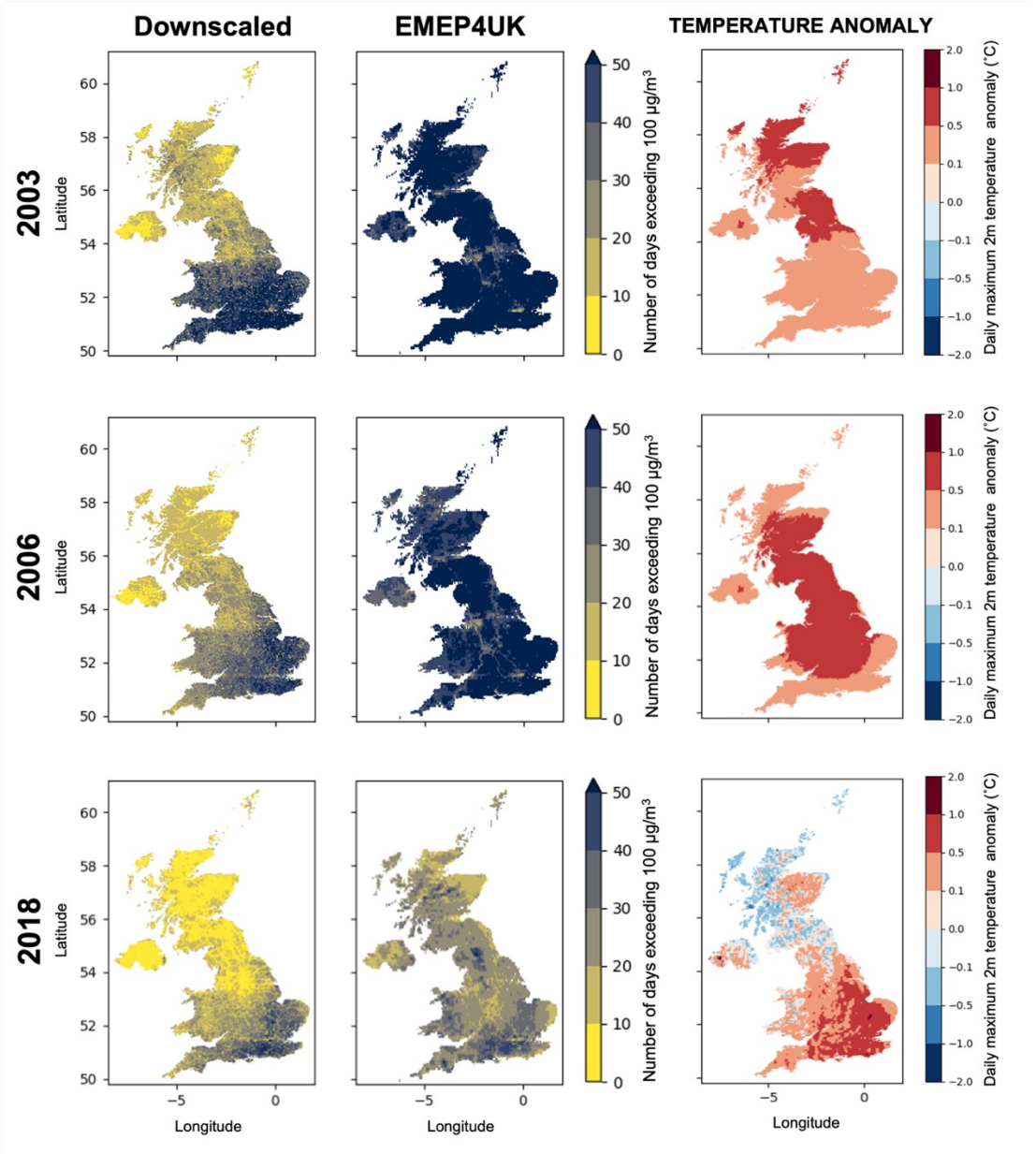

Figure 3.9: Number of days where each cell exceeds a level of 100 µg/m^3 in downscaled surface ozone (left) and EMEP4UK surface ozone (middle), for heatwaves years 2003, 2006 and 2018. Also given are the yearly mean daily maximum temperature anomalies compared to the 2001–2018 average (right) for each heatwave year.

concentrations, except rural areas in Scotland. Conversely, the more drastic NO_X reduction of 80% results in a decrease of annual mean ozone for much of the UK, particularly large portions of South West (England), Wales, Scotland, Northern Ireland, North West (England) and North East (England). In contrast to the trends seen in the annual mean, larger and more widely spread decreases are seen in the March–August mean for all NO_X reduction scenarios, suggesting the impact on spring and summer mean ozone is greater than on the annual mean. The change in 90^{th} percentile ozone is far more granular, largely due to the differences in tail behaviour of ozone at rural and urban locations. The relatively moderate reductions in NO_X concentrations of 20% and 40% lead to increases in 90^{th} percentile ozone in parts of the UK, whereas the more substantial reduction in NO_X concentrations of 80% results in only very urban areas having increases in 90^{th} percentile ozone, such as Manchester, Leeds, Sheffield, Birmingham, London, Newcastle, Edinburgh, Glasgow, and Aberdeen. Similar increases in high-level ozone due to NO_X reductions has been show for several cities in the US (Gao et al., 2013) and reflect the interdependence of ozone concentrations and NO_X mitigation strategies. Finally, a similar sensitivity analysis based on the unadjusted EMEP4UK output (Figure B.15) exhibits a very similar pattern of ozone response.

In summary, the above analysis demonstrates the applicability of our downscaled EMEP4UK surface to examine the sensitivity of UK ozone to changes in precursor emissions at high spatial resolution. These results also emphasise the challenges in controlling surface ozone (especially in urban areas) if NO_X emissions continue to decline substantially – an effect which few studies have demonstrated for the UK using models to date.

3.5 Conclusions

We have proposed a machine learning methodology to spatially downscale surface ozone output from the EMEP4UK chemical transport model from its native 5×5 km resolution to a 1×1 km resolution. Taking a 1×1 km interpolation of the

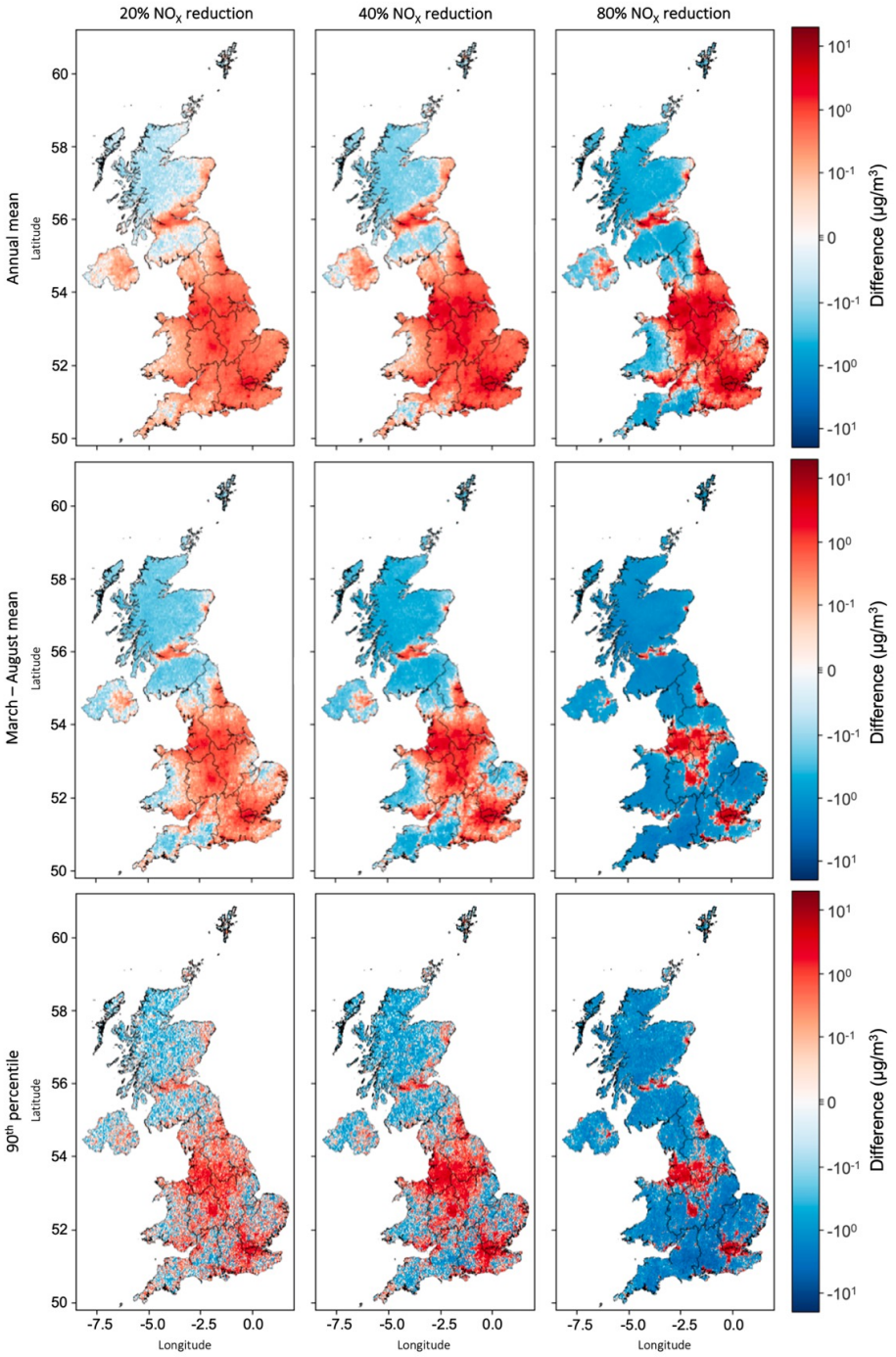

Figure 3.10: Difference in annual mean (top), March–August mean (middle), and 90th percentile (bottom) MDA8 ozone compared to 2018 for three UK NO_X scenarios: 20% reduction in NO_X (left), 40% reduction in NO_X (middle), 80% reduction in NO_X (right).

original EMEP4UK grid as input, our algorithm uses a gradient boosting tree to predict a high-resolution gridded ozone surface. The algorithm was trained to predict measurement data from sites across the UK, as well as 21 input feature variables. We find that the downscaled surface better represents the behaviour of measurement ozone, with a 128% improvement in R^2 and 37% reduction in RMSE compared to the EMEP4UK surface. The GBT allows replication of the behaviour of complex non-linear systems and the ability to work with high dimensional datasets. Producing the downscaled surface using the proposed methodology is far quicker and less computationally expensive than running a high-resolution CTM. We therefore consider this methodology to be a useful post-processing tool for CTMs that can efficiently produce higher resolution ozone surfaces and, as is the case for EMEP4UK, reduce biases by incorporating information from measurements. A further advantage of the proposed ML downscaling model is the ability to identify the most important features for the prediction of MDA8 ozone. Consistent with previous work, daily maximum 2m temperature is found to be the most important meteorological feature, with elevated temperatures strongly associated with high level ozone.

Our analysis on recent years (2014–2018) finds that South East (England) and South West (England) experience higher March–August concentrations of ozone than other regions. We find greater inter-region differences in spring/summer mean ozone concentrations than annual mean. There is a clear north-south difference of high percentile ozone in the downscaled surface, with high ozone concentrations in the south of the UK. Low percentile ozone has the greatest inter-region variation in the downscaled surface due to the particularly low 10^{th} percentile ozone concentration in London. This demonstrates the effect greater NO_X concentrations in highly urban areas in reducing background ozone concentrations through NO_X-titration.

We have estimated regional trends in various statistics of ozone using data from 2001–2018 for the three datasets: EMEP4UK, downscaled surface, and measurements. Annual and March–August mean ozone decreases for all regions

in the downscaled surface, while some regions have increasing trends in the measurements. EMEP4UK is the only dataset to estimate an increase in annual mean ozone for London. The proposed downscaling surface is useful when considering how UK ozone has changed over time as it is higher resolution than EMEP4UK, and provides more spatially complete coverage than measurements alone. The downscaling process also addresses the high bias present in EMEP4UK, resulting in a better reflection of high-level ozone relevant to health. We find an improved picture of high-level ozone when using the downscaled surface, with only 53% of the UK failing its government objective (to not exceed an ozone level of 100 µg/m^3 more than 10 times per year) in 2018, compared to 99% of the UK failing this objective in EMEP4UK. Further improvement in high-level ozone is apparent from considering trends in 90^{th} percentile ozone. We find significant reductions in 90^{th} percentile ozone for half of the regions considered in the downscaled surface, with the greatest reductions in the south of the UK, particularly for South East (England).

Through a sensitivity analysis, we considered the effect of three NO_X reduction scenarios on UK ozone concentrations downscaled using the proposed downscaling method. Moderate (20% and 40%) reductions in NO_X concentrations are shown to increase annual mean ozone for most of the UK, whereas significant (80%) reductions decrease annual mean ozone for large parts of the UK. More of the UK shows a decrease in March–August mean ozone for all NO_X scenarios, suggesting a stronger link between spring and summer ozone concentrations and NO_X than annual mean concentrations. The differences in the tail behaviour of ozone at urban and rural locations is made evident in the effect of NO_X reductions on 90^{th} percentile ozone. Very urban areas see the largest increases in 90^{th} percentile ozone when reducing NO_X concentrations by 80%, and this includes many of the UK's biggest cities. We determine it important to further understand the effect of NO_X reductions on UK ozone, as a considerable portion of the UK population live in these urban areas. These results reemphasise the broader challenges around NO_X mitigation strategies.

To conclude, machine learning based downscaling approaches offer a promising way to study pollutant trends and to assess the impact of policies for ozone and in principle other pollutants. A focus of future work will be to exploit the bias-corrected downscaled surfaces for the assessment of population exposure to poor air quality and to help quantify the resulting health impacts.

3.6 Acknowledgements

LG acknowledges the UK Engineering and Physical Research Council (EPSRC) for a PhD studentship (EP/R513076/1, project reference 2353903). RH was supported by a UK Natural Environment Research Council (NERC) Independent Research Fellowship (NE/N014375/1). EE and PJY were supported by the EPSRC-funded Data Science of the Natural Environment project (EP/R01860X/1). MV was supported by the Natural Environment Research Council (NE/R016429/1) as part of the UK-SCAPE programme delivering National Capability.

Chapter 4

Identifying the drivers of high-level ozone events using machine learning classification

Lily Gouldsbrough[1], Emma Eastoe[2], Ryan Hossaini[1,3], Paul J. Young[1,4]

[1]Lancaster Environment Centre, Lancaster University, Lancaster, UK

[2]Department of Mathematics and Statistics, Lancaster University, Lancaster, UK

[3]Centre of Excellence in Environmental Data Science, Lancaster University, Lancaster, UK

[4]JBA Risk Management Limited, Broughton Park, Skipton, BD23 3FD, UK

Corresponding author: Lily Gouldsbrough (lilygouldsbrough@outlook.com)

Statement of contribution. Lily Gouldsbrough led the methodology, data curation and analysis and conceived the study alongside Ryan Hossaini, Emma Eastoe. Lily Gouldsbrough drafted the manuscript, with guidance from Ryan Hossaini, Emma Eastoe and Paul J. Young.

Abstract

High concentrations of surface level ozone are harmful to human and plant health, making their reduction a global priority. At the same time, it is also essential to efficiently produce skilful short lead-time forecasts of high-level ozone concentrations to provide timely public health warnings. Both objectives require an understanding of the key drivers of high-level ozone and how these vary over different regions and local environments. We present a data-driven machine learning classification approach which is used to identify the most important factors contributing to high-level ozone concentrations in, and across, the UK. The classifier exploits input variables known to affect ozone concentrations, including meteorological, temporal, and spatial variables, to predict with high accuracy the occurrence of high-level ozone events. Through the implementation of SHAP measures, we can determine the extent to which each of these factors influences high-level ozone and identify their interactions. We use this driver information to develop a model that can inform the 'traffic light' warning system used by the UK government to advise the public of potential over-exposure. The model that we propose is a machine learning classification approach, trained to forecast days with an exceedance of the UK's 'moderate' ozone health threshold (MDA8 ozone>100 $\mu g/m^3$). We achieve high accuracy in forecasting exceedance days with hit rates ranging from 0.69 to 0.78 for the years considered: 2003, 2006, 2008 and 2018. Our results show that daily maximum 2m temperature is the most important predictor feature for both identifying high-level ozone events and forecasting moderate health threshold exceedance days.

4.1 Introduction

High concentrations of surface level ozone have detrimental impacts on human health (Ji et al., 2011; COMEAP, 2015; Nuvolone et al., 2018). Consequently, the monitoring, forecasting and reduction of ozone levels is a global priority in the management of air quality. Ozone is a particularly complex pollutant since it is not directly emitted; rather the formation and destruction of ozone involve complex chemical reactions that depend on both concentrations of precursor pollutants, NO_X and VOCs, and meteorological and atmospheric conditions at global, regional and local scales (Lelieveld and Dentener, 2000). To minimise morbidity and mortality rates during periods of poor air quality, many governments, including the UK government, have implemented a 'traffic light' system under which ozone levels are classed as normal, moderate, high or very high (see https://uk-air.defra.gov.uk/air-pollution/daqi?view=more-info). An advance public health warning is issued if the ozone forecast suggests that levels are likely to exceed one of the thresholds which would result in the potential for over-exposure. Accurate forecasting of high ozone levels is therefore key to the effectiveness of this system.

Atmospheric chemistry transport models (CTMs) have been used to produce operational air quality forecasts (Savage et al., 2013; A. Kumar et al., 2017; Spiridonov et al., 2019; Stortini et al., 2020). These models are well developed to accurately reproduce the expected behaviour of an array of different pollutants, at a high spatio-temporal resolution. However, developing and running CTMs can be complex and computationally expensive, often requiring specialist machines. The online air quality model AQUM (Air Quality in the Unified Model), developed and maintained by the UK Met Office, provides regional air quality forecasts for the UK, and has been shown to have a good level of performance in forecasting elevated ozone episode conditions (Savage et al., 2013).

To make models like AQUM useful in forecasting high-level events, we need to perform bias correction. Neal et al. (2014) proposed an automated air quality forecast bias correction scheme for the AQUM output, based on the short-term/

persistence of model bias with respect to recent observations to improve the accuracy of these forecasts. We propose an alternative data-driven approach based on machine learning models. The flexibility and computational efficiency of these models allows us to both identify the drivers behind high-level ozone and assess whether these are consistent across varying degrees of 'high-level'. We then show how, with minor modifications, the approach can be used to make short range forecasts of when an ozone episode is expected to occur. This forecast can then be used to directly inform whether a health warning is made.

The key advantage of this data-driven approach is that it allows us to learn from the data directly about the events in which we are most interested. Moreover, relative to CTMs, training a machine learning (ML) model is highly computationally efficient allowing for quick and agile model development, including easily updating should additional information become available in the future. Of the two types of data-driven model, statistical and machine learning, ML models offer a particularly good alternative to CTMs for air quality forecasting. In contrast to statistical models, they have much weaker constraints on the mathematical forms for the complex interactions and relationships between pollutants and their chemical and physical drivers. Further, ML models do not require that the form of these relationships be specified in advance of modelling; rather they are informed by the data. Lastly, ML models can leverage the large amount of measurement data available, as well as relevant meteorological and spatial information, in a computationally efficient manner. Gradient boosted trees (GBTs) are a particular type of ML model that can be address either regression or classification questions. As discussed below, we exploit GBTs to develop a classification algorithm, however they have previously been used in the air quality literature to build regression models to predict ozone exposure during wildfires (Watson et al., 2019), to create spatiotemporally resolved ozone surfaces for China (R. Li et al., 2020; Yuan Wang, Yuan, T. Li, et al., 2021; Hu et al., 2022; Song et al., 2022; Yuan Wang, Yuan, Zhu, et al., 2022; Chi et al., 2023; L. Liu et al., 2023), and to bias correct chemistry

transport models (Keller et al., 2021; Ortiz et al., 2021; Yin et al., 2021).

In this chapter we address the questions of driver identification and forecasting of ozone episodes in the UK where, despite reductions in the emissions of precursor pollutants in the UK and Europe, the mortality attributable to exposure to ozone between 1970 and 2010 increased by 17% (Carnell et al., 2019). The UK government has set health thresholds for the maximum daily 8-hr average (MDA8) ozone concentrations, with a moderate health threshold of $100\,\mu g/m^3$. Exceedances of these thresholds pose an increasing risk to public health, particularly for vulnerable groups such as children, the elderly and people with respiratory or cardiovascular conditions (Srebot et al., 2009). Rather than taking the first principles, physical science-based approach of a CTM, we train a ML GBT classifier on occurrences of health threshold exceedances derived from in situ measurements. This allows us to determine the factors most associated with these events. In the UK, ozone episodes have been shown to be affected by concentrations of NO_X and VOCs over north-west Europe (Jenkin, 2008); elevated temperatures (Finch and Palmer, 2020; Gouldsbrough et al., 2022); and synoptic weather conditions, particularly easterly flows, which transport precursor gases from mainland Europe, and dry, sunny and still anticyclonic conditions, which lead to greater ozone formation and decreased dispersion (Pope et al., 2016). These studies support our choice of drivers, which include both NO_X and relevant meteorological variables. Air quality measurements are obtained from AURN, a UK-wide monitoring network used by the UK government to monitor and assess air quality across the country, and to develop policies and measures to improve air quality and protect public health. Meteorological data is obtained from simulations of the Weather Research and Forecasting model (WRF) (C. Skamarock et al., 2008).

GBT classifiers are highly flexible and can identify highly non-linear relationships between input values and response, unlike simpler classifiers based on statistical models such as logistic regression. However, the resulting relationships are harder to visualise and the effect of each input variable on the response harder to describe.

To gain a better understanding of the significance of each input variable in the classification model, we use SHAP (SHapley Additive exPlanations), which permits interpretations of the output of a ML model by assessing the contribution of each input variable to the predicted outcome (Lundberg and S.-I. Lee, 2017). SHAP values enable us to quantify the relative importance of each input variable by providing insights into the contribution of each variable to the final prediction, and therefore identify the drivers of high-level ozone concentrations for the UK. To our knowledge, this is the first time that machine learning classification and SHAP have been used together to determine the drivers of high-level ozone for the UK.

This paper is structured as follows. Section 4.2 outlines the data used in the analysis and the features included with the ML models. Section 4.3 describes the methods used, including the ML models, and the driver and forecasting experiment setups. In Section 4.4 we determine the drivers of high-level ozone across the UK, focusing on synoptic weather conditions (Section 4.4.1), and on meteorological, spatial and temporal features (4.4.2). Section 4.5 presents results from four health threshold exceedance forecasting experiments. Finally, in Section 4.6 we present our conclusions.

4.2 Data

Our analysis used ozone and NO_X measurements, meteorological data from WRF, synoptic weather types and distance to roads data. These are described individually in the following subsections.

4.2.1 Ozone and NO_X measurements

Surface ozone and NO_X measurement data were obtained from the Automatic Urban and Rural Network (AURN: https://uk-air.defra.gov.uk/networks/network-info?view=aurn), via the R package openair (Carslaw and Ropkins, 2012). This network of air quality monitoring stations across the UK is operated by the UK

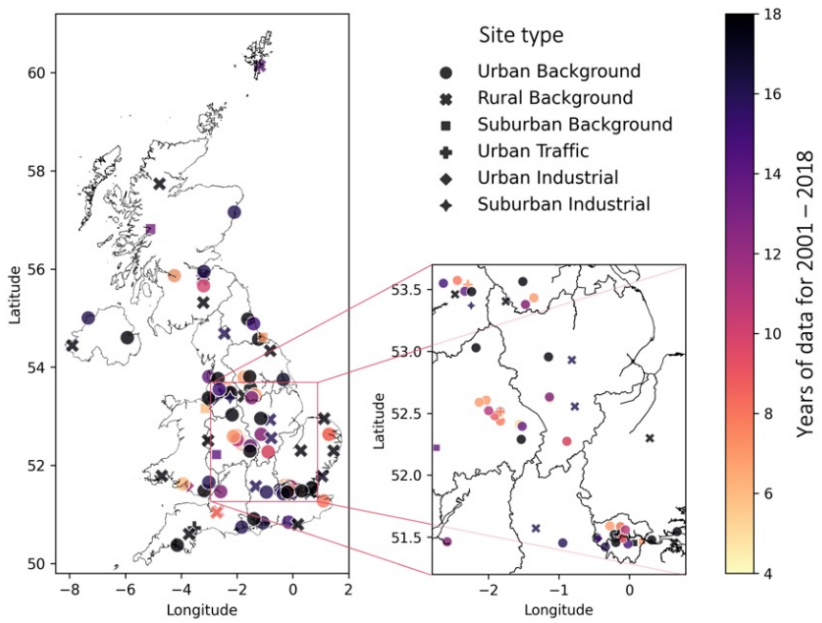

Figure 4.1: Map of AURN measurement stations used, showing station site type (symbols) and the number of years of data (colour). Central England (including London) is shown in an expanded view.

government's Department for Environment, Food and Rural Affairs (Defra) and provides continuous monitoring of air pollution levels in both urban and rural areas. We obtained data at each of 108 sites (see Figure 4.1) over the period 2001–2018. Variations in the start and end dates of monitoring result in differences in the length of records across the sites. The number of measurement stations in each UK region are listed in Table 4.1.

4.2.2 Meteorological variables

Meteorological variables for the years 2001–2018 and terrain height were obtained from archived output from previous simulations of the Weather Research and Forecasting (WRF) model, version 3.7.1 (C. Skamarock et al., 2008) for years 2001–2017 and WRF4.1 (W. C. Skamarock et al., 2019) for 2018. The WRF model is a numerical modelling system designed to simulate and predict weather

Table 4.1: Total number of measurement stations for each UK region with data in the period 2001–2018.

Region	**Number of measurement stations**
Central Scotland	5
East Midlands	8
Eastern	8
Greater London	16
Highland	3
North East	4
North East Scotland	1
North Wales	2
North West Merseyside	13
Northern Ireland	3
Scottish Borders	2
South East	8
South Wales	6
South West	8
West Midlands	13
Yorkshire Humberside	8

patterns and atmospheric processes (Powers et al., 2017). The WRF simulation in this work incorporates data from the numerical weather prediction model meteorological reanalysis of the US National Center for Environmental Prediction (NCEP)/National Center for Atmospheric Research (NCAR) Global Forecast System (GFS) (Environmental Prediction, 2000). The WRF model was run on a 5×5 km resolution and then we linearly interpolated this to a 1×1 km resolution. The WRF variables were then taken at the nearest grid cell locations to each measurement station.

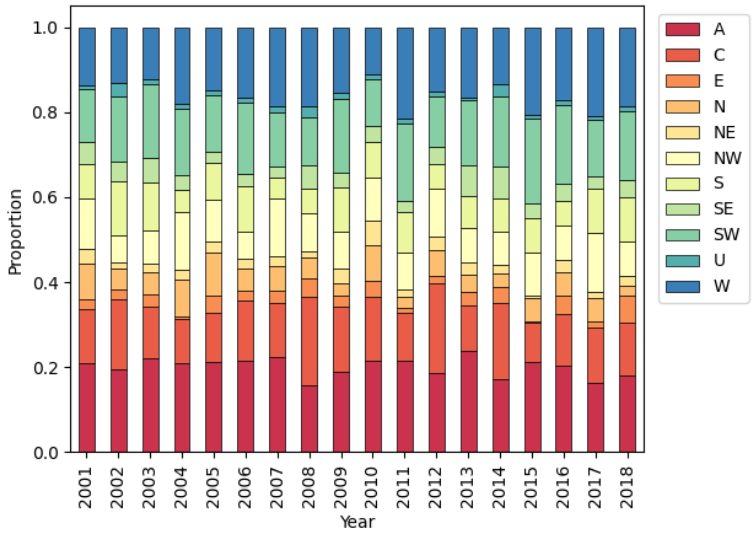

Figure 4.2: Proportion of days in a given year with each of our 11 Lamb weather types (see text and Table 4.2), for years 2001–2018.

4.2.3 Lamb weather types

Lamb weather types (LWTs) are an objective classification of the daily synoptic weather conditions across the UK. Rather than the original LWTs (Lamb, 1972), we use the recent classification proposed by Jenkinson and Collison (1997) maintained by the Climatic Research Unit at the University of East Anglia (https://crudata.uea.ac.uk/cru/data/lwt/). The 27 LWTs include 'pure' and 'hybrid' types; for instance, #0 is purely anticyclonic (A) whereas #1 is anticyclonic north-easterly (ANE), a hybrid between anticyclonic and north-easterly. We regroup the 27 LWTs into 11 types (A, C, NE, E, SE, S, SW, W, NW, N and U) to ensure sufficiently large sample sizes in each weather type. The resulting time fraction spent in each type is shown in Figure 4.2 for the years 2001–2018. We reassigned each day originally assigned a hybrid type, where the new type was assigned at random using a weighted sample from the types making up the hybrid; e.g. for the n_{ANE} days of type ANE, $w_1 n_{ANE}$ are classified as A and $w_2 n_{ANE}$ are classified as NE, where w_1 is the proportion of pure anticyclonic days and w_2 is the proportion of pure NE days. The total number of days under each LWT are listed in Table 4.2.

Table 4.2: Total number of days under each Lamb weather type (LWT) for 2001–2018.

LWT	**Long name**	**Number of days**
A	Anticyclonic	1325
C	Cyclonic	910
N	Northerly	362
NE	North-easterly	160
E	Easterly	183
SE	South-easterly	277
S	Southerly	550
SW	South-westerly	999
W	Westerly	1084
NW	North-westerly	634
U	Unclassified	90

4.2.4 Distance to roads

Distances to types of roads were obtained from the GRIP global roads database (Meijer et al., 2018). Distances were calculated for each AURN station and were used as input features to the classifier model. This feature is particularly useful for stations without NO_X measurements, since NO_X, and subsequently ozone concentrations, are strongly affected by road emissions (Granier and Brasseur, 2003).

4.3 Methods

4.3.1 Gradient boosted tree classifier (LightGBM)

For our ML classification model, we use the LightGBM Python package (https://lightgbm.readthedocs.io/en/latest/Python-API.html). LightGBM is a gradient boosting framework that uses a tree-based learning algorithm for machine learning

classification and regression tasks. The framework is an ensemble of decision trees that learn the relationships between input features and the target variable. The trees are combined using a gradient boosting algorithm, where new trees are added to the ensemble in such a way as to correct errors made by the previous trees. It is designed to be efficient, scalable and highly accurate, and is particularly useful for high-dimensional data with a large number of input features (Ke et al., 2017).

4.3.2 Balancing classes

Balancing classes is necessary to prevent the classification model from being dominated by the greater information available for the largest class, resulting in poor classification performance of the smallest class (Fernández et al., 2018). We use oversampling of the minority class to balance the classes i.e., random sampling the smallest class to be the same size as the largest class. The total post-balancing sample size is therefore twice the size of the largest class; in the context of threshold exceedances the consequence of this is that, whilst the number of exceedances decreases with the threshold, the post-balancing sample size increases.

4.3.3 Driver experiments

For each measurement station, we categorise high-level ozone days as days where MDA8 ozone exceeds the 90^{th}, 99^{th} and 99.9^{th} percentiles of MDA8 ozone at that location. We use site-specific percentiles instead of absolute thresholds to determine what drives the relatively high levels of ozone at each location, since the distribution of ozone levels differs across measurement sites. Table 4.3 shows the means and standard deviations for each exceedance group above the station-specific 50^{th}, 90^{th}, 99^{th} and 99.9^{th} percentile thresholds. Whilst not extreme, the 50^{th} percentile provides a useful baseline for our results. All experiments are run by first extracting the exceedances at each site and then pooling across sites to create a single dataset. Further, as our focus is on the interpretability of the trained ML model, not the ability to extrapolate to unseen cases, these are within-sample experiments. Table

4.4 lists the input features used for the classification experiments in Section 4.4.2.

Table 4.3: Means and standard deviations for MDA8 ozone measurements above station-specific 50^{th}, 90^{th}, 99^{th} and 99.9^{th} percentile thresholds and then pooled across stations.

Percentile	**Mean of exceedances ($\mu g/m^3$)**	**Standard Deviation of exceedances ($\mu g/m^3$)**
50^{th}	75.12	17.61
90^{th}	97.94	18.39
99^{th}	132.75	21.04
99.9^{th}	162.07	23.04

4.3.4 Model explanations (SHAP)

We use SHAP (SHapely Additive exPlanations) to determine the importance of the input features in the classification model, and hence the drivers of high-level ozone. SHAP provides a unified framework for interpreting the output of any machine learning model. The SHAP values represent the marginal contribution of each feature to the model output, relative to the expected value of the model output. By aggregating the SHAP values for each feature across a set of instances, it is possible to understand the overall importance of each feature in the model (Lundberg and S.-I. Lee, 2017). SHAP values can be positive or negative, with a positive value implying that the corresponding feature increases the likelihood of a threshold exceedance day, and a negative one implying an increased likelihood of a non-exceedance day. We define a 'driver' of high-level ozone as a feature with a positive SHAP value, and an 'inhibitor' as a feature with a negative SHAP value.

4.3.5 Forecasting moderate health threshold exceedances

In the UK, the moderate health threshold for MDA8 ozone is set by the UK government at $100\,\mu g/m^3$. We propose an ML-based health threshold exceedance

Table 4.4: Input features used to train the classifier models in Section 4.4.2.

Input feature	**Data source**
Year	Measurements
Month	Measurements
Latitude	Measurements
Longitude	Measurements
QVAPOR (water vapor mixing ratio)	WRF
SWDOWN (downward short-wave flux at ground surface)	WRF
T2_min (daily minimum 2m temperature)	WRF
T2_max (daily maximum 2m temperature)	WRF
PBLH (planetary boundary layer height)	WRF
PSFC (surface pressure)	WRF
HGT (terrain height)	WRF
Wind_speed	WRF
Wind_angle	WRF
Distance to motorways	GRIP
Distance to primary roads	GRIP
Distance to secondary roads	GRIP
Distance to tertiary roads	GRIP
Distance to local roads	GRIP

forecasting model, and perform four experiments on years with a high number of exceedances in the AURN data: 2003, 2006, 2008 and 2018.

The forecasting model is trained across the full network of stations, using all data prior to the forecasting year. We use 43 predictor features comprising all from Table 4.4 excepting distance to secondary and local roads – excluded due to their negligible impact on model output - and the additional features listed in Table 4.5. Unlike the models used to identify drivers (see above), we now include various lagged input features, including lagged measurements of MDA8 ozone from as recent

as the day before prediction. The parameters for the LightGBM model were tuned manually and set as `n_estimator=10000`, `colsample_bytree=0.87`, `max_depth=410`, `min_child_samples=213`, `min_child_weight=10.0`, `num_leaves=47`, `reg_alpha=7`, `reg_lambda=50`. All forecasting experiments are out-of-sample classification as the model has not been trained on data for the forecasting year.

4.3.6 Evaluation metrics

To evaluate the classifier models, we use three standard ML classification metrics: precision, recall and F1-score. Precision measures the accuracy of positive predictions and is calculated as $Precision = \frac{TP}{TP+FP}$, where TP is the number of correctly forecast exceedances and FP the number of falsely predicted exceedances. Recall of the exceedance class measures the ability of the model to correctly identify exceedance samples and is calculated as $Recall = \frac{TP}{TP+FN}$, where TP is as above and FN is the number of exceedances incorrectly classed as non-exceedances; an analogous measure can be calculated for the non-exceedance class as the ratio of correctly classified non-exceedances to the total number of observed non-exceedances. The F1-score is the harmonic mean of precision and recall, and summarises the overall performance of the model; it is defined $F1\text{–}score = \frac{2(Precision*Recall)}{Precision+Recall}$.

Note that the alternative evaluation metrics, hit rate and false alarm rate, which are often used in air quality forecasting literature, can both be derived from the classification metrics described above. The hit rate is the proportion of exceedances captured by the forecasting model and is the equivalent of recall of the exceedance class. The false alarm rate is the proportion of misidentified exceedance days (predicting an exceedance when none occurred) and is calculated as $1 - Recall_{exceedances}$.

Table 4.5: Additional input features to classifier models for forecasting health threshold exceedances in Section 4.5.

Input feature	**Data source**
Day of month	Measurements
Day of week	Measurements
1-day lagged MDA8 ozone	Measurements
2-day lagged MDA8 ozone	Measurements
3-day lagged MDA8 ozone	Measurements
4-day lagged MDA8 ozone	Measurements
10-day lagged MDA8 ozone	Measurements
14-day lagged MDA8 ozone	Measurements
1-day lagged daily mean ozone	Measurements
2-day averaged lagged daily mean ozone	Measurements
1-day lagged daily mean NOX	Measurements
3-day averaged lagged daily mean NOX	Measurements
7-day averaged lagged daily mean NOX	Measurements
1-day lagged T2_max	WRF
2-day lagged T2_max	WRF
1-day lagged T2_min	WRF
2-day lagged T2_min	WRF
1-day lagged QVAPOR	WRF
2-day lagged QVAPOR	WRF
1-day lagged SWDOWN	WRF
3-day average lagged SWDOWN	WRF
1-day lagged wind angle	WRF
1-day lagged wind speed	WRF
1-day lagged PSFC	WRF
3-day averaged lagged PSFC	WRF
1-day lagged PBLH	WRF
3-day averaged lagged PBLH	WRF

4.4 Finding drivers of high-level ozone using a machine learning classifier

4.4.1 Lamb weather types (within-sample classification)

Table 4.6 shows the results obtained when using only LWTs to classify MDA8 ozone into exceedances or non-exceedances of station-specific 50^{th}, 60^{th}, 70^{th}, 80^{th}, 90^{th}, 95^{th}, 99^{th} and 99.9^{th} percentile MDA8 ozone. The classifier was trained separately for each percentile. Average metrics only are reported since average and macro metrics are equivalent when class sizes are balanced. We find that LWTs are a better predictor of classifying high percentile ozone, with the average F1-score increasing from 0.51 for the 50^{th} to 0.71 for the 99.9^{th} percentile exceedance classifications. Average precision and recall scores also increase with classification threshold, from 0.53 to 0.73 and 0.53 to 0.72, respectively, for the 50^{th} and 99.9^{th} percentiles. Model performance still improves for the higher percentile thresholds despite the smaller number of threshold exceedances and the larger size of the balanced dataset required to compensate for this, which would typically lead to a poorer model fit. To evaluate this, we run the same experiments with equal sample sizes of 529,298 (the smallest sample size of the original experiments), with results shown in Table 4.7. We find similar improvement in performance as the percentile threshold increases, with average F1-score increasing from 0.52 for the 50^{th} to 0.71 for the 90^{th} percentiles, respectively. Therefore, we find LWTs explain high-level ozone behaviour better than general behaviour.

Our subsequent analysis focuses on classifying exceedances of the 50^{th}, 90^{th}, 99^{th} and 99.9^{th} site-specific percentiles only. Figure 4.3 shows the SHAP values for these four percentile thresholds. The E LWT is the greatest driver in the 50^{th} percentile classification; however, all SHAP values are relatively small compared to the higher percentile experiments, further suggesting that LWTs are a better predictor of high-level ozone than general behaviour. The E LWT is also the greatest driver for the 90^{th} percentile experiment, whereas SE is the greatest driver for the highest events

Table 4.6: Classification results of classifying MDA8 ozone above each measurement stations 50^{th} to 99.9^{th} percentile of MDA8 ozone using LWTs as input features. Class 0 are the non-exceedance days and class 1 are the exceedance days.

Percentile (sample size - balanced)	**Average Precision [class 0 | class 1]**	**Average Recall [class 0, class 1]**	**Average F1-score [class 0, class 1]**
50^{th} (529,298)	0.53 [0.52 | 0.54]	0.53 [0.68 | 0.37]	0.51 [0.59 | 0.44]
60^{th} (635,224)	0.53 [0.53 | 0.53]	0.53 [0.58 | 0.48]	0.53 [0.55 | 0.51]
70^{th} (740,720)	0.55 [0.53 | 0.56]	0.54 [0.72 | 0.37]	0.53 [0.61 | 0.45]
80^{th} (846,130)	0.57 [0.56 | 0.58]	0.57 [0.63 | 0.50]	0.57 [0.60 | 0.54]
90^{th} (951,408)	0.62 [0.62 | 0.62]	0.62 [0.63 | 0.61]	0.62 [0.62 | 0.61]
95^{th} (1,004,066)	0.66 [0.67 | 0.65]	0.66 [0.62 | 0.70]	0.66 [0.65 | 0.67]
99^{th} (1,046,088)	0.71 [0.76 | 0.67]	0.71 [0.61 | 0.80]	0.70 [0.68 | 0.73]
99.9^{th} (1,055,562)	0.73 [0.77 | 0.68]	0.72 [0.61 | 0.82]	0.71 [0.68 | 0.74]

occurring above the 99^{th} and 99.9^{th} percentiles. This is not unexpected: easterly flow transports air masses from mainland Europe that typically have elevated levels of ozone precursor pollutants, such as NO_X and VOCs (e.g. Pope et al., 2016). NW is the most important inhibitor of exceedances for all three high percentiles. This is consistent with our understanding that westerly flow typically brings cleaner maritime air from the Atlantic to over the UK. Other key inhibitors in high-level

Table 4.7: Classification results of classifying MDA8 ozone above each measurement stations 50^{th} to 99.9^{th} percentile of MDA8 ozone using LWTs as input features, and using a consistent sample size across the experiments. Class 0 are the non-exceedance days and class 1 are the exceedance days.

Percentile (529,298 samples)	**Average Precision [class 0 | class 1]**	**Average Recall [class 0 | class 1]**	**Average F1-score [class 0 | class 1]**
50^{th}	0.53 [0.52 | 0.54]	0.53 [0.68 | 0.37]	0.52 [0.59 | 0.44]
60^{th}	0.53 [0.53 | 0.53]	0.53 [0.58 | 0.48]	0.53 [0.55 | 0.51]
70^{th}	0.55 [0.53 | 0.56]	0.54 [0.72 | 0.37]	0.53 [0.61 | 0.45]
80^{th}	0.57 [0.56 | 0.58]	0.57 [0.63 | 0.50]	0.57 [0.60 | 0.54]
90^{th}	0.62 [0.62 | 0.62]	0.62 [0.63 | 0.61]	0.62 [0.62 | 0.61]
95^{th}	0.66 [0.67 | 0.65]	0.66 [0.62 | 0.69]	0.66 [0.65 | 0.67]
99^{th}	0.71 [0.76 | 0.67]	0.71 [0.61 | 0.80]	0.70 [0.68 | 0.73]
99.9^{th}	0.73 [0.78 | 0.68]	0.72 [0.61 | 0.82]	0.71 [0.68 | 0.74]

ozone include W, C, N and SW LWTs.

In contrast to the results found by Pope et al. (2016), we do not find anticyclonic conditions to be a key driver of high-level ozone, most likely due to differences in the data used. Specifically, MDA8 ozone data across the entire year is used here instead of the summertime noon ozone data used by Pope et al. (2016). To confirm this, we performed the 99.9^{th} percentile classification using approximately the same subset of

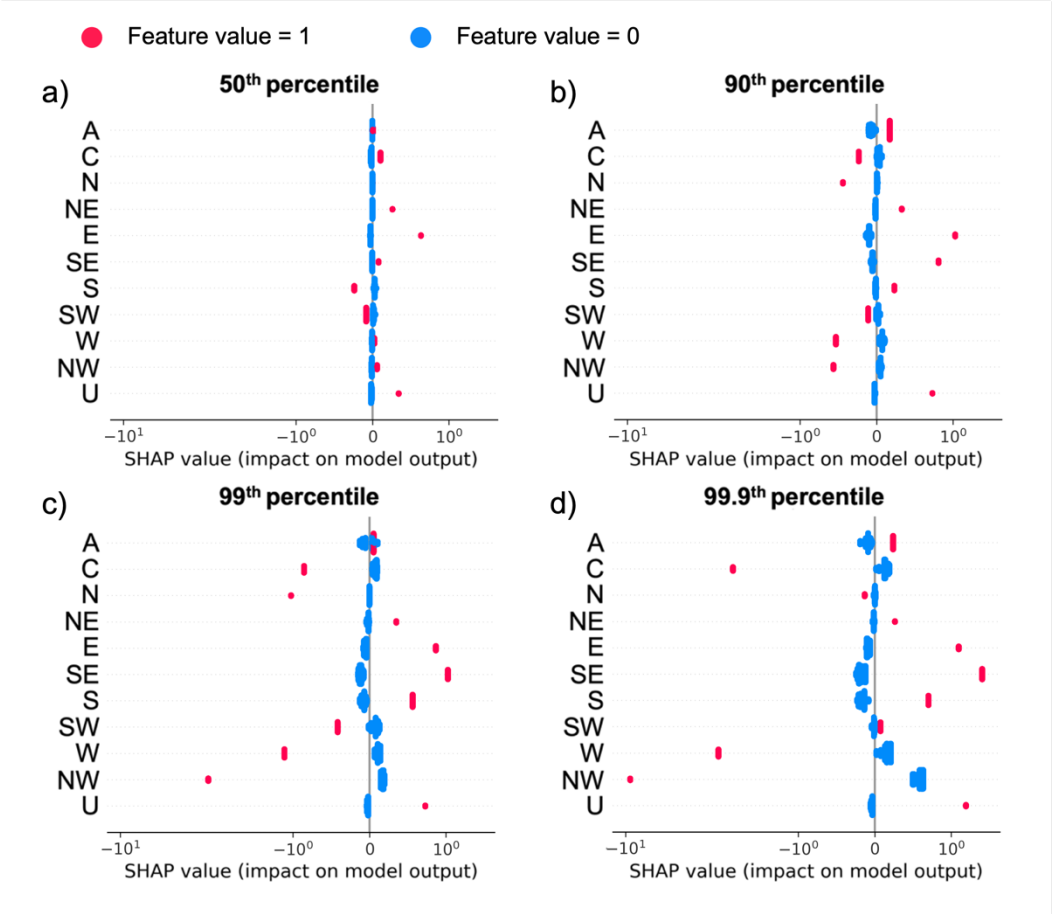

Figure 4.3: SHAP values for the 50^{th}, 90^{th}, 99^{th} and 99.9^{th} percentile classification. Red dots are days which are the LWT, blue dots are days which are not the LWT. Positive SHAP values have a positive impact on model output i.e., contributing to threshold exceedance days.

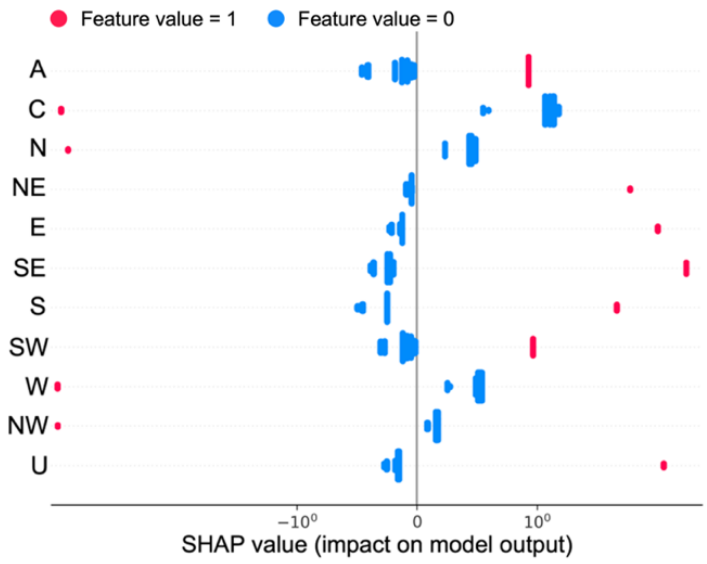

Figure 4.4: SHAP values for the 99.9$^{\text{th}}$ percentile classification, using only data from April–Sept 2006–2010 at background locations. Red dots are days which are the LWT, blue dots are days which are not the LWT. Positive SHAP values have a positive impact on model output i.e., contributing to threshold exceedance days.

our data as used by Pope et al. (2016); i.e., data from only background locations for April–September over 2006–2010. Using this subset, we find anticyclonic conditions to be a driver of very-high level ozone as seen in Figure 4.4. Further, Pope et al. (2016) found that under anticyclonic conditions, UK surface ozone ranges from 50 to 80 µg/m^3, whereas larger concentrations of between 60 and 100 µg/m^3 are sampled under south-easterly flow regime. As noted above, our results also show that the highest concentrations of ozone are driven mostly by easterly weather regimes. The LWT drivers of high-level ozone that we identify are the types that occur more infrequently compared with the most common LWTs, namely A, W, SW and C (see Table 4.2).

4.4.2 Determining meteorological and temporal drivers (within-sample classification)

While LWTs provide a high-level picture of the meteorological drivers of high-level ozone, there is a risk that they oversimplify the complex interactions between the atmosphere and ozone. Further, since they are defined on a country-level scale, the resulting model lacks the ability to identify changes in the behaviour of high-level ozone that occur on higher resolution spatial scales. This is a fundamental flaw when it is well known that ozone levels and production vary over the country (e.g. Pope et al., 2016). To address this, we show results for a classification that makes use of a much larger, and more detailed, set of input variables.

As in Section 4.4.1, we consider classifications of the 50^{th}, 90^{th}, 99^{th} and 99.9^{th} percentiles but now also use ML classification models trained on the input features listed in Table 4.5. Table 4.8 shows the classification metrics obtained when using only meteorological variables as input features. Based on these metrics, the new models are a substantial improvement over the LWTs classifiers, with all three metrics higher than the comparable values in Table 4.6. For both models the lowest average precision, recall and F1-scores all occur for the 50^{th} percentile, with all three metrics being 0.81 for the meteorological classifier compared to 0.53, 0.53 and 0.52, respectively, for the LWT classifier. Improvement continues for the higher percentile classifications, with perfect average precision, recall and F1-score for the 99.9^{th} percentile classifications. We conclude that the greater detail available in the meteorological input features better explains ozone behaviour, particularly high-level ozone behaviour, than using LWTs alone.

Figure 4.5 shows the five most important features for the four classifier models. Month is the most important feature for the 50^{th} and 90^{th} percentile threshold classifiers, and the second most important feature for the 99^{th} percentile, reflecting the strong seasonal signal in ozone concentrations. Daily maximum 2m temperature, T2_max, is also key as the most important feature for the two highest percentile the second most for the 90^{th} percentile and the fifth most for the 50^{th} percentile. This

Table 4.8: Classification results of classifying MDA8 ozone above each measurement stations 50^{th} to 99.9^{th} percentile of MDA8 ozone using meteorological, spatial and temporal variables as input features. Class 0 are the non-exceedance days and class 1 are the exceedance days.

Percentile (number of samples)	Average Precision [class 0 | class 1]	Average Recall [class 0 | class 1]	Average F1-score [class 0 | class 1]
50^{th} (529,298)	0.81 [0.81 | 0.82]	0.81 [0.82 | 0.80]	0.81 [0.82 | 0.81]
90^{th} (951,408)	0.87 [0.90 | 0.84]	0.86 [0.82 | 0.90]	0.86 [0.86 | 0.87]
99^{th} (1,046,088)	0.97 [0.99 | 0.95]	0.97 [0.95 | 0.99]	0.97 [0.97 | 0.97]
99.9^{th} (1,055,562)	1.00 [1.00 | 0.99]	1.00 [0.99 | 1.00]	1.00 [1.00 | 1.00]

is consistent with our understanding that very high-level ozone episodes co-occur with the warmest temperatures (e.g. Gouldsbrough et al., 2022). We see also that the daily minimum 2m temperature, T2_min, is the third and second most important feature for, respectively, the 99^{th} and 99.9^{th} percentile classifications, but fails to make the top five for the lower two thresholds. We believe this to be due to the highest ozone events mostly occurring during warm spells, when daily minimum temperatures are atypically high. Related to this, shortwave downward radiation flux from the sun, SWDOWN, is either the third or fourth most important feature across all percentiles, demonstrating the importance of sunlight in the net photochemical production of ozone. Lastly, planetary boundary layer height, PBLH, is the second most important feature in the baseline 50^{th} percentile classification which we suspect is due to the role of low-level jets and nocturnal mixing processes (Klein et al., 2014).

Figure 4.6 shows the relationships between the input features year, month,

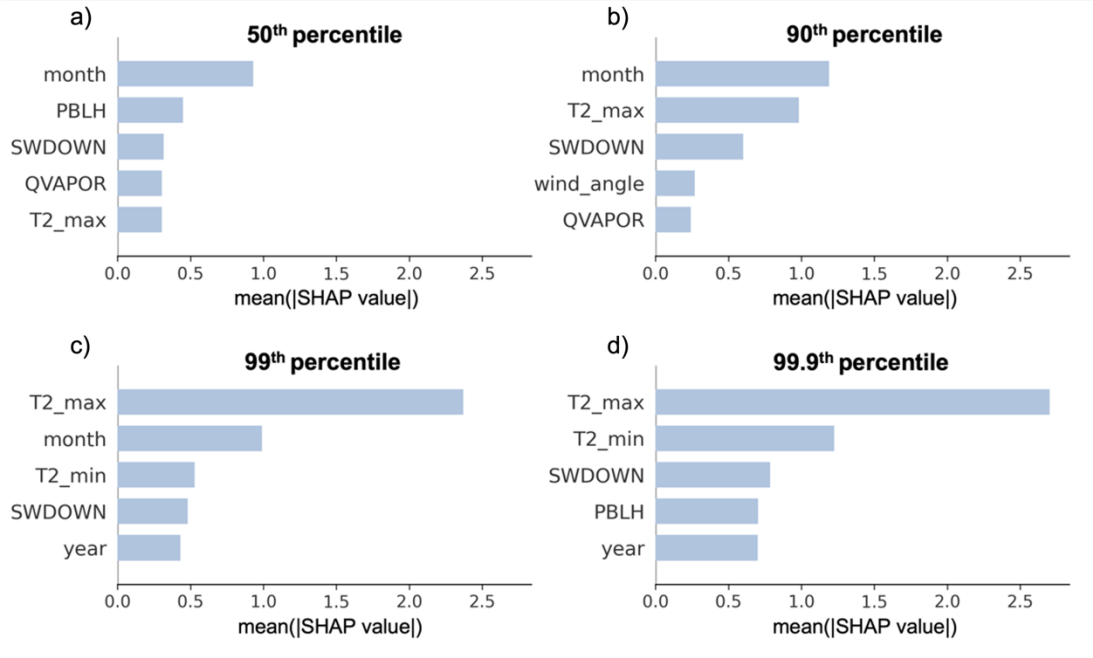

Figure 4.5: The five most important features for the a) 50^{th} percentile threshold, b) 90^{th} percentile threshold, c) 99^{th} percentile threshold, and d) 99.9^{th} percentile threshold classification models fitted to all UK data.

T2_max, T2_min and their respective SHAP values. SHAP values for year are greatest for the 99th and 99.9th percentiles, peaking around 2003, a year in which Europe experienced a remarkable heatwave (e.g. Vieno, Dore, et al., 2010) and with a considerable number of ozone health threshold exceedances (as shown in Figure 4.8). Across all percentiles, SHAP values for month peak in May and are lowest at the end of autumn / beginning of winter, reflecting the well-known seasonal cycle of UK ozone (Diaz et al., 2020). The smallest monthly SHAP values are seen in the highest percentile classifications, suggesting that month is a more important indicator for exceedances of 50th and 90th percentiles than for the highest ozone events above the 99.9th percentile. There is a clear positive relationship between T2_max and SHAP values. The 50th percentile experiment shows the weakest relationship, with the SHAP values being positive for values above ∼18 °C and zero for lower temperatures, and a similar crossing point is found for the 90th percentile experiment. The 99th and 99.9th percentile experiments show higher crossing points, at ∼20 and ∼22 °C, further illustrating that the highest ozone events typically occur during the warmest weather. For T2_min, a crossing point for all percentiles is found at ∼12 °C. However, the form and magnitude of the relationship between SHAP value and T2_min relationship is different across the percentiles: the 50th and 90th percentiles show an almost identical weak positive relationship, while the 90th and 99.9th show a stronger positive relationship. Low daily minimum temperature is an inhibitor of high-level ozone whilst high daily minimum temperature is a driver. Daily minimum temperature has little impact on lower ozone levels with the key temperature driver being daily maximum temperature.

Figure 4.7 shows the SHAP value relationships with the remaining features identified as key drivers in Figure 4.5: PBLH, SWDOWN, QVAPOR and wind_angle. The SHAP relationship for PBLH differs across percentiles, with a clear positive trend for the 50th percentile, a weaker positive trend for the 90th percentile and negative trends for the two higher percentiles. The negative trend is strongest for the 99.9th percentile. This suggests that increases in boundary layer height led

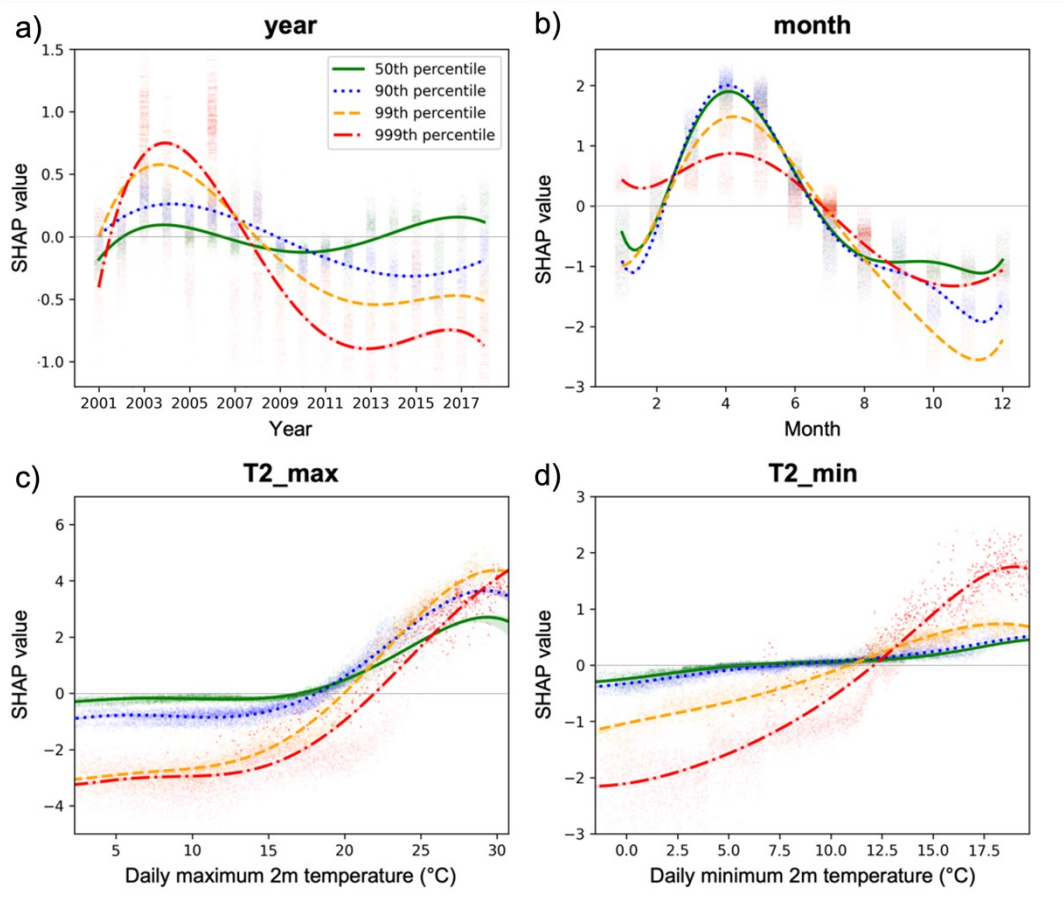

Figure 4.6: Scatterplots showing the relationships between SHAP value and feature value for a) year, b) month, c) T2_max, and d) T2_min. Locally Weighted Scatterplot Smoothing (LOWESS) curves are added to each plot for each percentile classifications; 50^{th} percentile (green, solid line), 90^{th} percentile (blue, dotted line), 99^{th} percentile (yellow, dashed line) and 99.9^{th} percentile (red, dashed-dotted line).

to increases in moderate to high ozone concentrations but reduced incidences of very high-level ozone. While a positive correlation between PBLH and ozone have been observed by a number of studies these relationships can be weak and the underpinning mechanism may represent both dynamics (e.g. mixing and dilution of airmasses), deposition, and the chemistry/availability of ozone and its precursors (e.g. Reddy et al., 2012; Haman et al., 2014; Su et al., 2018). While our results support a complex ozone-PBLH relationship, interestingly the crossing point from non-exceedance to exceedance is ∼450m for all percentiles. SWDOWN shows a positive SHAP value relationship for all percentiles; however, greater short-wave flux is needed to lead to 99th and 99.9th percentile exceedances (∼200 and ∼225 W/m^2, respectively) compared to 50th and 90th percentile exceedances (∼150 W/m^2 for both). The SHAP value relationships for QVAPOR are the most consistent across percentiles than for any other variable, with a negative impact on exceedances after ∼0.03, demonstrating that higher humidity levels lead to lower concentrations of ozone. In line with this, humidity and ozone have been shown to be anticorrelated by numerous studies with the underlying mechanism linked to ozone dry deposition (e.g. Kavassalis and Murphy, 2017). Finally, we consider the relationship between SHAP value and wind_angle. Wind_angle has the lowest impact on exceedances of the 50th percentile, and the greatest impact at the 99.9th percentile experiment. Wind angles between 50°and 160°, i.e., winds from the NE, E and SE,, have the greatest impact on 99.9th percentile threshold exceedances due to transported air masses from Europe (Tudor, 2022).

We lastly fit the classifier models separately to each UK region to assess consistency and identify any spatial trends in the feature importances (Table 4.4). From Table 4.9, we find that, for the 50th percentile, month is the most important driver in all regions; the same is true for most regions at the 90th percentile. This confirms the strong seasonal cycle in present in mean level ozone concentrations across the UK (Diaz et al., 2020). T2_max is the most important feature for the 99th and 99.9th percentiles for most regions, consistent with understanding of

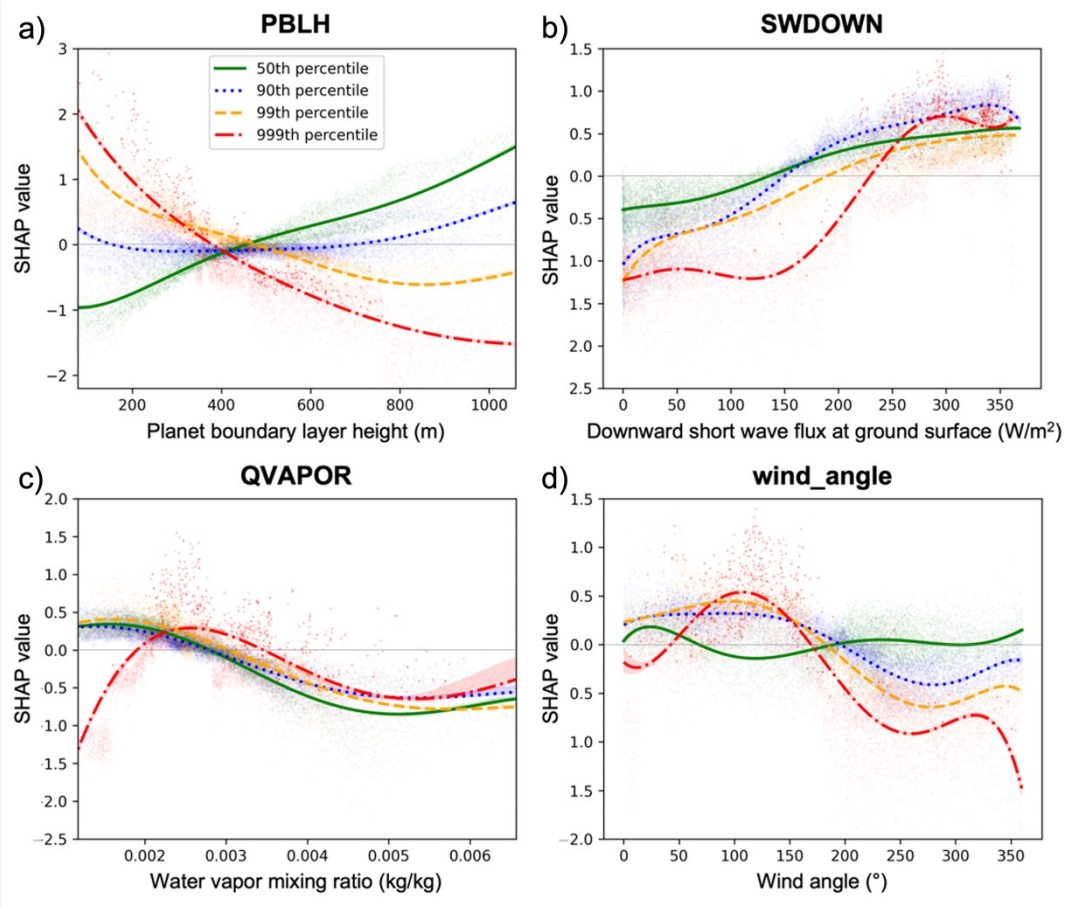

Figure 4.7: Scatterplots showing the relationships between SHAP value and feature value for a) PBLH, b) SWDOWN, c) QVAPOR, and d) wind_angle. Locally Weighted Scatterplot Smoothing (LOWESS) curves are added to each plot for each percentile classifications; 50^{th} percentile (green, solid line), 90^{th} percentile (blue, dotted line), 99^{th} percentile (yellow, dashed line) and 99.9^{th} percentile (red, dashed-dotted line).

the relationship between extreme ozone and temperature. We find the Eastern region to be the only region with SWDOWN as the most important feature in classifying the 99.9^{th} percentile ozone, suggesting that sunshine is most important in driving high-level ozone episodes for this region. Wind angle and PBLH are the most important features for North East Scotland in classifying 99^{th} and 99.9^{th} percentile ozone; however, there is only one measurement station in this region (Table 4.1) at Aberdeen and so caution is required in over-generalising this result. Similarly, the Highland region in the north of Scotland has three monitoring stations, and its highest-level ozone episodes (99.9^{th} percentile) are driven by wind angle, signifying that this region is more impacted by the transport of air masses than ozone formation.

4.5 Forecasting moderate health threshold exceedances

We now demonstrate the use of a ML classifier to forecast moderate health threshold exceedances of MDA8 ozone, i.e., exceedances of $100\,\mu g/m^3$. We use this threshold since at-risk individuals are recommended to reduce strenuous physical activity, particularly outdoors, when MDA8 ozone exceeds this level. We consider 2003, 2006, 2008 and 2018, as these years had higher than normal numbers of moderate health threshold exceedances (as shown in Figure 4.8). For each year, the ML model is trained only once using all previous years of data. The model is then used to make one-day-ahead operational forecasts, based on the previous day's pollution measurements and a combination of past and forecast meteorological variables, for each day of the year.

Table 4.10 shows the classification results from our UK-wide forecasting experiments. We report macro averages, the arithmetic mean of the class results, instead of weighted averages to avoid biasing the results towards the greatest sized class (non-exceedances days). The forecasting model is well balanced between precision

Table 4.9: Most important feature for each region and for each percentile threshold classification, fitted to data from each region.

Region	**50th percentile**	**90th percentile**	**99th percentile**	**99.9th percentile**
Central Scotland	month	month	T2_max	T2_max
East Midlands	month	T2_max	T2_max	T2_max
Eastern	month	month	T2_max	SWDOWN
Greater London	month	T2_max	T2_max	T2_max
Highland	month	month	T2_max	wind_angle
North East	month	month	T2_max	T2_max
North East Scotland	month	month	wind_angle	PBLH
North Wales	month	month	T2_max	T2_max
North West Merseyside	month	month	T2_max	T2_max
Northern Ireland	month	month	T2_max	T2_max
Scottish Borders	month	month	T2_max	T2_max
South East	month	T2_max	T2_max	T2_max
South Wales	month	month	T2_max	T2_max
South West	month	month	T2_max	T2_max
West Midlands	month	T2_max	T2_max	T2_max
Yorkshire Humberside	month	month	T2_max	T2_max

and recall across all years. F1-scores are highest for 2006, at 0.86, and still relatively high for 2003, 2008 and 2018 at 0.83, 0.84 and 0.84, respectively. The larger number of training samples in the 2018 forecasting experiment seem to have little impact on model performance compared with 2006 and 2008; however, model performance is improved for 2006 onwards when compared to 2003 with only 2 years of training data. The hit rates of forecasting moderate health threshold exceedances are good at 0.69, 0.78, 0.73 and 0.71 for 2003, 2006, 2008 and 2018, respectively. The false

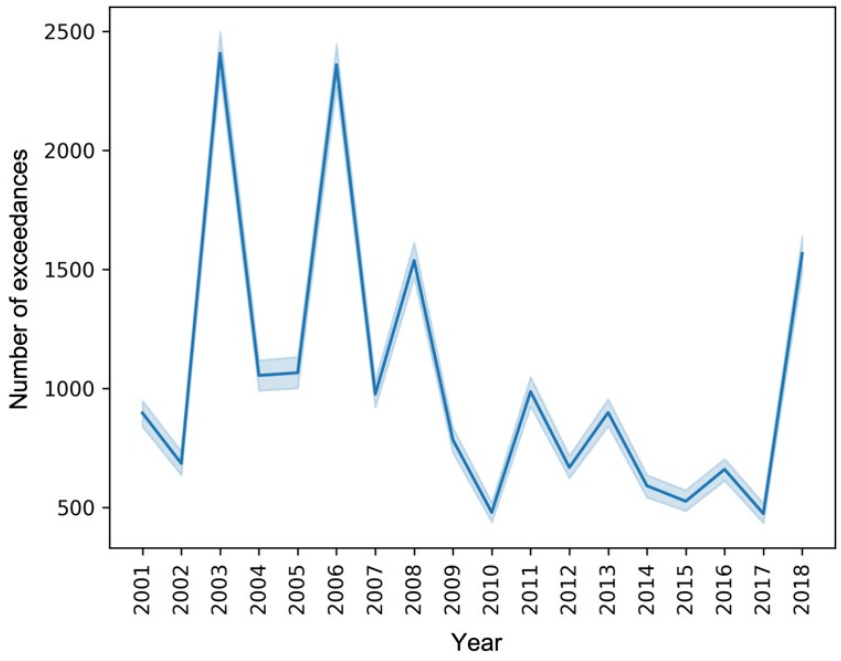

Figure 4.8: Total number of exceedances across the measurement network for each year.

alarm rates are low for all experiment years: 0.03, 0.02, 0.02 and 0.02 for 2003, 2006, 2008 and 2018, respectively.

Finally, we consider the importance of the input features in predicting moderate health threshold exceedances of MDA8 ozone, as shown in Figure 4.9 for a model trained on years 2001–2018 with balanced classes. Here, T2_max is the most important predictor; 1-day lagged MDA8 ozone is the second most important predictor, showing the temporal persistence of ozone concentrations; and month is the third most important feature, followed by QVAPOR, year, T2_min, 1-day lagged daily mean ozone, SWDOWN, PSFC and wind_angle. Whilst the most important features are similar to those seen in Figure 4.5, there are differences in their relative importance. This is most likely due to using an absolute threshold of $100\,\mu g/m^3$ instead of relative definitions of high-level ozone (e.g. percentile thresholds) as well as the inclusion of lagged pollutant measurement data. The features of least importance are 2-day averaged lagged daily mean ozone, 3-day

Table 4.10: Classification results from forecasting 2003, 2006, 2008 and 2018 moderate health threshold exceedances of MDA8 ozone. Class 0 are the non-exceedance days and class 1 are the exceedance days.

Year [0 samples | 1 samples]	**Years of training data**	**Macro Average Precision [class 0, class 1]**	**Macro Average Recall [class 0, class 1]**	**Macro Average F1-score [class 0, class 1]**
2003 [27,481 | 2,408]	2	0.83 [0.97 | 0.68]	0.83 [0.97 | 0.69]	0.83 [0.97 | 0.69]
2006 [29,915 | 2,360]	5	0.85 [0.98 | 0.72]	0.88 [0.98 | 0.78]	0.86 [0.98 | 0.75]
2008 [27,910 | 1,537]	7	0.83 [0.99 | 0.68]	0.86 [0.98 | 0.73]	0.84 [0.98 | 0.71]
2018 [24,574 | 1,566]	17	0.84 [0.98 | 0.70]	0.85 [0.98 | 0.71]	0.84 [0.98 | 0.71]

average lagged SWDOWN and 1-day lagged SWDOWN, with SHAP values of $\sim$ 0.1, 0.12 and 0.13. These features add little additional information to the model that is not already present in the other features.

Overall, our machine learning classification approach, coupled with SHAP's interpretability capabilities, provides a tool for identifying and understanding the drivers of high-level ozone concentrations in the UK. Further, treating the forecasting of health threshold exceedance days as a machine learning classification task shows promise as an alternative to current methods for forecasting ozone episodes.

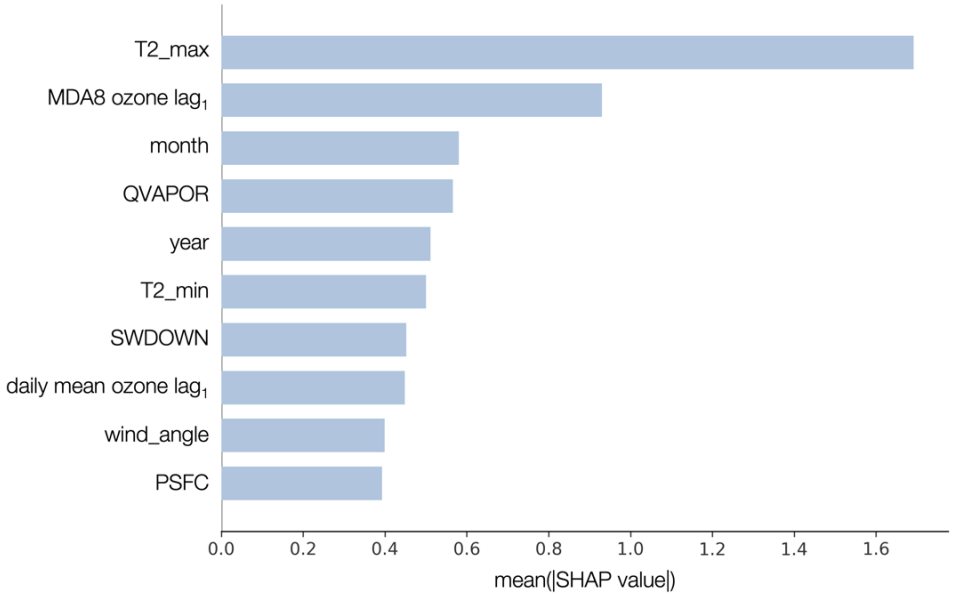

Figure 4.9: The ten most important input features to the forecasting classification model.

4.6 Conclusion

Our study presents a novel machine learning classification approach for the identification of factors contributing to high-level ozone concentrations in the UK, at both a national and regional level. The classifier uses input variables known to affect ozone concentrations. These variables vary either over space and time (e.g. temperature, wind speed and direction, and atmospheric pressure), over space only (e.g. distance from major roads), or over time only (e.g. month of the year). By training the classifier on appropriate datasets, we were able to identify both drivers and inhibitors of high-level ozone events and to predict with very high accuracy the occurrence of such events. We used SHAP (SHapley Additive exPlanations), a powerful machine learning interpretability tool, to assess the impact of each predictor variable on the model's output, allowing us to rank the importance of these factors and determine the extent to which they influence high-level ozone. Additionally, SHAP enabled us to identify the interactions between these predictor variables, providing a more nuanced understanding of the underlying drivers of high-level

ozone episodes in the UK and how these drivers change as the magnitude of the ozone event is increased from moderate to high to very high (defined by the UK's air quality index).

As a proof of concept, our first classifier used Lamb weather types (LWTs) to classify maximum daily 8-hour average (MDA8) ozone concentrations into exceedances or non-exceedances of various percentile thresholds. We find that LWTs are a better predictor of high-level ozone than they are of mean behaviour, with average F1-score increasing from 0.51 to 0.71 for the 50^{th} and 99.9^{th} percentile exceedance classifications. Weather patterns with south-easterly and easterly flows are the greatest drivers for the 99^{th} and 99.9^{th} percentile classifications, indicating an association with high-level ozone events due to the transport of polluted air masses from mainland Europe.

Whilst classifiers based on LWTs are helpful in demonstrating that a data-driven model, when provided with suitable input features, can reproduce observed events with a reasonable degree of accuracy, the LWTs contain only synoptic scale information. To improve classification performance, to a level of accuracy that the model could be used as an operational forecasting tool to inform public air quality warnings, required us to include more input features including meteorological, temporal and spatial variables. The results of the updated classifiers showed that the models trained with meteorological, spatial and temporal input features outperformed the LWT classifiers, with higher average precision, recall, and F1-scores for all percentile classifications. The SHAP values showed that month is the most important driver of 50^{th} and 90^{th} percentile exceedances, while daily 2m maximum temperature is the most important feature for driving very high-level ozone above the 99^{th} and 99.9^{th} percentiles. The importance of year in driving high-level ozone is greatest in 2003, a year with a high number of ozone episodes. We also find a strong seasonal cycle present with the likelihood of an exceedance peaking in spring; we also find that the importance of season decreases as the percentile threshold increases. Daily 2m minimum temperature replaces season as

a driving factor in the highest percentile experiments, possibly because the highest-level ozone events typically occur during periods when night-time temperatures are unusually high. Daily 2m maximum has a clear positive relationship with the likelihood of a threshold exceedance with the crossing point temperature between an exceedance and non-exceedance day increasing for the higher percentile classifications, showing that it is elevated temperatures (temperatures greater than 20 °C) that lead to the most severe ozone episodes. The relationship between planetary boundary layer height and its SHAP values is negative for the 99^{th} and 99.9^{th} percentile classifications, suggesting that higher boundary layer height allow for better dispersion of ozone concentrations.

Finally, we propose a machine learning classification approach to forecast moderate health threshold exceedance days of MDA8 ozone. We evaluated the capability of the classifier in years with raised moderate health threshold exceedance days: 2003, 2006, 2008, and 2018. The experiments use previous day's pollution measurements as a proxy for one-day-ahead operational forecasting. We find that we can forecast health threshold exceedance days with moderate accuracy, achieving hit rates (rate of capturing exceedance days) of 0.69 to 0.78 in each forecasting experiment. We also achieve low false alarm rates of 0.02 to 0.03. Daily maximum 2m temperature is the most important predictor feature, followed by 1-day lagged MDA8 ozone, which shows the temporal dependence of ozone concentrations. We anticipate further study extending these forecasting experiments to include additional relevant features, such as information from neighbouring stations, and directly comparing the performance of the ML model against CTM forecasts.

4.7 Acknowledgements

We are grateful for the assistance of Iain Walmsley and Massimo Vieno in supplying WRF data.

4.8 Funding statement

LG was supported by the UK Engineering and Physical Research Council (EPSRC) for a PhD studentship (EP/R513076/1, project reference 2353903). RH was supported by a UK Natural Environment Research Council (NERC) Independent Research Fellowship (NE/N014375/1). EE and PJY were supported by the EPSRC-funded Data Science of the Natural Environment project (EP/R01860X/1).

4.9 Author contributions

Conceptualization: L.G; E.E.; R.H. Methodology: L.G; E.E; R.H. Data curation: L.G. Data visualisation: L.G. Writing original draft: L.G; E.E; R.H.; P.Y. All authors approved the final submitted draft.

Chapter 5

Conclusions

This thesis explored the application and development of data science methods to address gaps in the knowledge and understanding of UK surface-level ozone. By using sophisticated statistical and machine learning techniques, the work in this thesis has improved understanding of the probability, magnitude, frequency, spatial distribution, trends and drivers of high-level ozone, and this knowledge has been used to create trustworthy forecasts of high-level ozone events. The contributions are motivated by the need to better understand the risk and drivers of high-level ozone across the UK in a changing climate, and to develop more effective methods of modelling surface level ozone. Despite the relatively rich monitoring data available in the UK, studies of UK air quality have yet to fully benefit from recent advances in data science approaches. The aims of the thesis have been addressed as follows:

1. **To quantify extreme ozone events evaluating their magnitude, frequency, and likelihood and examining how these events have changed over time**: Chapter 2 presents an extreme value analysis (EVA) of UK surface-level ozone using a temperature-dependent extremes model. We estimated the magnitude, frequency, and likelihood of extreme ozone events at in-situ measurement stations across the UK and analysed how these properties have changed over time. From 2010 to 2019, we found that over 90% of sites have a 1-year MDA8 ozone return level greater than the 'moderate' ozone

threshold of 100 µg/m^3, and that one-third of sites are expected to exceed the UK government's air quality objective of not exceeding 100 µg/m^3 for more than 10 days a year. Further, the probability of ozone levels exceeding 100 µg/m^3 is between <0.1% and 5.4%, depending on location, with average levels of 2.7% in rural areas and 1.6% in urban areas. An analysis into decadal changes of extreme ozone risk showed that the likelihood of extreme ozone episodes in the UK has significantly decreased over time, with 1-year return levels in the 1980s now being similar to 10-year return levels at present. Nevertheless, the strong positive correlation between temperature and ozone exceedances means that hotter summers due to climate change could undermine these improvements.

2. **Determining the variability of ozone in both space and time and identifying long-term trends that are relevant to policymaking**: Chapter 3 presents a novel machine learning (ML) methodology to downscale EMEP4UK ozone output from a 5×5 km to 1×1 km resolution over the 2001–2018 period. The methodology is based on a gradient boosted tree and the downscaled product better represents measured ozone, particularly at high levels, and the method is applicable to other CTMs. An analysis was subsequently performed on the downscaled surface, original CTM surface, and measurement data, to quantify both the current state of UK ozone, including regional variability, adherence to government objectives, and trends over time. The results showed that the unadjusted CTM overestimated surface-level ozone concentrations in the UK and that the downscaled surface improved the representation of high-level ozone. This difference in high-level ozone representation results in vastly altered estimates of the proportion of the UK failing the government objective to not exceed 100 µg/m^3 more than 10 times per year: 99% of the UK fails this objective in the EMEP4UK surface using a 2014–2018 average, compared to just 27% in the downscaled surface. A sensitivity analysis into three NO_X reduction scenarios showed that moderate

reductions in UK NO_X emissions (i.e., 20% or 40%) will lead to increased ozone concentrations for substantial portions of the UK. Severe reductions in NO_X (i.e., 80%) will lead to decreased ozone concentrations in some parts of the UK, however urban areas will be most affected by increased concentrations.

3. **To assess the association of high concentrations of ozone with various meteorological, spatial and temporal factors**: Chapter 4 presents a ML classification-based experimental analysis into the drivers of high concentrations of measurement ozone using gradient boosted trees. The impacts of synoptic weather, meteorological, spatial, and temporal features on high-level ozone were identified. While challenges remain in understanding the underpinning processes controlling ozone concentrations, the proposed ML approach efficiently identifies the most important input features. Daily maximum temperature is found to be a key driver of high-level ozone, consistent with similar associations reported in other parts of the world (Bloomer, Stehr, et al., 2009; Gu et al., 2020; Zheng et al., 2023).

4. **To produce data-driven forecasts for the occurrence of high concentrations of ozone**: Chapter 4 also presents a ML classification method to forecast the presence or absence of high-level ozone events, demonstrating a potential for operational utility. The results showed that while synoptic weather patterns (Lamb Weather Types) are useful in determining the presence of high-level ozone, more information can be gained from more specific meteorological variables.

The research presented in this thesis advances our understanding of high-level ozone in the UK by quantifying trends and characteristics of extreme ozone events. Additionally, the association of meteorological, spatial and temporal factors with varying definitions of high-level ozone has been examined, and the potential of machine learning for real-time operational forecasts has been shown.

While this thesis presents significant contributions to the understanding of

high-level ozone events in the UK and demonstrates where advanced statistical and ML methods may be of particular use, it is important to acknowledge the limitations of the research. Firstly, the research in this thesis is dependent on ozone measurement data throughout (principally from the UK's AURN network), which can lead to uncertainty in determining ozone behaviour across the UK due to the sparsity of the measurements in some regions, as well as the representativeness of measurement sites. In addition to missing data, inaccuracies may still be present despite the data being ratified, therefore any model fitted is subject to a degree of measurement uncertainty i.e., the AURN ozone monitors are calibrated against an ozone photometer to ensure a relative uncertainty of <15% (Stevenson et al., 2009). Secondly, the weather and atmospheric data used in Chapters 3 and 4 are from a relatively coarse gridded process-based model output, which was chosen due to the ease of implementation and lack of data processing needed. Observational weather and atmospheric data may provide a more accurate view of conditions at a specific location. However, observational data is sparse in some parts of the UK, and is not necessarily complete over time. Thirdly, none of the methods considered in this thesis directly model the spatial behaviour of ozone, rather the spatial structure is imposed indirectly through spatial covariates, as in Chapters 3 and 4. Finally, while machine learning has proved to be a useful tool for modelling ozone concentrations, it's essential to recognise its limitations. The effectiveness of machine learning models can be significantly influenced by the quality and quantity of available data. In the context of UK air quality networks, limitations may arise from gaps in monitoring, and the challenge of capturing the full complexity of atmospheric processes. Additionally, machine learning models are not immune to overfitting, especially when dealing with relatively small datasets.

The work in this thesis has highlighted several areas for future work; these are listed by chapter but note that there is some overlap between the methods that could be used to implement the extensions.

Firstly, an extension of the work in Chapter 2 would be to undertake a spatial

extreme value analysis. Currently the analysis is undertaken separately for each site; incorporating nearby measurement data, could provide more precise estimates of extreme ozone at measurement sites, and, by modelling the relationship between ozone levels at different sites, neighbouring information can be used to fill in gaps in measurement records. Furthermore, a spatial model permits the spatial intensity and extent of high-level ozone episodes to be estimated, allowing the identification of regions where ozone episodes are interconnected, e.g. examining the likelihood of an ozone episode at both location A and location B simultaneously.

Another possible extension to the extremes model proposed in Chapter 2 is to include additional covariates. This would then enable the estimation of extreme ozone risk conditional on other variables, e.g. concentrations of NO_X, humidity, time of day or year, or to test for long-term trends or anomalous years.

The downscaling framework and analysis in Chapter 3 could be expanded through integration of additional data sources into the ML-downscaling model. Examples of such sources include satellite data, additional CTMs and measurement data from community networks. Whilst this additional information could be useful in improving the accuracy of the downscaled output, care would be needed to ensure the reliability of the data, as it may give rise to fresh issues such as measurement inaccuracies, discrepancies in spatial and temporal resolution, or non-uniform spatial grids.

A second useful addition to the downscaling framework would be the inclusion of relevant precursor pollutant data such as measurement NO_X, which might improve the accuracy of the downscaling results. Also, it may no longer be necessary to include road variables in the downscaling model since these are known to be correlated with NO_X. Instead, NO_X data could be used as a more informative and direct proxy for traffic-related emissions. Including NO_X data within the downscaling framework would allow investigation of the impact of emission scenarios on UK surface ozone directly from the downscaling framework instead of relying on further CTM scenario runs, and further NO_X scenario analyses could be explored

as an extension to Chapter 3.

Lastly, we note that the proposed downscaling methodology can be applied to other gridded outputs, such as CTM and reanalysis products, and the high-resolution downscaled surface can be used to study the historical health implications of high-level ozone across the UK, e.g. using the downscaled surface as input into models that relate air quality and health outcomes (Pannullo et al., 2017).

The forecasting model in Chapter 4 could also be extended to include additional spatial information, particularly as ozone and its precursors may be transported from polluted regions. Unlike in Chapter 2, where we are interested in spatial extent at a given point in time, here we are interested in relating the occurrence of an ozone episode at one location, given meteorological conditions such as wind direction, to the likelihood of an ozone episode elsewhere.

An alternative to the point-specific forecasting model proposed in Chapter 4 is to apply the bias-correction downscaling methodology from Chapter 3 to CTM forecasts. By integrating the drivers of high-level ozone identified in Chapter 4 (e.g. temperature and lagged measurement data) into the bias-correction process, it would be feasible to produce accurate spatial forecasts of ozone. This would involve training a ML model to learn and correct biases from CTM forecasts using information from recent measurement data, and then using the trained ML model to bias correct CTM forecasts in real-time; this is similar to the statistical post-processing bias correction used by the Met Office (Neal et al., 2014).

A key priority of ozone forecasting is to provide trustworthy predictions of high-level events, i.e., ozone health threshold exceedances and episodes; this is known to challenge both CTM and ML-based forecasting methods. Consequently, further development of the ML-based forecasting model to better capture the full tail behaviour of ozone is essential in creating accurate data-driven forecasts of ozone episodes. Ideally this forecasting model will be capable of forecasting both the presence of high-level ozone episodes and the magnitude of the ozone concentrations. The methods considered in this thesis are not limited to the UK, but can be

applied to other countries where elevated levels of ozone are a pressing issue, such as China and India. The key requirement for this is that the country has one or more substantial ozone monitoring networks. By applying these methods in other countries, we can gain a better understanding of the factors contributing to high levels of ozone and improve our ability to forecast and mitigate potential ozone episodes in these locations. This could have a significant impact on public health and the environment in these countries, as well as contribute to global efforts to address the issue of ozone pollution.

As this thesis has demonstrated, the interdisciplinary nature of data science can provide benefits to progress within atmospheric science. By combining expertise from different fields, we can gain a more complete understanding of the complex processes that govern our atmosphere. Moreover, the use of advanced data science techniques is complementary to traditional atmospheric science methods, enabling us to extract and fuse information from both new and existing data sources, and use this information to create more accurate and representative models. It is therefore crucial that we continue to collaborate across disciplines and embrace the opportunities that data science presents to advance our understanding of the atmosphere.

Appendix A

Supplementary material to Chapter 2

A.1 Supplementary tables

Table A.1: Information for each monitoring site used within our study.

				Ozone data record		**Ozone season (month)**	
Code	**Site Name**	**Site Type**	**Region**	**Start**	**End**	**Start**	**End**
CLL	Central London	Urban Background	Greater London	1972	1990	3	9
SIB	Sibton	Rural Background	Eastern	1973	ongoing	3	9
STE	Stevenage	Suburban Background	Eastern	1976	1994	3	9
HAR	Harwell	Rural Background	South East	1976	2015	3	9
BOT	Bottesford	Rural Background	East Midlands	1977	2016	3	9

Continued on next page...

				Ozone data record		**Ozone season (month)**	
Code	**Site Name**	**Site Type**	**Region**	**Start**	**End**	**Start**	**End**
WC	Wharleycroft	Rural Background	North West Merseyside	1985	1995	2	9
BUSH	Bush Estate	Rural Background	Central Scotland	1986	ongoing	2	7
ESK	Eskdalemuir	Rural Background	Scottish Borders	1986	ongoing	2	7
GDF	Great Dun Fell	Rural Background	North West Merseyside	1986	2016	2	8
AH	Aston Hill	Rural Background	North Wales	1986	ongoing	2	9
LH	Lullington Heath	Rural Background	South East	1986	ongoing	3	9
SV	Strathvaich	Rural Background	Highland	1987	ongoing	1	6
LN	Lough Navar	Rural Background	Northern Ireland	1987	ongoing	1	7
YW	Yarner Wood	Rural Background	South West	1987	ongoing	2	8
HM	High Muffles	Rural Background	Yorkshire Humberside	1987	ongoing	2	9
GLAZ	Glazebury	Rural Background	North West Merseyside	1988	ongoing	3	8
LB	Ladybower	Rural Background	East Midlands	1988	ongoing	2	8
BRI	London Bridge Place	Urban Background	Greater London	1990	1999	3	9
CLL2	London Bloomsbury	Urban Background	Greater London	1992	ongoing	3	9
NEWC	Newcastle Centre	Urban Background	North East	1992	ongoing	2	8
BEL2	Belfast Centre	Urban Background	Northern Ireland	1992	ongoing	2	7

Continued on next page...

				Ozone data record		**Ozone season (month)**	
Code	**Site Name**	**Site Type**	**Region**	**Start**	**End**	Start	End
BIRM	Birmingham Centre	Urban Background	West Midlands	1992	2009	3	8
CARD	Cardiff Centre	Urban Background	South Wales	1992	ongoing	3	9
ED	Edinburgh Centre	Urban Background	Central Scotland	1992	2003	2	9
LEED	Leeds Centre	Urban Background	Yorkshire Humberside	1993	ongoing	3	8
BRIS	Bristol Centre	Urban Background	South West	1993	2005	3	8
LIVR	Liverpool Centre	Urban Background	North West Merseyside	1993	2002	2	8
BIR2	Birmingham East	Urban Background	West Midlands	1993	2004	3	9
SOUT	Southampton Centre	Urban Background	South East	1994	ongoing	3	9
LEIC	Leicester Centre	Urban Background	East Midlands	1994	2013	3	9
HULL	Hull Centre	Urban Background	Yorkshire Humberside	1994	2002	3	9
BEX	London Bexley	Suburban Background	Greater London	1994	2007	3	9
SWAN	Swansea	Urban Background	South Wales	1994	2006	2	9
MID	Middlesbrough	Urban Industrial	North East	1995	ongoing	3	9
MAN3	Manchester Piccadilly	Urban Background	North West Merseyside	1995	ongoing	3	9
WOLV	Wolverhampton Centre	Urban Background	West Midlands	1995	2007	2	9

Continued on next page...

				Ozone data record		**Ozone season (month)**	
Code	**Site Name**	**Site Type**	**Region**	**Start**	**End**	**Start**	**End**
SHE2	Sheffield Centre	Urban Background	Yorkshire Humberside	1995	2013	3	8
ROCH	Rochester Stoke	Rural Background	South East	1996	ongoing	3	9
BREN	London Brent	Urban Background	Greater London	1996	2007	3	9
WA2	London Wandsworth	Urban Background	Greater London	1996	2007	2	9
SUT3	London Sutton	Suburban Background	Greater London	1996	2002	2	8
KC1	London N. Kensington	Urban Background	Greater London	1996	ongoing	3	9
LON6	London Eltham	Suburban Background	Greater London	1996	ongoing	3	9
HG2	London Haringey	Urban Background	Greater London	1996	2012	3	9
EX	Exeter Roadside	Urban Traffic	South West	1996	ongoing	2	9
LEAM	Leamington Spa	Urban Background	West Midlands	1996	ongoing	3	9
GLA3	Glasgow Centre	Urban Background	Central Scotland	1996	2012	2	8
HIL	London Hillingdon	Urban Background	Greater London	1996	ongoing	2	9
TED	London Teddington	Urban Background	Greater London	1996	2016	3	9
THUR	Thurrock	Urban Background	Eastern	1996	ongoing	3	9
NOTT	Nottingham Centre	Urban Background	East Midlands	1996	ongoing	3	9

Continued on next page...

				Ozone data record		**Ozone season (month)**	
Code	**Site Name**	**Site Type**	**Region**	**Start**	**End**	**Start**	**End**
MAN4	Manchester South	Suburban Industrial	North West Merseyside	1996	2016	2	9
HK4	London Hackney	Urban Background	Greater London	1997	2007	3	9
PT	Port Talbot	Urban Industrial	South Wales	1997	2007	3	9
PEMB	Narberth	Rural Background	South Wales	1997	ongoing	1	7
BURY	Bury Roadside	Urban Traffic	North West Merseyside	1997	2007	1	8
BOLT	Bolton	Urban Background	North West Merseyside	1997	2008	3	8
SK1	London Southwark	Urban Background	Greater London	1997	2007	3	9
STOK	Stoke-on-Trent Centre	Urban Background	West Midlands	1997	ongoing	3	9
ECCL	Salford Eccles	Urban Background	North West Merseyside	1997	2013	3	9
LW1	London Lewisham	Urban Background	Greater London	1997	2007	3	9
DERY	Derry	Urban Background	Northern Ireland	1997	2016	1	7
ROTH	Rotherham Centre	Urban Background	Yorkshire Humberside	1997	2007	2	8
REDC	Redcar	Suburban Background	North East	1997	2007	3	9
BAR3	Barnsley Gawber	Urban Background	Yorkshire Humberside	1997	ongoing	2	8
READ	Reading	Urban Background	South East	1997	2003	3	9

Continued on next page...

				Ozone data record		**Ozone season (month)**	
Code	**Site Name**	**Site Type**	**Region**	**Start**	**End**	**Start**	**End**
MY1	London Marylebone Road	Urban Traffic	Greater London	1997	ongoing	1	9
NOR2	Norwich Centre	Urban Background	Eastern	1997	2008	3	9
PLYM	Plymouth Centre	Urban Background	South West	1997	ongoing	2	9
WFEN	Wicken Fen	Rural Background	Eastern	1997	ongoing	3	9
BRAD	Bradford Centre	Urban Background	Yorkshire Humberside	1997	2007	2	8
WBRO	Sandwell West Bromwich	Urban Background	West Midlands	1998	2011	3	9
TRAN	Wirral Tranmere	Urban Background	North West Merseyside	2000	ongoing	3	8
PRES	Preston	Urban Background	North West Merseyside	2000	ongoing	3	8
SEND	Southend-on-Sea	Urban Background	Eastern	2000	ongoing	3	9
BLAC	Blackpool	Urban Background	North West Merseyside	2000	2004	3	8
WEYB	Weybourne	Rural Background	Eastern	2001	ongoing	3	9
HORS	London Westminster	Urban Background	Greater London	2001	2013	3	9
OSY	St Osyth	Rural Background	Eastern	2002	ongoing	3	9
HUL2	Hull Freetown	Urban Background	Yorkshire Humberside	2002	ongoing	3	9
BORN	Bournemouth	Urban Background	South West	2003	ongoing	3	9

Continued on next page...

				Ozone data record		Ozone season (month)	
Code	**Site Name**	**Site Type**	**Region**	**Start**	**End**	**Start**	**End**
NTON	Northampton	Urban Background	East Midlands	2003	2012	3	9
PMTH	Portsmouth	Urban Background	South East	2003	ongoing	3	9
CWMB	Cwmbran	Urban Background	South Wales	2003	2020	2	8
LVP	Liverpool Speke	Urban Industrial	North West Merseyside	2003	ongoing	3	8
ABD	Aberdeen	Urban Background	North East Scotland	2003	ongoing	1	7
REA1	Reading New Town	Urban Background	South East	2003	ongoing	3	9
ED3	Edinburgh St Leonards	Urban Background	Central Scotland	2003	ongoing	2	7
MKTH	Market Harborough	Rural Background	East Midlands	2003	2019	3	8
HRL	London Harlington	Urban Industrial	Greater London	2004	ongoing	3	9
BIR1	Birmingham Tyburn	Urban Background	West Midlands	2004	2016	3	9
WIG5	Wigan Centre	Urban Background	North West Merseyside	2004	ongoing	3	8
BRT3	Brighton Preston Park	Urban Background	South East	2004	ongoing	3	9
SUN2	Sunderland Silksworth	Urban Background	North East	2004	ongoing	2	7
LERW	Lerwick	Rural Background	Highland	2005	ongoing	1	6
BLC2	Blackpool Marton	Urban Background	North West Merseyside	2005	ongoing	3	7

Continued on next page...

				Ozone data record		**Ozone season (month)**	
Code	**Site Name**	**Site Type**	**Region**	**Start**	**End**	**Start**	**End**
LEOM	Leominster	Suburban Background	West Midlands	2005	ongoing	3	9
BRS8	Bristol St Paul's	Urban Background	South West	2006	ongoing	3	9
FW	Fort William	Suburban Background	Highland	2006	ongoing	1	6
ACTH	Auchencorth Moss	Rural Background	Central Scotland	2006	ongoing	1	6
PT4	Port Talbot Margam	Urban Industrial	South Wales	2007	ongoing	2	9
MACK	Charlton Mackrell	Rural Background	South West	2008	ongoing	2	9
BIRT	Birmingham Tyburn Roadside	Urban Traffic	West Midlands	2009	2016	3	8
NO12	Norwich Lakenfields	Urban Background	Eastern	2009	ongoing	2	9
PEEB	Peebles	Urban Background	Scottish Borders	2009	ongoing	2	6
MOLD	Mold	Suburban Background	North Wales	2009	2013	3	9
CANT	Canterbury	Urban Background	South East	2011	ongoing	2	9
AGRN	Birmingham Acocks Green	Urban Background	West Midlands	2011	ongoing	3	9
WAL4	Walsall Woodlands	Urban Background	West Midlands	2012	ongoing	2	9
NTN3	Northampton Kingsthorpe	Urban Background	East Midlands	2012	2017	2	9
HG4	London Haringey Priory Park South	Urban Background	Greater London	2012	ongoing	3	9

Continued on next page...

				Ozone data record		**Ozone season (month)**	
Code	**Site Name**	**Site Type**	**Region**	**Start**	**End**	**Start**	**End**
LECU	Leicester University	Urban Background	East Midlands	2013	ongoing	2	9
GLKP	Glasgow Townhead	Urban Background	Central Scotland	2013	ongoing	2	8
SHDG	Sheffield Devonshire Green	Urban Background	Yorkshire Humberside	2013	ongoing	2	8
COAL	Coventry Allesley	Urban Background	West Midlands	2014	ongoing	2	9
CHBO	Chilbolton Observatory	Rural Background	South East	2016	ongoing	2	9
MAHG	Manchester Sharston	Suburban Industrial	North West Merseyside	2016	ongoing	2	8
DERR	Derry Rosemount	Urban Background	Northern Ireland	2016	ongoing	1	6
BIRR	Birmingham A4540 Roadside	Urban Traffic	West Midlands	2016	ongoing	2	8

Table A.2: Summary of 1- and 10-year return levels of MDA8 ozone (µg/m^3) in different decades for the seven monitoring sites with ozone records dating back to the 1980s. The 95% confidence intervals of our estimates are reported as [lower bound, upper bound].

	1-year return level (µg/m^3)			
Site	**1980s**	**1990s**	**2000s**	**2010s**
AH	154 [137, 173]	184 [170, 195]	142 [134, 151]	123 [118, 129]
BOT	159 [151, 169]	138 [130, 146]	135 [128, 142]	114 [110, 118]
BUSH	133 [120, 151]	118 [112, 125]	110 [106, 114]	111, [107, 116]
ESK	152 [131, 178]	145 [133, 155]	120 [114, 128]	121 [114, 126]
GDF	189 [162, 213]	208 [180, 233]	146 [135, 155]	107 [103, 111]
HAR	166 [155, 178]	172 [163, 190]	148 [139, 157]	120 [115, 125]
SIB	184 [170, 200]	164 [154, 175]	138 [133, 144]	131 [125, 137]
	10-year return level (µg/m^3)			
Site	**1980s**	**1990s**	**2000s**	**2010s**
AH	212 [176, 241]	269 [244, 304]	193 [175, 210]	159 [148, 172]
BOT	213 [193, 229]	186 [172, 202]	173 [161, 185]	132 [125, 138]
BUSH	185 [160, 225]	156 [143, 177]	138 [126, 146]	136 [125, 148]
ESK	227 [188, 264]	215 [190, 237]	158 [146, 176]	155 [141, 171]
GDF	295 [228, 356]	330 [264, 383]	204 [180, 223]	128 [118, 139]
HAR	219 [201, 239]	243 [226, 282]	195 [179, 210]	144 [135, 153]
SIB	275 [246, 321]	225 [204, 242]	172 [162, 186]	165 [153, 176]

Table A.3: Probability of MDA8 exceeding 100 µg/m^3 in different decades for the seven monitoring sites with ozone records dating back to the 1980s. The 95% confidence intervals of our estimates are reported as [lower bound, upper bound].

Site	**1980s**	**1990s**	**2000s**	**2010s**
AH	0.04 [0.03, 0.06]	0.06 [0.05, 0.08]	0.05 [0.04, 0.06]	0.02 [0.02, 0.03]
BOT	0.06 [0.05, 0.07]	0.03 [0.02, 0.04]	0.04 [0.03, 0.05]	0.02 [0.02, 0.03]
BUSH	0.03 [0.02, 0.04]	0.02 [0.02, 0.03]	0.01 [0.01, 0.02]	0.02 [0.01, 0.02]
ESK	0.04 [0.02, 0.06]	0.03 [0.03, 0.05]	0.02 [0.02, 0.03]	0.02 [0.02, 0.03]
GDF	0.06 [0.04, 0.08]	0.07 [0.05, 0.08]	0.03 [0.03, 0.04]	0.01 [0.01, 0.01]
HAR	0.06 [0.05, 0.08]	0.06 [0.05, 0.07]	0.05 [0.04, 0.06]	0.03 [0.02, 0.04]
SIB	0.06 [0.05, 0.07]	0.06 [0.05, 0.06]	0.05 [0.04, 0.06]	0.04 [0.04, 0.05]

A.2 Supplementary figures

Figure A.1: The regions of the UK as defined by the Department for Environment, Food and Rural Affairs.

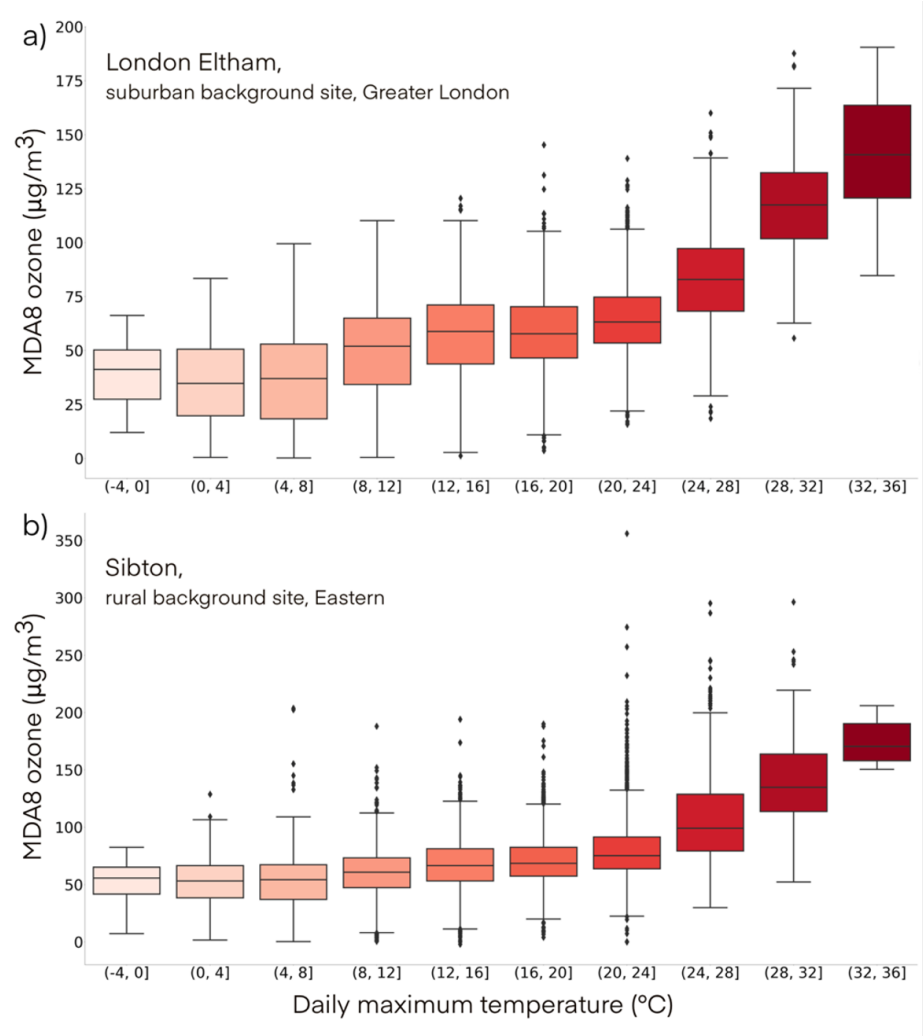

Figure A.2: 4 °C temperature binned boxplots, showing the non-linear relationship between daily maximum temperature and MDA8 ozone for two example sites: (a) London Eltham, and (b) Sibton.

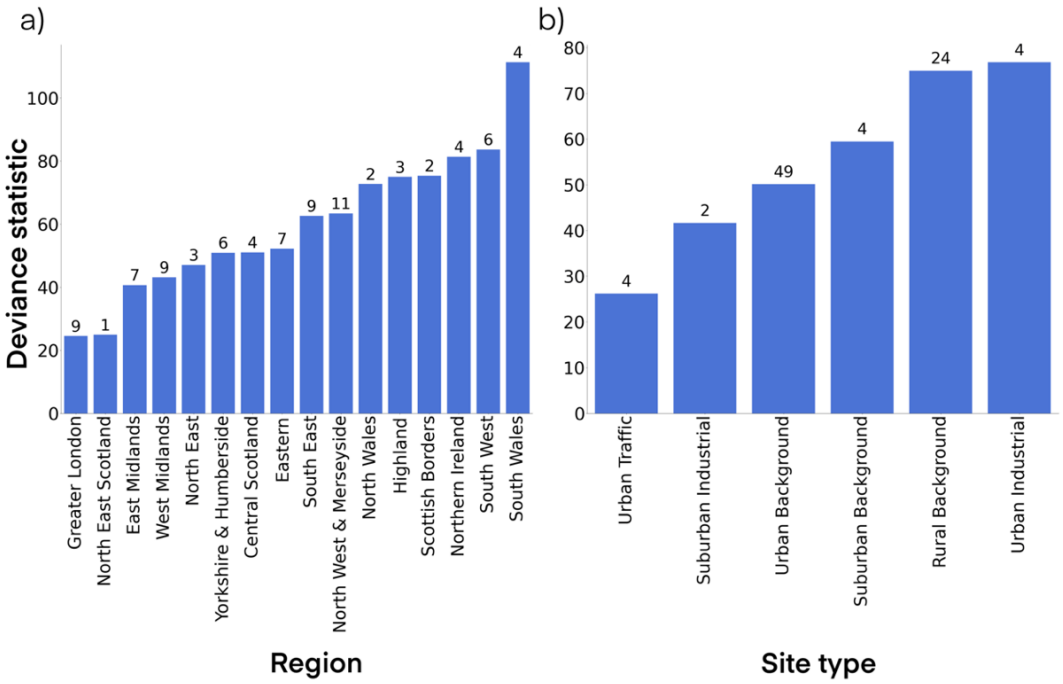

Figure A.3: The mean deviance test statistic grouped by region (a) and site type (b). The number above each bar is the number of sites for the group.

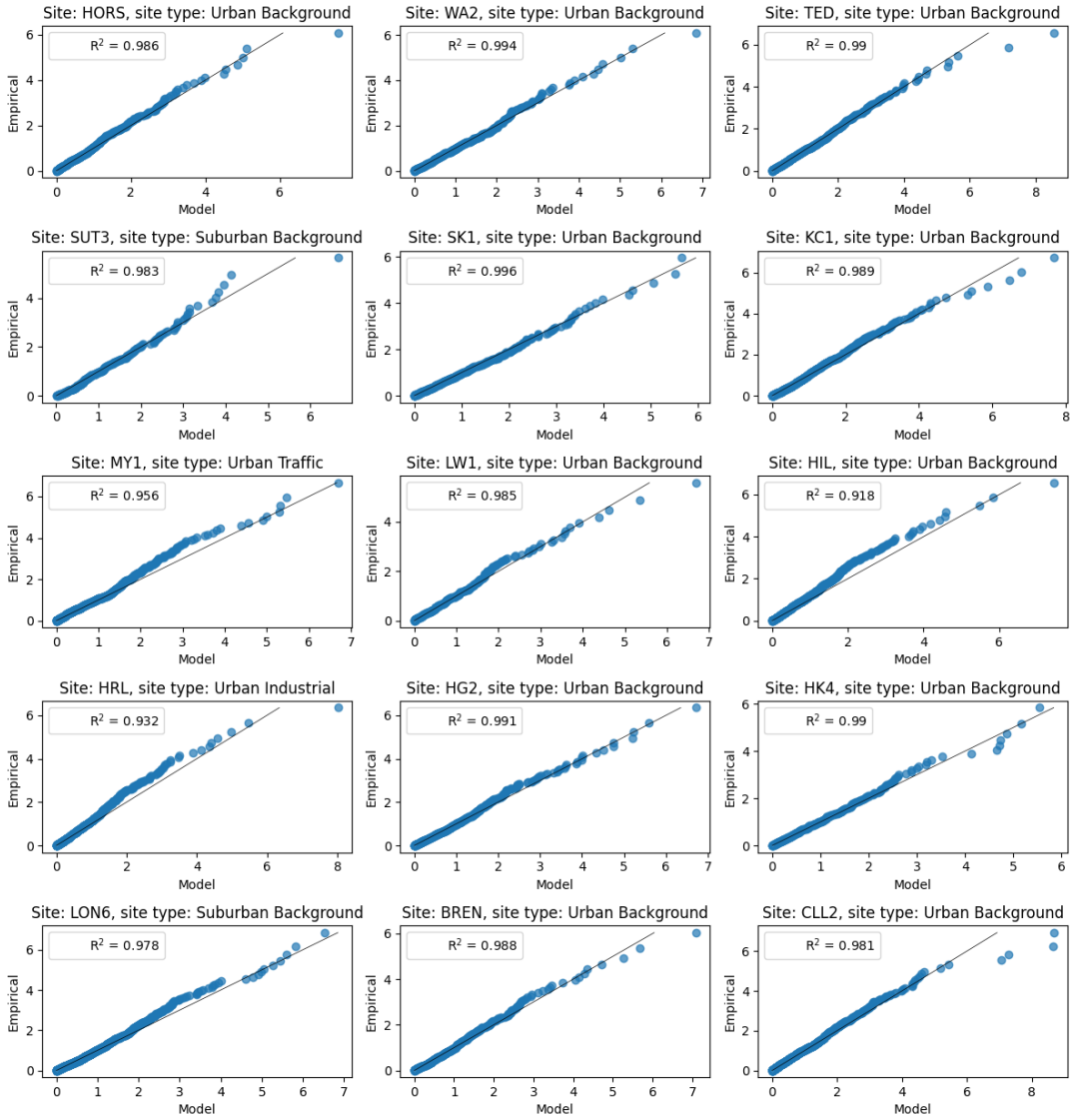

Figure A.4: Example quantile-quantile plots showing the 2000–2009 model fit for sites in the Greater London region. Each subplot is a residual (MDA8 ozone exceeding the modelling threshold) quantile plot on the exponential scale. The x-axes are the model output quantiles. The y-axes are the empirical quantiles from the MDA8 ozone observations. The R^2 values for each site's model fit are also shown. For a good fit, points should lie on the line $x = y$.

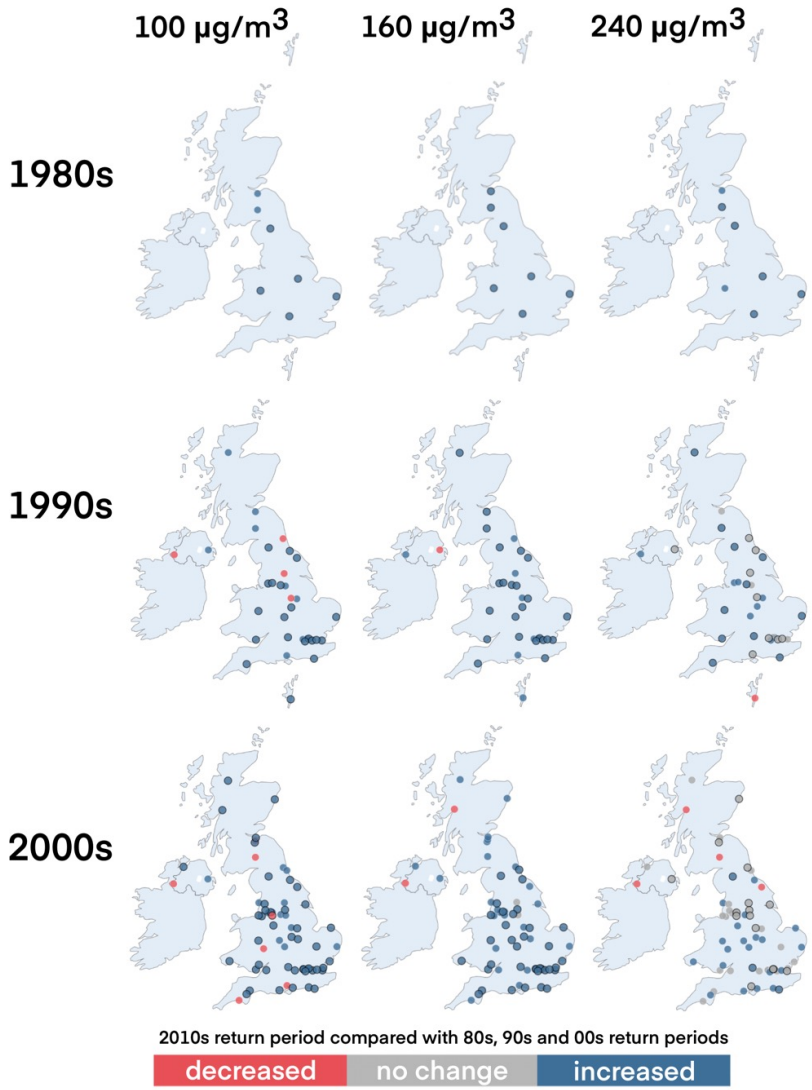

Figure A.5: The change in the 2010s return period for health threshold levels of ozone when compared with the 1980s, 1990s and 2000s return periods for each UK measurement site, found by using the simulation method in Section 2.3.4. Left column: moderate health threshold of 100 µg/m^3; middle column: high health threshold of 160 µg/m^3; right column: very high threshold of 240 µg/m^3. A decrease (or increase) in return period means that a site exceeds the health threshold level more often (or less often) in the 2010s than in the compared decade. Sites are outlined if the change is significant (p=0.05).

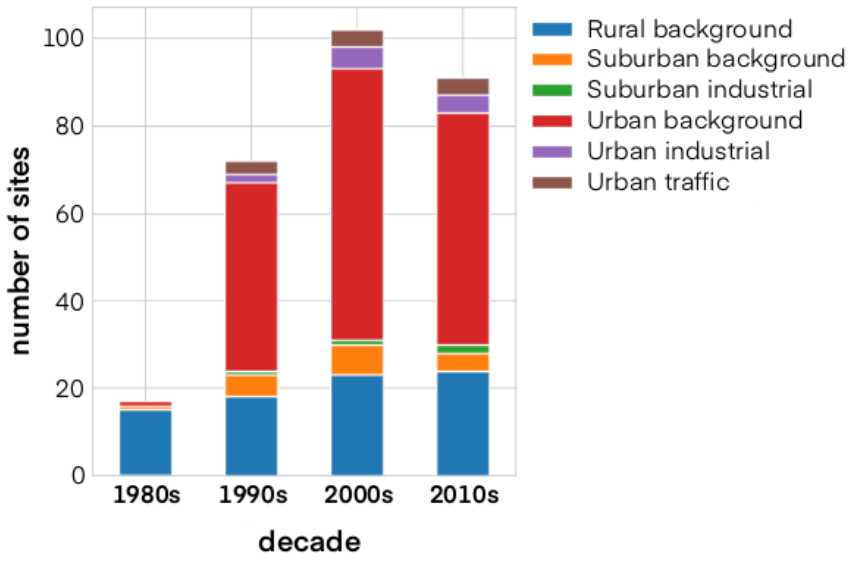

Figure A.6: The number of AURN sites for each decade and site type included within this study.

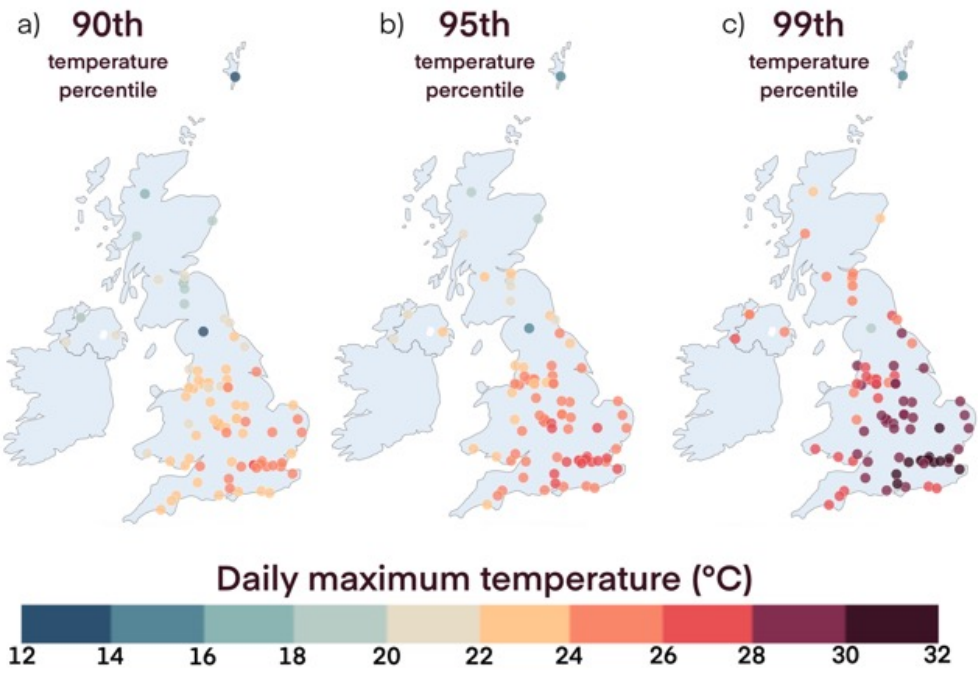

Figure A.7: The daily maximum temperature values (°C) for each site for the 90th (a), 95th (b) and 99th (c) temperature percentile statistics, calculated using the ozone season temperature values.

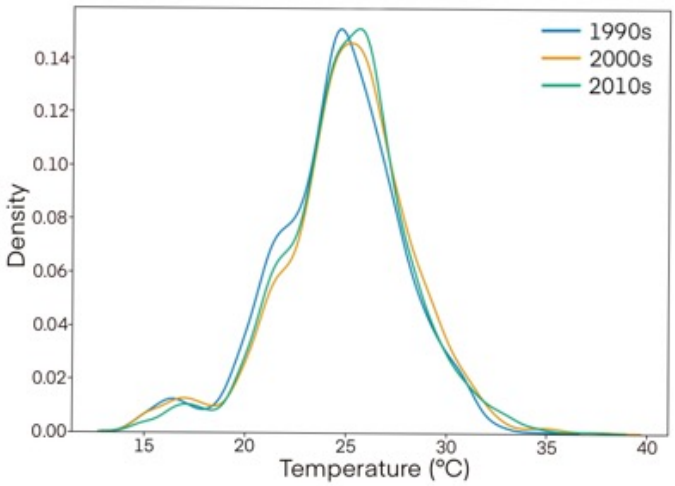

Figure A.8: Density plots of daily maximum temperatures (°C), grouped by decade, using daily maximum temperatures above the site-specific 90^{th} temperature percentile. The sites included are those monitoring from the 1990s through to the 2010s.

A.3 Data simulation method

Simulation of 1000 years of threshold-exceeding ozone measurements is accomplished by the following. First, simulating 1000 years of daily temperatures by sampling from the empirical distribution of the temperature data. In practice this means sampling with replacement from the observed temperature values. Second, using the trained logistic regression model to generate a binary value for each simulated temperature value, thereby assigning whether a threshold exceedance has occurred. Third, if an exceedance has occurred, the magnitude of the exceedance is simulated from the generalised Pareto distribution using the probability integral transformation. In practice this involves simulating a data point U at random on the interval [0, 1] and then transforming this to the appropriate scale for the trained generalised Pareto distribution $\frac{\sigma_{\mathrm{T}}}{\xi} * \left[(1-\mathrm{U})^{(-\xi)} - 1\right] + u_{\mathrm{T}}$ where σ_{T} is the temperature-dependent scale parameter, ξ is the shape parameter, and u_{T} is the temperature-dependent threshold.

Appendix B

Appendix to Chapter 3

B.1 Additional tables

Table B.1: Annual mean MDA8 (µg/m^3) per region for 2001–2018, with 95% confidence intervals of the mean estimate shown in square brackets.

Region	**Downscaled**	**EMEP4UK**	**Measurements**
East Midlands (England)	62.38 [62.37, 62.38]	75.02 [74.99, 75.05]	62.15 [61.76, 62.54]
East of England	64.96 [64.95, 64.96]	75.89 [75.86, 75.93]	64.55 [64.23, 64.87]
London	57.37 [57.34, 57.39]	70.41 [70.31, 70.54]	48.56 [48.37, 48.75]
North East (England)	60.68 [60.68, 60.69]	77.31 [77.27, 77.35]	58.79 [58.30, 59.24]
North West (England)	59.72 [59.72, 59.73]	76.39 [76.35, 76.42]	58.76 [58.46, 59.10]
Northern Ireland	59.16 [59.15, 59.16]	77.44 [77.42, 77.46]	59.40 [58.97, 59.85]
Scotland	62.77 [62.77, 62.77]	78.30 [78.29, 78.31]	66.14 [65.92, 66.37]
South East (England)	66.24 [66.24, 66.25]	76.82 [76.79, 76.84]	66.49 [66.28, 66.73]
South West (England)	67.11 [67.11, 67.12]	78.71 [78.70, 78.74]	65.43 [65.05, 65.81]
Wales	66.61 [66.61, 66.62]	79.29 [79.26, 79.31]	64.66 [64.36, 64.92]
West Midlands (England)	62.67 [62.66, 62.67]	75.00 [74.95, 75.03]	60.90 [60.58, 61.22]
Yorkshire and The Humber	59.27 [59.26, 59.27]	74.90 [74.88, 74.93]	58.97 [58.57, 59.35]
All-region mean	**62.41 [62.40, 62.42]**	**76.29 [76.26, 76.33]**	**61.23 [60.91, 61.56]**
All-region sdev. (% of mean)	**3.13 (5.01)**	**2.37 (3.10)**	**4.99 (8.15)**

Table B.2: March-August MDA8 (µg/m^3) per region for 2001–2018, with 95% confidence intervals of the mean estimate shown in square brackets.

Region	**Downscaled**	**EMEP4UK**	**Measurements**
East Midlands (England)	71.12 [71.11, 71.13]	86.42 [86.38, 86.44]	71.35 [70.85, 71.82]
East of England	75.14 [75.13, 75.14]	88.07 [88.04, 88.10]	74.17 [73.78, 74.59]
London	68.90 [68.87, 68.93]	84.37 [84.27, 84.49]	58.04 [57.78, 58.34]
North East (England)	66.09 [66.08, 66.10]	85.18 [85.13, 85.22]	66.16 [65.59, 66.74]
North West (England)	65.57 [65.56, 65.58]	85.20 [85.17, 85.23]	66.48 [66.12, 66.90]
Northern Ireland	62.76 [62.75, 62.77]	82.61 [82.57, 82.65]	63.35 [62.78, 63.94]
Scotland	66.30 [66.29, 66.30]	83.66 [83.65, 83.68]	70.64 [70.34, 70.95]
South East (England)	75.56 [75.55, 75.57]	87.86 [87.83, 87.89]	75.07 [74.79, 75.34]
South West (England)	73.05 [73.05, 73.06]	86.12 [86.10, 86.15]	71.91 [71.38, 72.46]
Wales	71.51 [71.50, 71.51]	85.84 [85.81, 85.87]	69.21 [68.81, 69.58]
West Midlands (England)	70.40 [70.39, 70.40]	85.06 [85.03, 85.09]	69.32 [68.89, 69.71]
Yorkshire and The Humber	66.69 [66.69, 66.70]	85.19 [85.16, 85.22]	67.02 [66.45, 67.52]
All-region mean	**69.42 [69.41, 69.43]**	**85.47 [85.43, 85.50]**	**68.56 [68.13, 68.99]**
All-region sdev. (% of mean)	**4.04 (5.82)**	**1.57 (1.83)**	**4.75 (6.93)**

Table B.3: Regional average 90th percentiles of MDA8 ozone (µg/m^3) for 2001–2018, with 95% confidence intervals of the mean estimate shown in square brackets.

Region	**Downscaled**	**EMEP4UK**	**Measurements**
East Midlands (England)	82.58 [82.59, 82.56]	97.22 [97.17, 97.26]	84.87 [84.27, 85.57]
East of England	86.52 [86.53, 86.51]	98.11 [98.08, 98.14]	87.32 [86.78, 87.91]
London	82.52 [82.57, 82.48]	96.71 [96.44, 96.94]	77.19 [76.88, 77.47]
North East (England)	77.84 [77.85, 77.83]	96.18 [96.13, 96.24]	80.22 [79.38, 80.74]
North West (England)	78.32 [78.34, 78.32]	96.90 [96.86, 96.95]	81.34 [81.01, 81.76]
Northern Ireland	75.61 [75.62, 75.60]	95.59 [95.55, 95.63]	79.00 [78.28, 79.56]
Scotland	79.03 [79.04, 79.03]	95.41 [95.40, 95.42]	86.20 [85.90, 86.50]
South East (England)	87.86 [87.87, 87.84]	99.48 [99.43, 99.53]	89.48 [89.13, 89.77]
South West (England)	85.53 [85.54, 85.52]	98.99 [98.95, 99.03]	89.63 [88.94, 90.07]
Wales	83.88 [83.89, 83.87]	98.72 [98.69, 98.75]	86.41 [86.00, 86.84]
West Midlands (England)	81.90 [81.91, 81.89]	97.22 [97.18, 97.28]	84.07 [83.59, 84.78]
Yorkshire and The Humber	78.38 [78.39, 78.37]	96.23 [96.19, 96.28]	82.52 [81.60, 83.26]
All-region mean	**81.66 [81.65, 81.68]**	**97.23 [97.17, 97.29]**	**84.02 [83.48, 84.52]**
All-region sdev (% of mean)	**3.86 (4.72)**	**1.34 (1.37)**	**4.04 (4.80)**

Table B.4: Regional average 10th percentiles of MDA8 ozone (µg/m^3) for 2001–2018, with 95% confidence intervals of the mean estimate shown in square brackets.

Region	**Downscaled**	**EMEP4UK**	**Measurements**
East Midlands (England)	41.52 [41.50, 41.53]	52.02 [51.94, 52.09]	37.09 [36.38, 37.79]
East of England	42.05 [42.04, 42.07]	51.79 [51.71, 51.87]	39.25 [38.69, 40.04]
London	29.18 [29.12, 29.23]	40.57 [40.25, 40.79]	16.96 [16.65, 17.27]
North East (England)	45.32 [45.31, 45.33]	58.97 [58.88, 59.04]	36.11 [34.80, 36.80]
North West (England)	41.71 [41.70, 41.73]	56.10 [56.01, 56.17]	31.52 [30.72, 32.29]
Northern Ireland	44.42 [44.41, 44.43]	60.97 [60.92, 61.00]	38.88 [38.27, 39.61]
Scotland	48.31 [48.31, 48.31]	61.49 [61.47, 61.51]	45.34 [44.99, 45.66]
South East (England)	44.13 [44.12, 44.15]	54.18 [54.10, 54.24]	41.09 [40.63, 41.55]
South West (England)	49.22 [49.21, 49.22]	60.95 [60.91, 60.99]	38.72 [37.78, 39.48]
Wales	50.28 [50.27, 50.29]	62.08 [62.04, 62.11]	39.90 [39.17, 40.42]
West Midlands (England)	42.27 [42.26, 42.29]	53.48 [53.40, 53.55]	35.86 [35.37, 36.63]
Yorkshire and The Humber	40.31 [40.30, 40.33]	52.82 [52.74, 52.89]	34.63 [33.80, 35.45]
All-region mean	**43.23 [43.21, 43.24]**	**55.45 [55.36, 55.52]**	**36.28 [35.60, 36.92]**
All-region sdev (% of mean)	**5.49 (12.70)**	**6.13 (11.05)**	**7.00 (19.30)**

Table B.5: Annual mean trends of MDA8 ozone (µg/m^3/yr) per region for 2001–2018, with 95% confidence intervals of the mean estimate shown in brackets. Significant trends are in bold.

Region	**Downscaled**	**EMEP4UK**	**Measurements**
East Midlands (England)	-0.15 [-0.42, 0.13]	0.05 [-0.16, 0.26]	0.24 [-0.08, 0.55]
East of England	-0.16 [-0.46, 0.15]	-0.07 [-0.27, 0.12]	0.09 [-0.19, 0.36]
London	-0.06 [-0.38, 0.26]	**0.43 [0.20, 0.66]**	-0.23 [-0.50, 0.05]
North East (England)	-0.19 [-0.41, 0.03]	-0.10 [-0.30, 0.09]	-0.18 [-0.42, 0.06]
North West (England)	-0.09 [-0.34, 0.15]	0.03 [-0.18, 0.24]	0.29 [-0.02, 0.59]
Northern Ireland	-0.10 [-0.32, 0.12]	-0.16 [-0.36, 0.04]	-0.06 [-0.27, 0.15]
Scotland	-0.17 [-0.40, 0.05]	-0.16 [-0.38, 0.06]	0.18 [-0.04, 0.39]
South East (England)	-0.26 [-0.56, 0.04]	-0.04 [-0.26, 0.17]	0.06 [-0.24, 0.37]
South West (England)	-0.25 [-0.55, 0.04]	-0.11 [-0.32, 0.10]	0.19 [-0.10, 0.47]
Wales	-0.25 [-0.50, 0.00]	-0.14 [-0.34, 0.05]	-0.13 [-0.36, 0.11]
West Midlands (England)	-0.17 [-0.45, 0.11]	0.03 [-0.18, 0.23]	0.28 [-0.01, 0.56]
Yorkshire and The Humber	-0.13 [-0.35, 0.10]	0.01 [-0.17, 0.19]	**0.33 [0.02, 0.64]**

Table B.6: March–August mean trends of MDA8 ozone (µg/m^3/yr) per region for 2001–2018, with 95% confidence intervals of the mean estimate shown in brackets. Significant trends are in bold.

Region	**Downscaled**	**EMEP4UK**	**Measurements**
East Midlands (England)	**-0.47 [-0.86, -0.07]**	**-0.32 [-0.56, -0.08]**	-0.01 [-0.42, 0.40]
East of England	**-0.46 [-0.88, -0.04]**	**-0.43 [-0.65, -0.21]**	-0.21 [-0.60, 0.18]
London	-0.34 [-0.77, 0.10]	0.17 [-0.06, 0.39]	**-0.51 [-0.89, -0.13]**
North East (England)	**-0.40 [-0.71, -0.08]**	**-0.40 [-0.62, -0.18]**	**-0.44 [-0.76, -0.12]**
North West (England)	-0.32 [-0.65, 0.02]	**-0.28 [-0.50, -0.06]**	0.11 [-0.26, 0.47]
Northern Ireland	-0.22 [-0.53, 0.09]	**-0.36 [-0.59, -0.13]**	-0.24 [-0.51, 0.02]
Scotland	-0.32 [-0.67, 0.02]	**-0.37 [-0.62, -0.11]**	0.00 [-0.33, 0.33]
South East (England)	**-0.58 [-1.02, -0.15]**	**-0.39 [-0.62, -0.16]**	-0.21 [-0.63, 0.2]
South West (England)	**-0.52 [-0.95, -0.09]**	**-0.41 [-0.64, -0.18]**	-0.02 [-0.46, 0.41]
Wales	**-0.46 [-0.82, -0.10]**	**-0.44 [-0.65, -0.23]**	**-0.40 [-0.77, -0.02]**
West Midlands (England)	**-0.46 [-0.85, -0.06]**	**-0.33 [-0.57, -0.10]**	0.02 [-0.38, 0.41]
Yorkshire and The Humber	**-0.35 [-0.68, -0.02]**	**-0.29 [-0.50, -0.08]**	0.15 [-0.29, 0.59]

Table B.7: 90th percentile trends of MDA8 ozone (µg/m^3/yr) per region for 2001–2018, with 95% confidence intervals of the mean estimate shown in brackets. Significant trends are in bold.

Region	**Downscaled**	**EMEP4UK**	**Measurements**
East Midlands (England)	**-0.57 [-1.12, -0.02]**	**-0.55 [-0.90, -0.20]**	-0.17 [-0.63, 0.28]
East of England	**-0.61 [-1.18, -0.04]**	**-0.75 [-1.09, -0.40]**	-0.50 [-1.02, 0.02]
London	-0.48 [-1.14, 0.19]	-0.16 [-0.48, 0.16]	**-0.51 [-0.95, -0.07]**
North East (England)	**-0.39 [-0.77, -0.01]**	**-0.52 [-0.79, -0.25]**	**-0.59 [-0.95, -0.22]**
North West (England)	-0.29 [-0.69, 0.11]	**-0.42 [-0.70, -0.14]**	-0.19 [-0.57, 0.20]
Northern Ireland	-0.16 [-0.49, 0.16]	**-0.26 [-0.50, -0.02]**	**-0.33 [-0.62, -0.03]**
Scotland	-0.27 [-0.62, 0.09]	**-0.43 [-0.74, -0.13]**	-0.11 [-0.45, 0.22]
South East (England)	**-0.74 [-1.35, -0.12]**	**-0.63 [-0.97, -0.28]**	-0.36 [-0.87, 0.14]
South West (England)	**-0.55 [-1.06, -0.05]**	**-0.48 [-0.79, -0.18]**	-0.03 [-0.47, 0.40]
Wales	**-0.48 [-0.90, -0.06]**	**-0.47 [-0.74, -0.19]**	-0.37 [-0.80, 0.05]
West Midlands (England)	-0.50 [-1.02, 0.02]	**-0.41 [-0.74, -0.09]**	0.01 [-0.42, 0.45]
Yorkshire and The Humber	-0.39 [-0.82, 0.05]	**-0.42 [-0.70, -0.15]**	-0.04 [-0.53, 0.46]

Table B.8: 10th percentile trends of MDA8 ozone (µg/m^3/yr) per region for 2001–2018, with 95% confidence intervals of the mean estimate shown in brackets. Significant trends are in bold.

Region	**Downscaled**	**EMEP4UK**	**Measurements**
East Midlands (England)	**0.42 [0.14, 0.69]**	**0.75 [0.40, 1.09]**	**0.80 [0.45, 1.15]**
East of England	**0.40 [0.10, 0.69]**	**0.60 [0.27, 0.92]**	**0.71 [0.42, 1.01]**
London	**0.53 [0.22, 0.84]**	**1.19 [0.75, 1.62]**	0.17 [-0.04, 0.37]
North East (England)	0.07 [-0.16, 0.31]	**0.37 [0.05, 0.68]**	0.28 [-0.03, 0.59]
North West (England)	0.24 [-0.07, 0.54]	**0.61 [0.27, 0.96]**	**0.81 [0.42, 1.20]**
Northern Ireland	-0.01 [-0.30, 0.27]	-0.03 [-0.30, 0.24]	0.20 [-0.12, 0.52]
Scotland	-0.06 [-0.29, 0.17]	0.04 [-0.21, 0.29]	**0.57 [0.27, 0.86]**
South East (England)	**0.34 [0.03, 0.64]**	**0.68 [0.30, 1.07]**	**0.70 [0.44, 0.96]**
South West (England)	0.10 [-0.18, 0.38]	**0.40 [0.05, 0.75]**	**0.53 [0.24, 0.82]**
Wales	-0.01 [-0.26, 0.24]	0.23 [-0.08, 0.53]	**0.21 [0.01, 0.41]**
West Midlands (England)	**0.36 [0.03, 0.69]**	**0.78 [0.40, 1.16]**	**0.83 [0.48, 1.18]**
Yorkshire and The Humber	**0.26 [0.03, 0.49]**	**0.57 [0.27, 0.88]**	**0.86 [0.53, 1.20]**

B.2 Additional figures

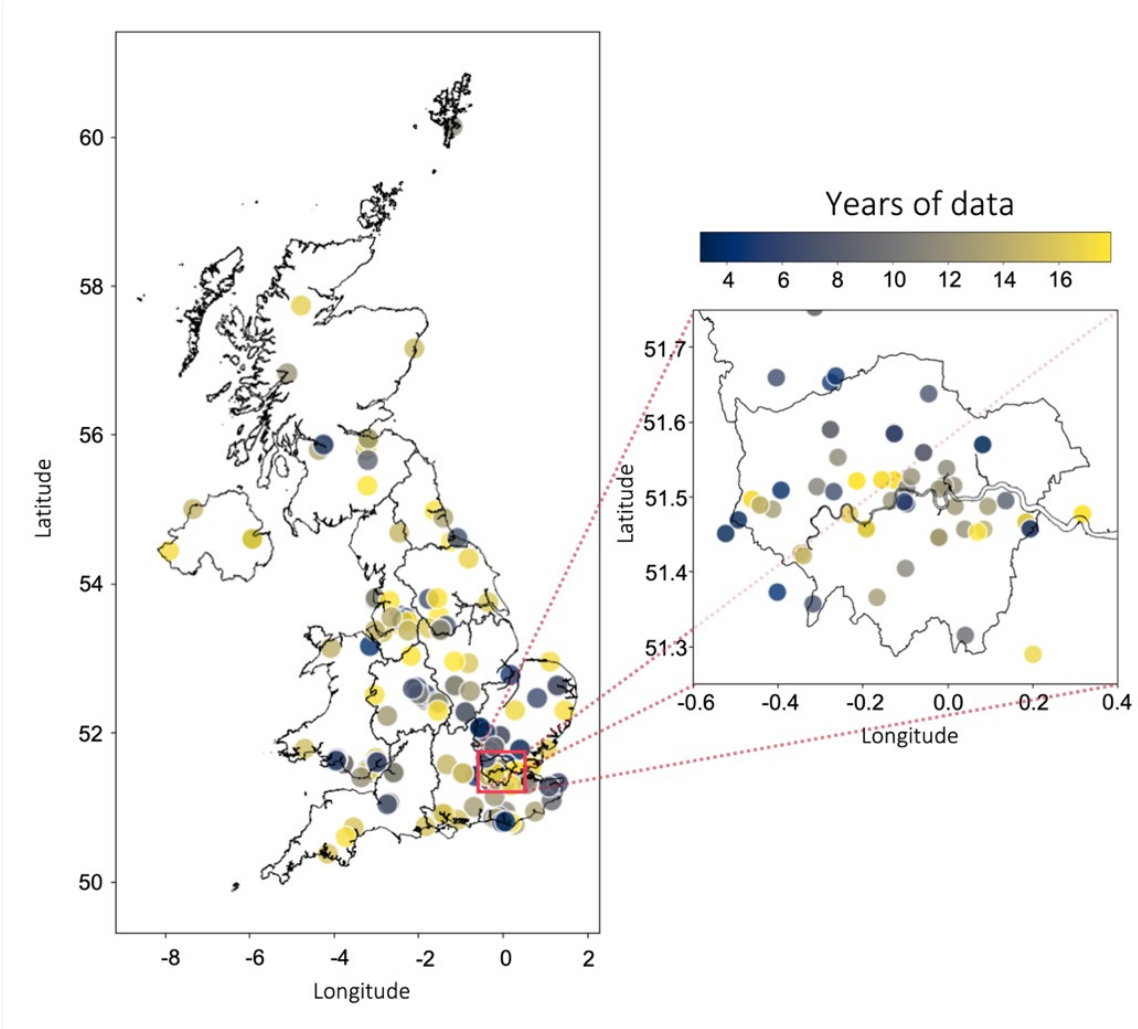

Figure B.1: Measurement station map with number of measurement days in record for each station.

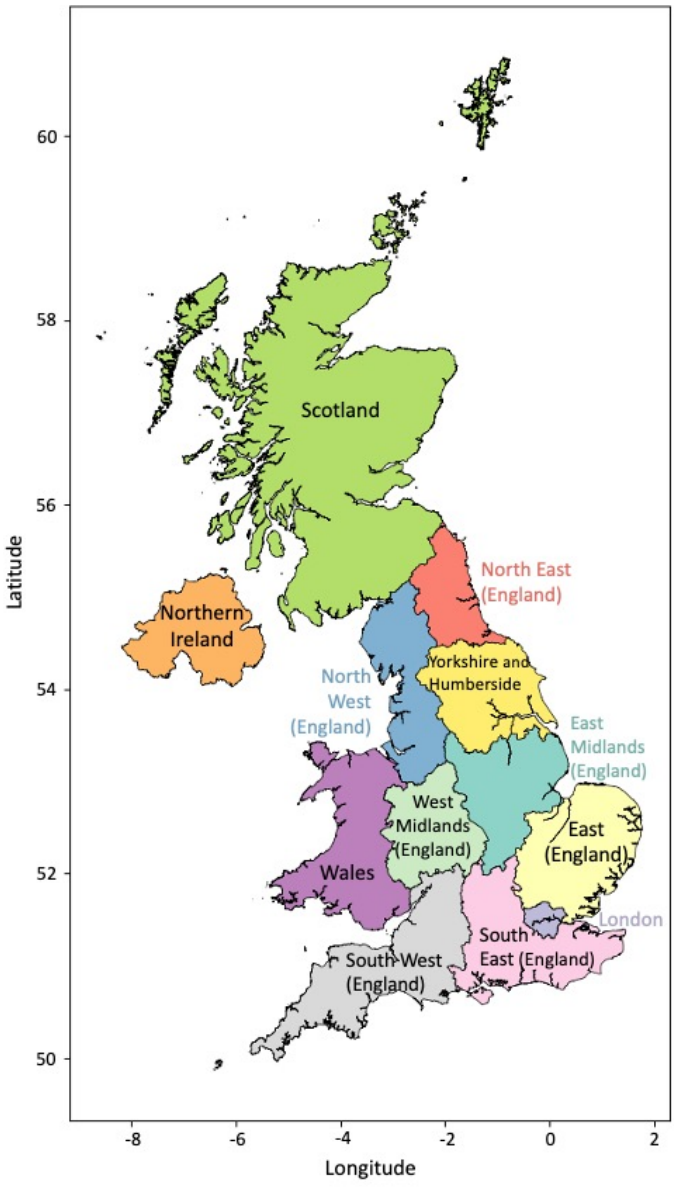

Figure B.2: The region definitions for this paper (Level 1 Nomenclature of Territorial Units for Statistics).

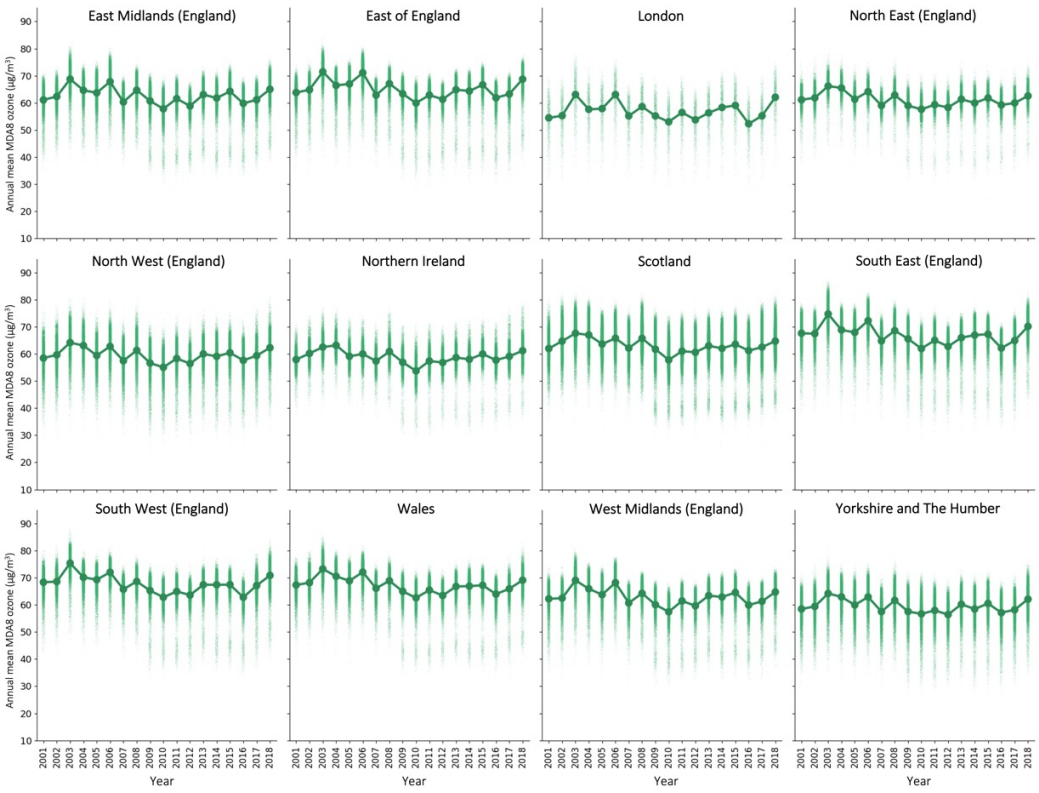

Figure B.3: Regional annual mean MDA8 ozone for the downscaled surface, for 2001–2018. Background dots are individual cell estimates, larger foreground dots are the yearly average.

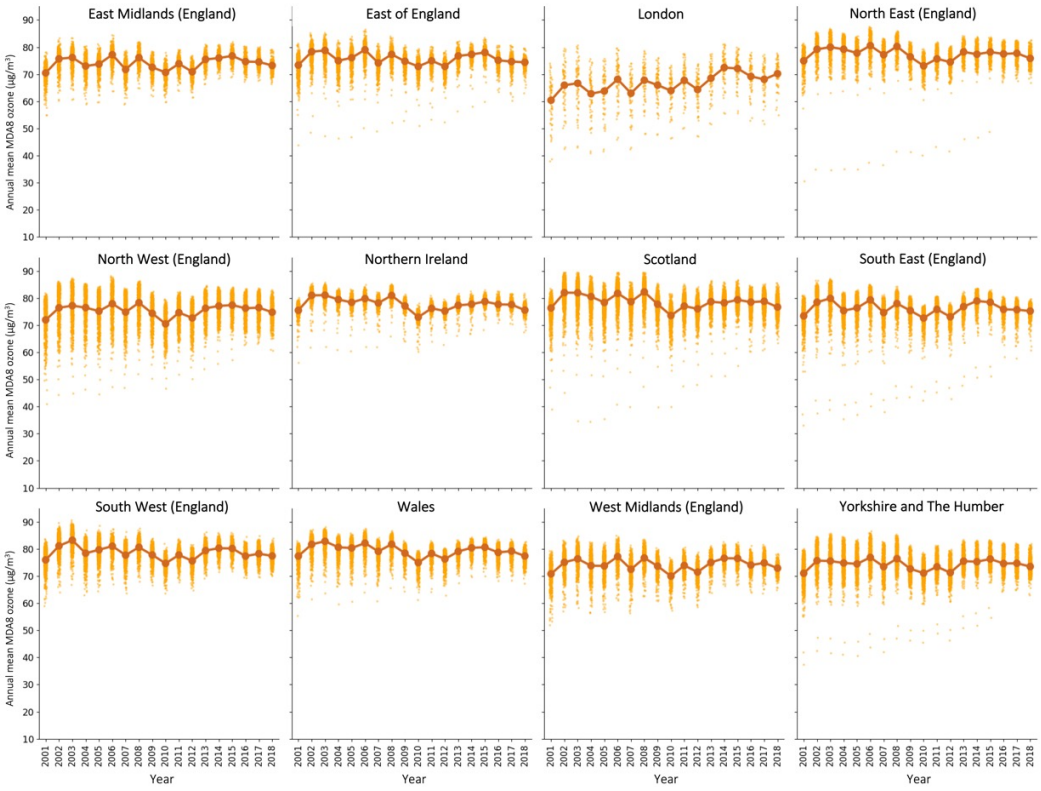

Figure B.4: Regional annual mean MDA8 ozone for the EMEP4UK surface, for 2001–2018. Background dots are individual cell estimates, larger foreground dots are the yearly average.

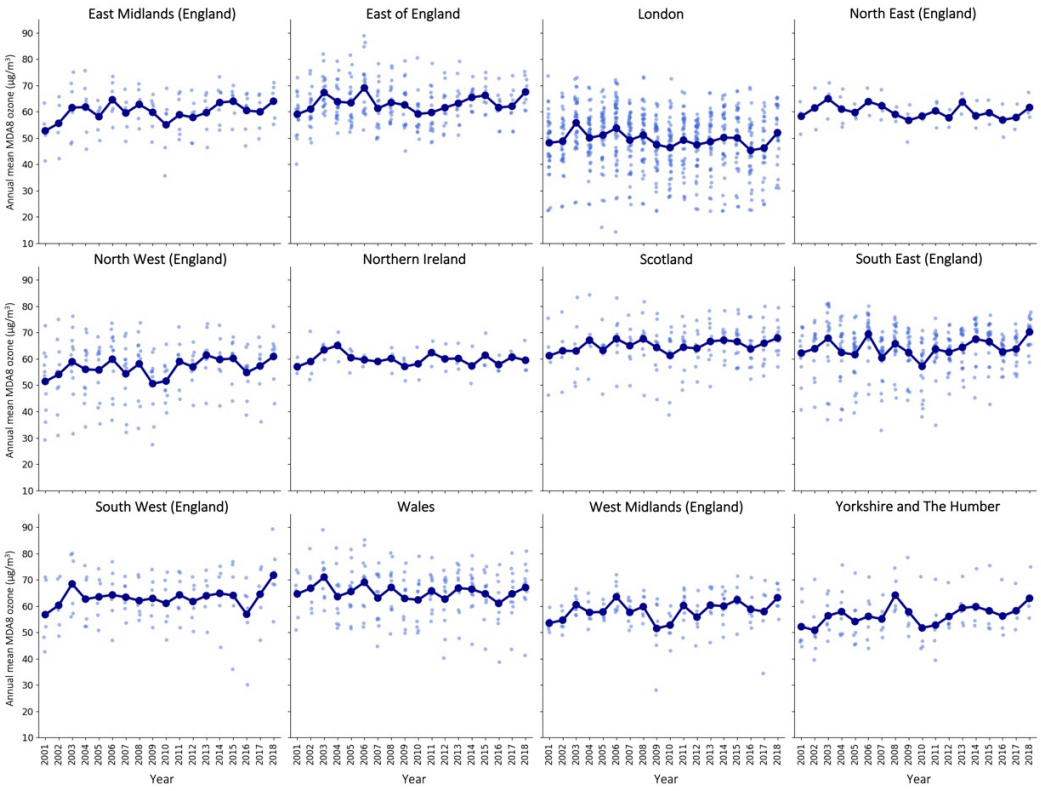

Figure B.5: Regional annual mean MDA8 ozone for the measurement data, for 2001–2018. Background dots are individual cell estimates, larger foreground dots are the yearly average.

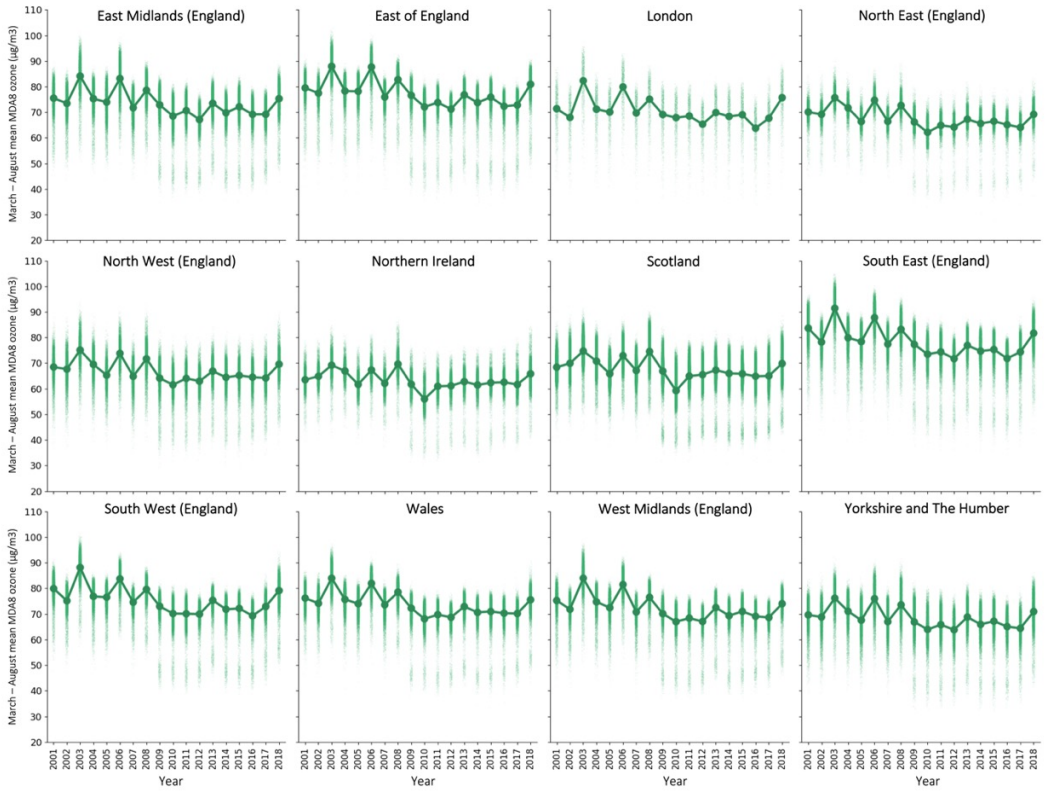

Figure B.6: Regional March–August mean MDA8 ozone for the downscaled surface, for 2001–2018. Background dots are individual cell estimates, larger foreground dots are the yearly average.

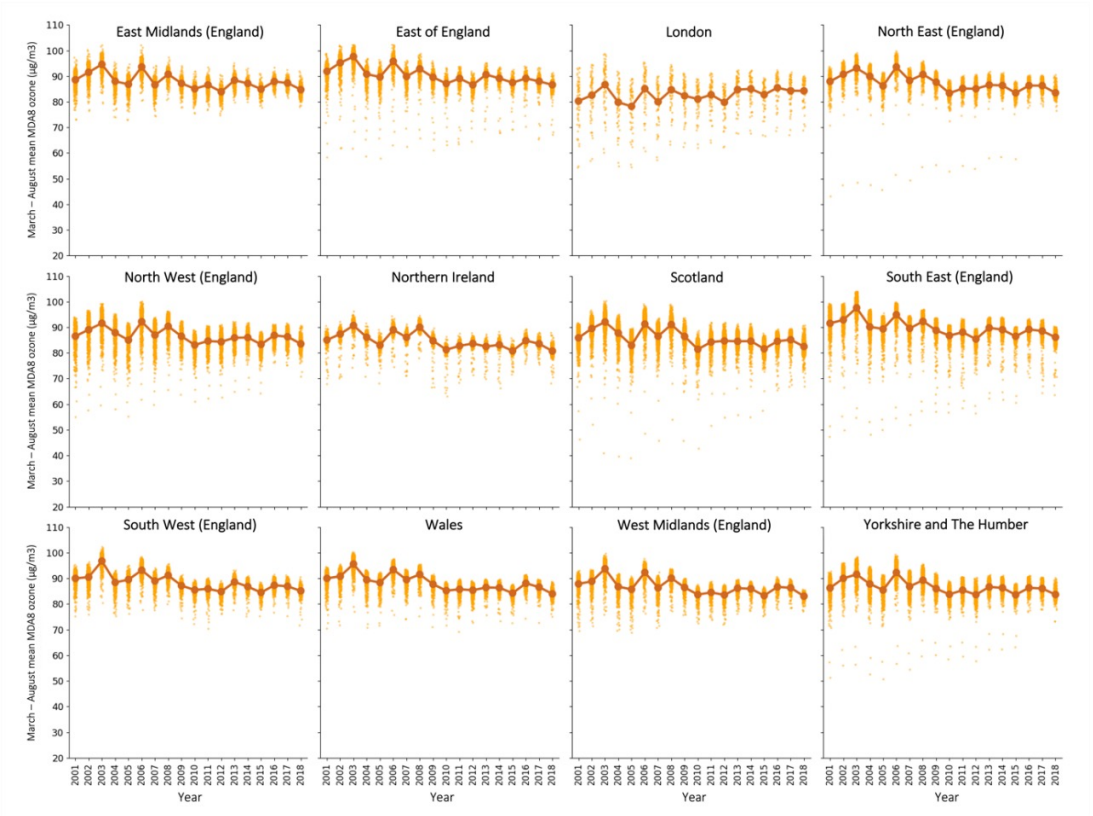

Figure B.7: Regional March–August mean MDA8 ozone for the EMEP4UK surface, for 2001–2018. Background dots are individual cell estimates, larger foreground dots are the yearly average.

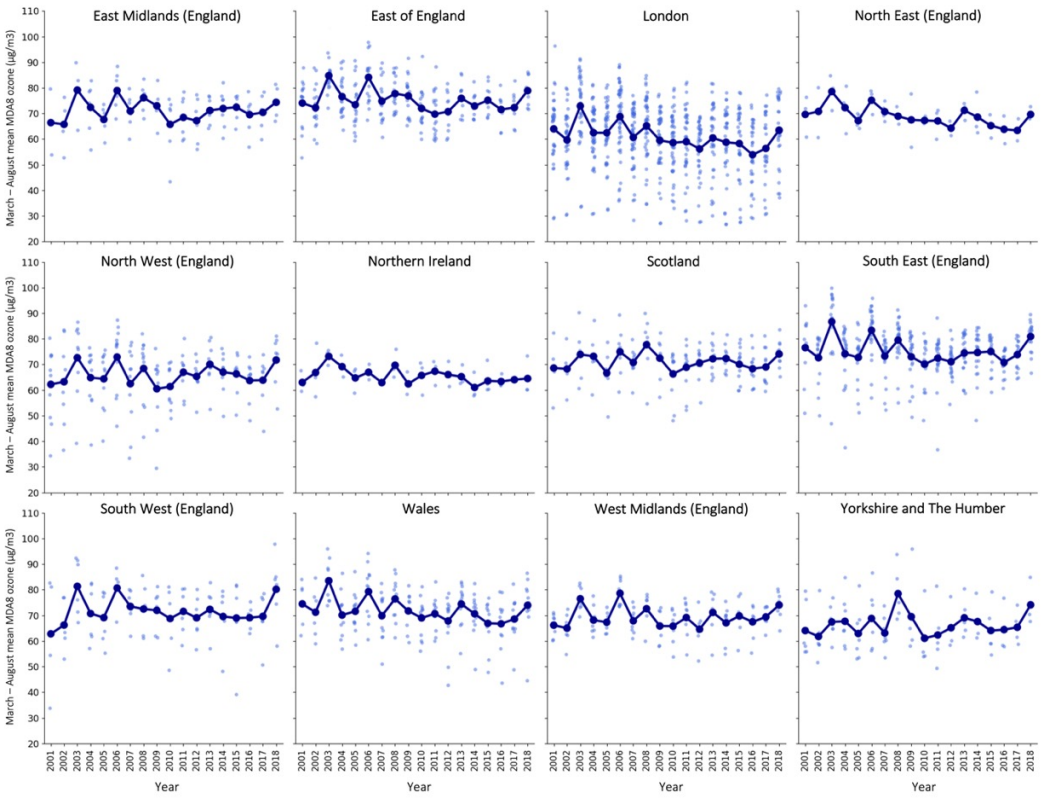

Figure B.8: Regional March–August mean MDA8 ozone for the measurement data, for 2001–2018. Background dots are individual cell estimates, larger foreground dots are the yearly average.

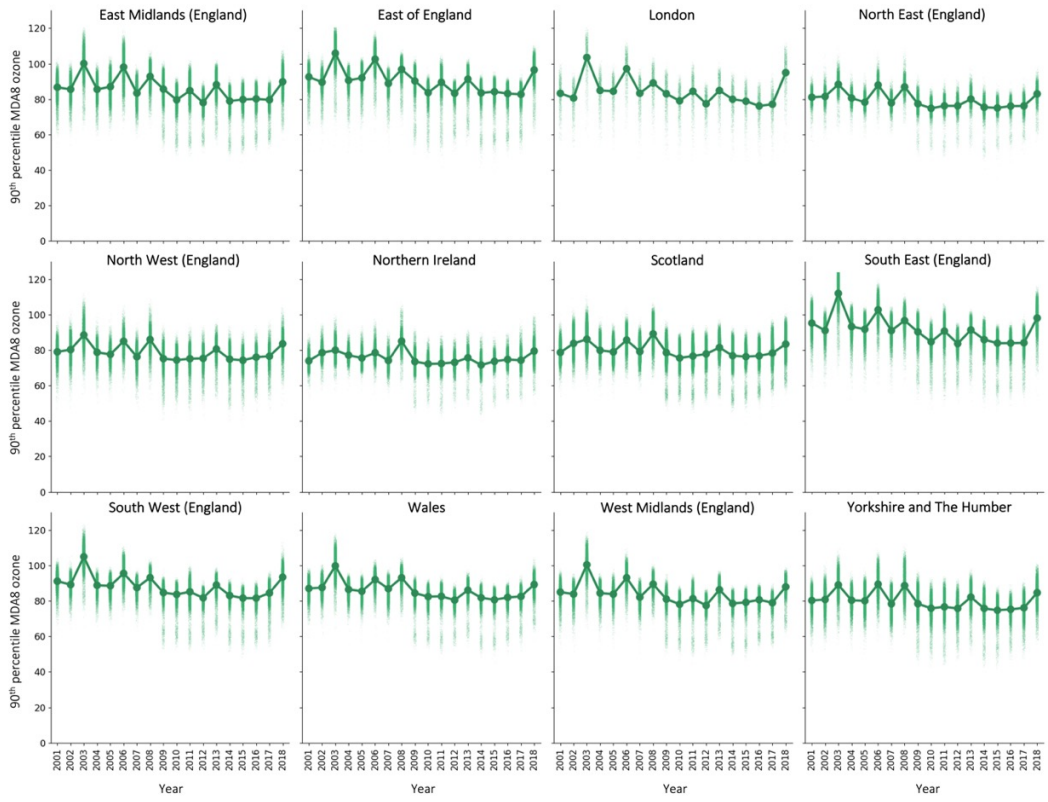

Figure B.9: Regional 90$^{\text{th}}$ percentile MDA8 ozone for the downscaled surface, for 2001–2018. Background dots are individual cell estimates, larger foreground dots are the yearly average.

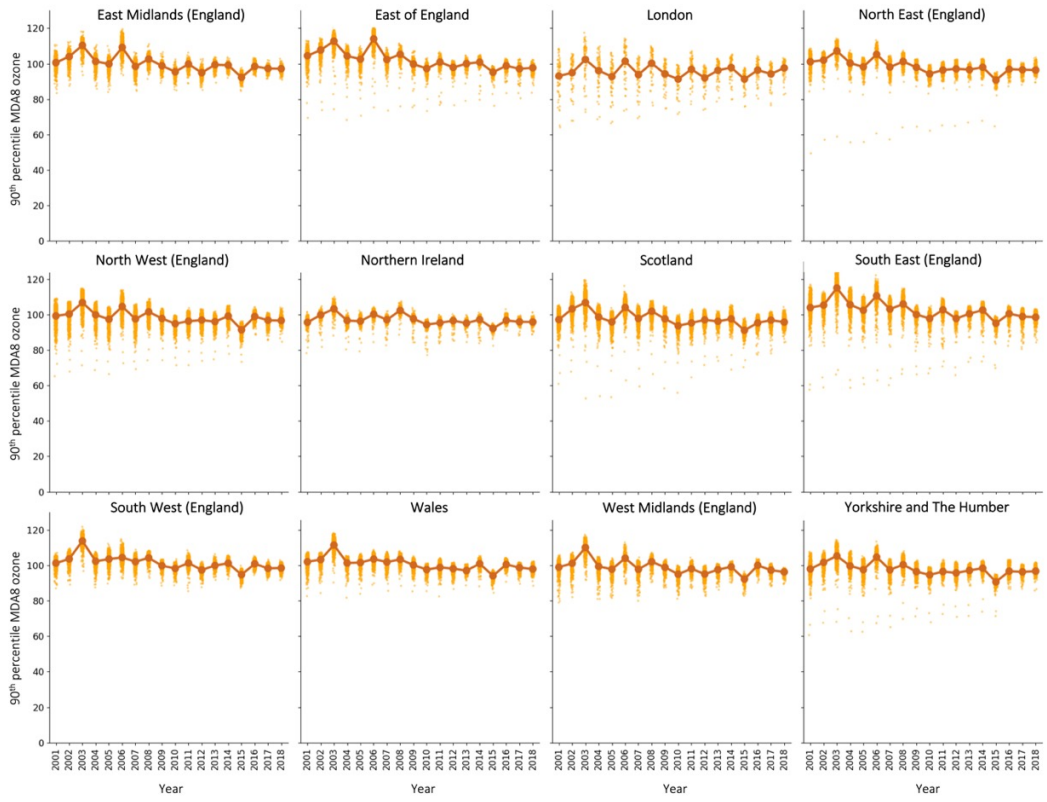

Figure B.10: Regional 90$^{\text{th}}$ percentile MDA8 ozone for the EMEP4UK surface, for 2001–2018. Background dots are individual cell estimates, larger foreground dots are the yearly average.

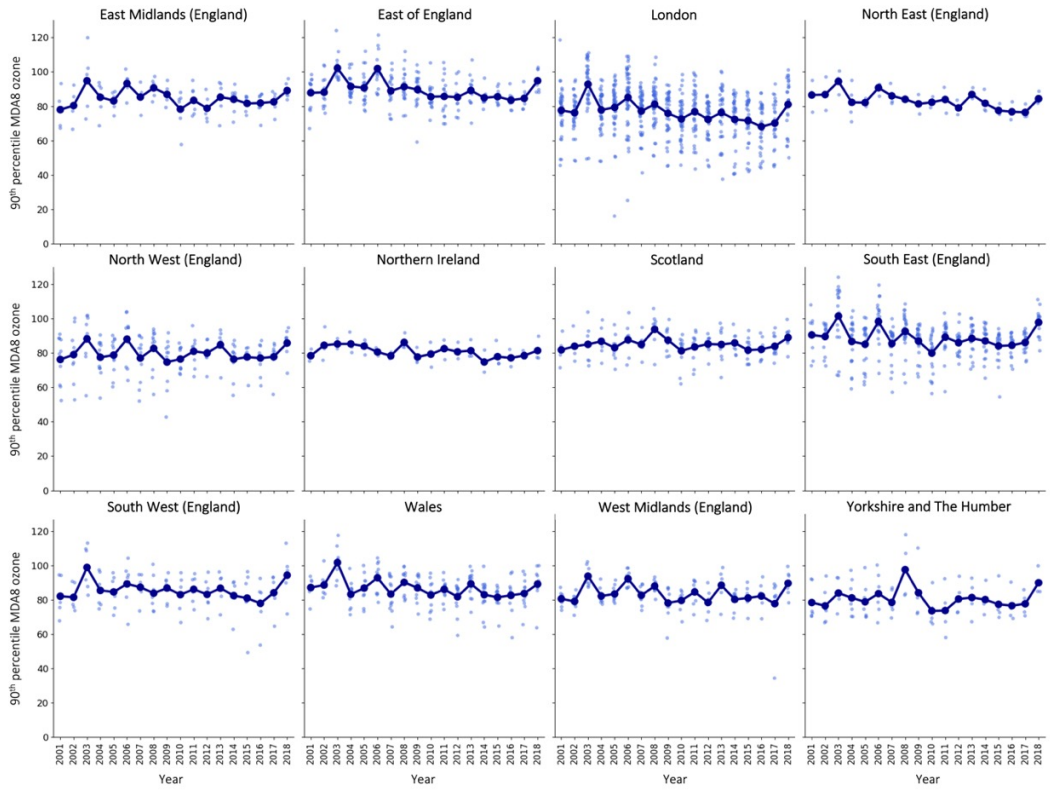

Figure B.11: Regional 90$^{\text{th}}$ percentile MDA8 ozone for the measurement data, for 2001–2018. Background dots are individual cell estimates, larger foreground dots are the yearly average.

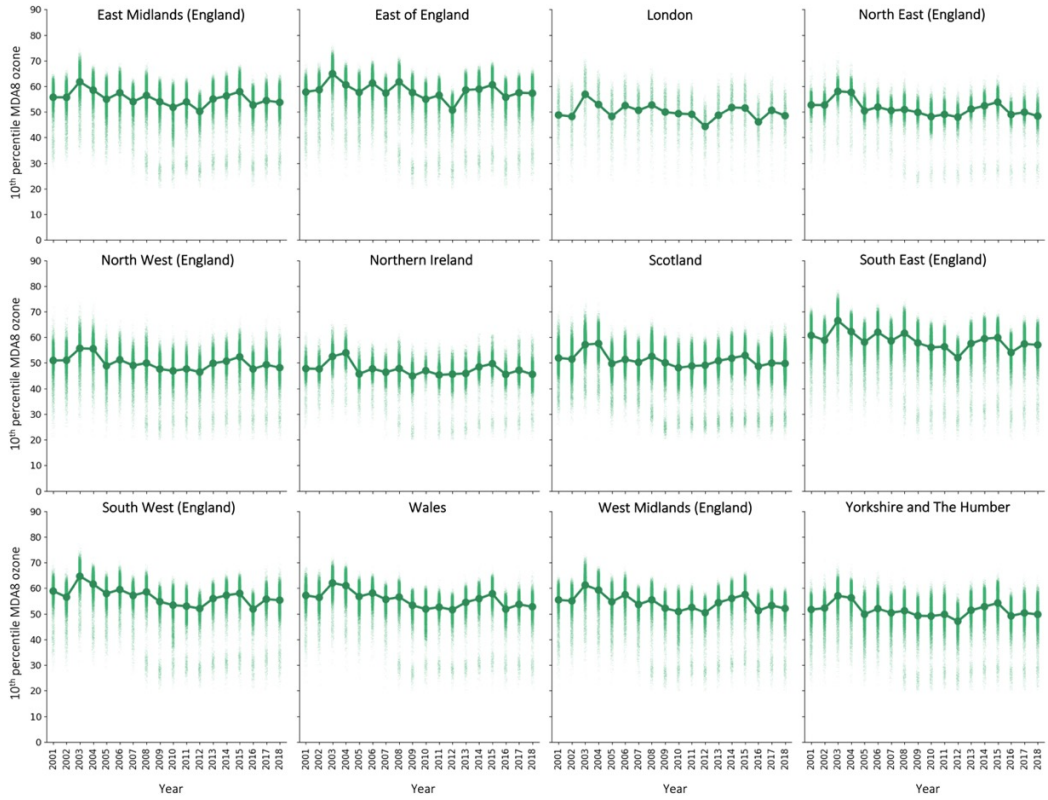

Figure B.12: Regional 10$^{\text{th}}$ percentile MDA8 ozone for the downscaled surface, for 2001–2018. Background dots are individual cell estimates, larger foreground dots are the yearly average.

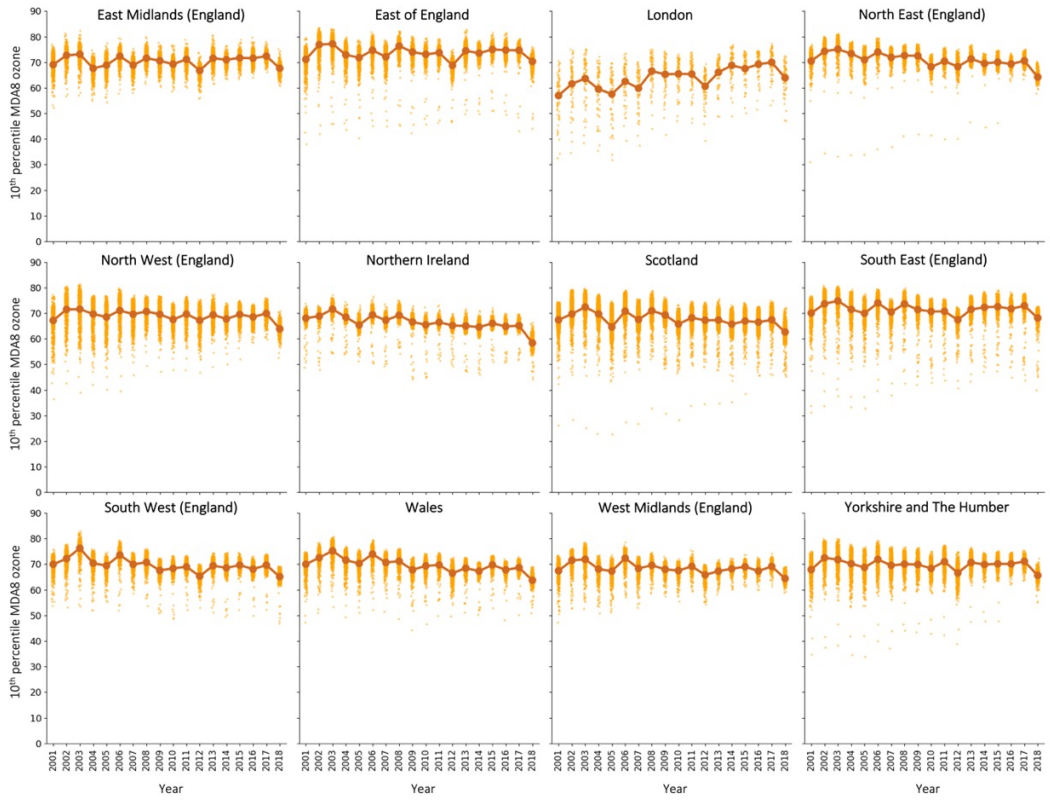

Figure B.13: Regional 10$^{\text{th}}$ percentile MDA8 ozone for the EMEP4UK surface, for 2001–2018. Background dots are individual cell estimates, larger foreground dots are the yearly average.

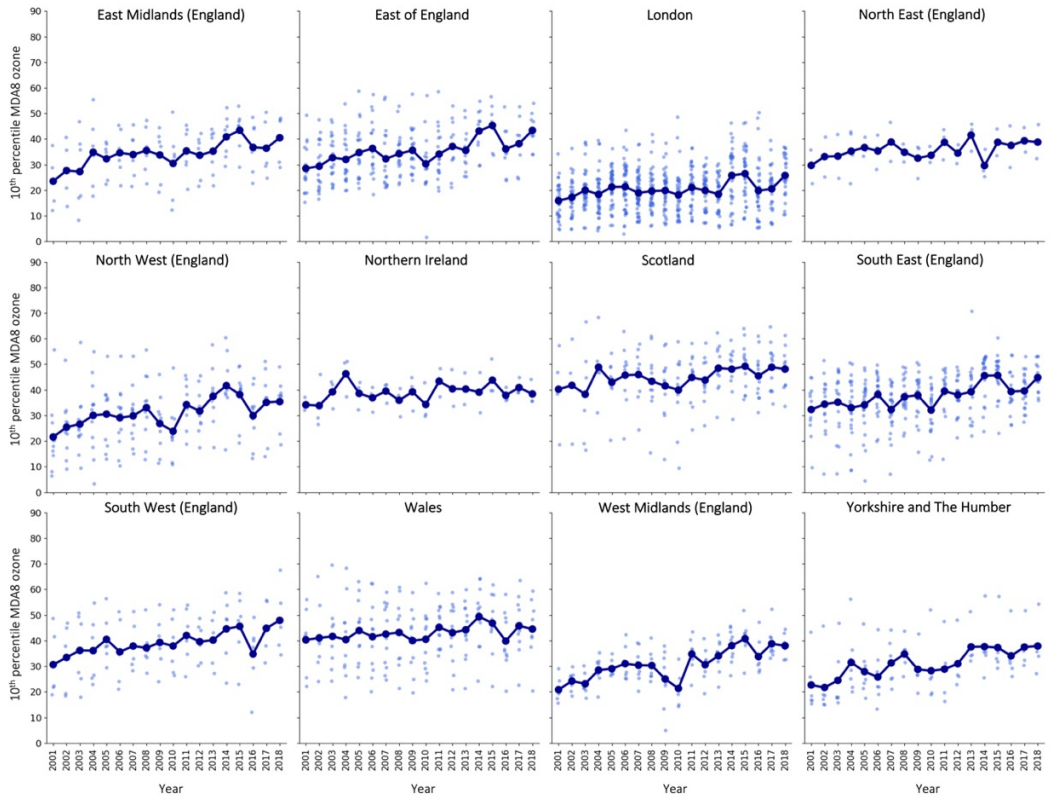

Figure B.14: Regional 10$^{\text{th}}$ percentile MDA8 ozone for the measurement data, for 2001–2018. Background dots are individual cell estimates, larger foreground dots are the yearly average.

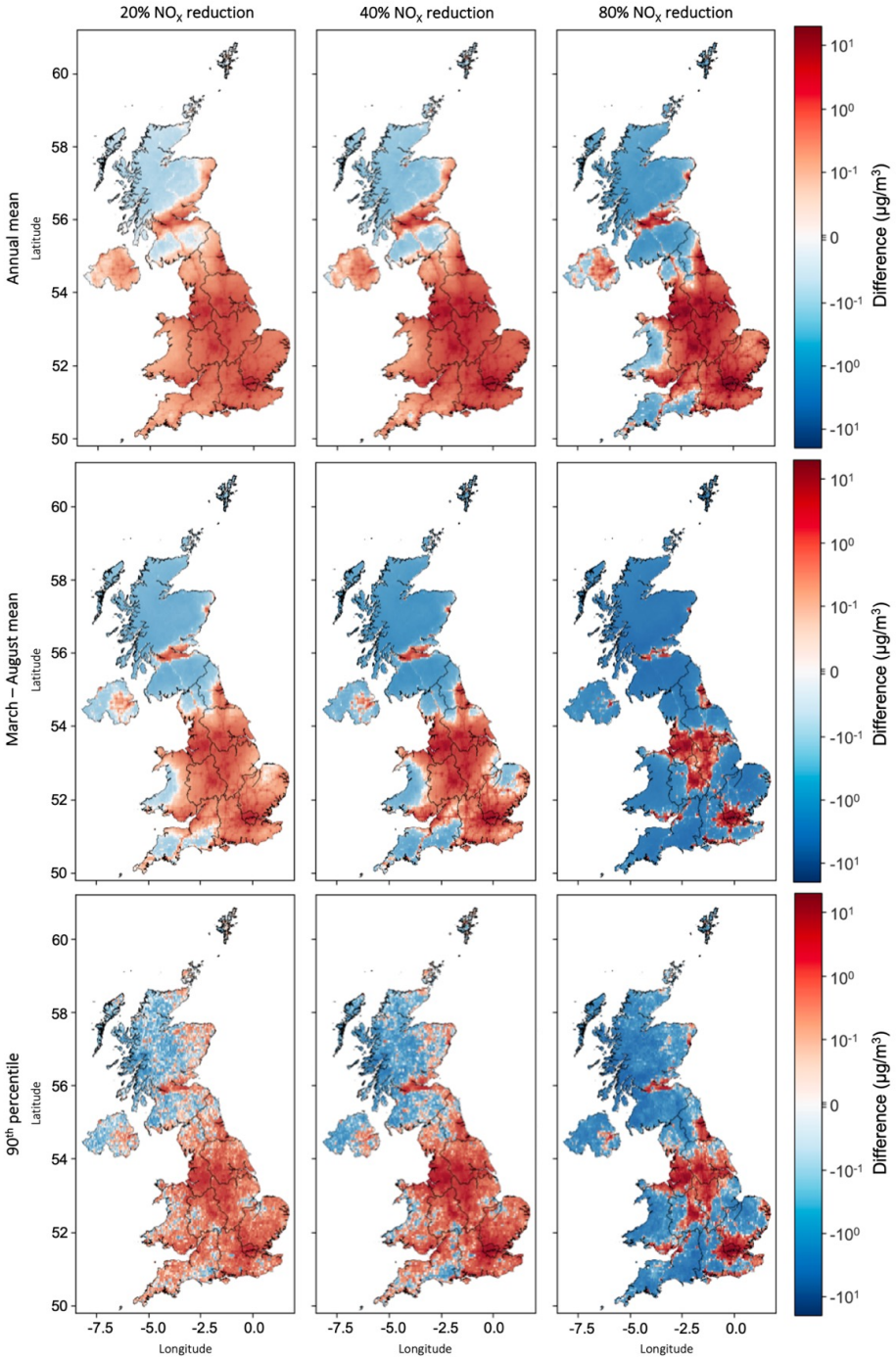

Figure B.15: Original EMEP4UK difference in annual mean (top), March-August mean (middle), and 90th percentile (bottom) MDA8 ozone compared to 2018 for three UK NO_X scenarios: 20% reduction in NO_X (left), 40% reduction in NO_X (middle), 80% reduction in NO_X (right).

References

Abdi-Oskouei, M. et al. (2020). "Sensitivity of Meteorological Skill to Selection of WRF-Chem Physical Parameterizations and Impact on Ozone Prediction During the Lake Michigan Ozone Study (LMOS)". In: *Journal of Geophysical Research: Atmospheres* 125.5, e2019JD031971. DOI: 10.1029/2019JD031971.

Abdul-Wahab, Sabah A., Charles S. Bakheit, and Saleh M. Al-Alawi (2005). "Principal component and multiple regression analysis in modelling of ground-level ozone and factors affecting its concentrations". In: *Environmental Modelling & Software* 20.10, pp. 1263–1271. DOI: 10.1016/j.envsoft.2004.09.001.

Adam, -Poupart Ariane et al. (2014). "Spatiotemporal Modeling of Ozone Levels in Quebec (Canada): A Comparison of Kriging, Land-Use Regression (LUR), and Combined Bayesian Maximum Entropy–LUR Approaches". In: *Environmental Health Perspectives* 122.9, pp. 970–976. DOI: 10.1289/ehp.1306566. (Visited on 04/04/2023).

Ainsworth, Elizabeth A. et al. (2012). "The Effects of Tropospheric Ozone on Net Primary Productivity and Implications for Climate Change". In: *Annual Review of Plant Biology* 63.1. _eprint: https://doi.org/10.1146/annurev-arplant-042110-103829, pp. 637–661. DOI: 10.1146/annurev-arplant-042110-103829.

Alkuwari, Farha A., Serge Guillas, and Yuhang Wang (2013). "Statistical downscaling of an air quality model using Fitted Empirical Orthogonal Functions". In: *Atmospheric Environment* 81, pp. 1–10. DOI: 10.1016/j.atmosenv.2013.08.031.

Allu, Sarat Kumar et al. (2020). "Seasonal ground level ozone prediction using multiple linear regression (MLR) model". In: *Modeling Earth Systems and Environment* 6.4, pp. 1981–1989. DOI: 10.1007/s40808-020-00810-0.

Anderson, H. Ross et al. (2004). "Meta-analysis of time-series studies and panel studies of particulate matter (PM) and ozone (O3) : report of a WHO task group". In: Accepted: 2014-03-19T23:06:05Z Number: EUR/04/5046026 Publisher: Copenhagen : WHO Regional Office for Europe. URL: https://apps.who.int/iris/handle/10665/107557 (visited on 11/16/2020).

AQEG (2021). *Ozone in the UK - Recent Trends and Future Projections*. URL: https://uk-air.defra.gov.uk/assets/documents/reports/cat09/2112200932_Ozone_in_the_UK_Recent_Trends_and_Future_Projections.pdf.

AQEG et al. (2013). *Mitigation of United Kingdom PM2.5 Concentrations*, p. 50. URL: https://uk-air.defra.gov.uk/assets/documents/reports/cat11/1508060903_DEF-PB14161_Mitigation_of_UK_PM25.pdf.

Avnery, Shiri et al. (2011). "Global crop yield reductions due to surface ozone exposure: 1. Year 2000 crop production losses and economic damage". In: *Atmospheric Environment* 45.13, pp. 2284–2296. DOI: 10.1016/j.atmosenv.2010.11.045.

Bandyopadhyay, G. and S. Chattopadhyay (2007). "Single hidden layer artificial neural network models versus multiple linear regression model in forecasting the time series of total ozone". In: *International Journal of Environmental Science & Technology* 4.1, pp. 141–149. DOI: 10.1007/BF03325972.

Bell, Michelle L., Devra L. Davis, and Tony Fletcher (2004). "A retrospective assessment of mortality from the London smog episode of 1952: the role of influenza and pollution." In: *Environmental Health Perspectives* 112.1. Publisher: Environmental Health Perspectives, pp. 6–8. DOI: 10.1289/ehp.6539.

Bergstra, J, D Yamins, and D D Cox (2013). "Making a Science of Model Search: Hyperparameter Optimization in Hundreds of Dimensions for Vision Architectures". In: Proceedings of the 30 th International Conference on Machine Learning, vol. 28. Atlanta, Georgia, USA: JMLR: W&CP, p. 9. URL: https://proceedings.mlr.press/v28/bergstra13.html.

Bertrand, Jean-Maxime et al. (2022). "Technical note: Improving the European air quality forecast of Copernicus Atmosphere Monitoring Service using machine learning techniques". In: *Atmospheric Chemistry and Physics Discussions*, pp. 1–28. DOI: 10.5194/acp-2022-767.

Betancourt, Clara et al. (2022). "Global, high-resolution mapping of tropospheric ozone – explainable machine learning and impact of uncertainties". In: *Geoscientific Model Development* 15.11, pp. 4331–4354. DOI: 10.5194/gmd-15-4331-2022.

Bloomer, Bryan J., Jeffrey W. Stehr, et al. (2009). "Observed relationships of ozone air pollution with temperature and emissions". In: *Geophysical Research Letters* 36.9. _eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2009GL037308. DOI: 10.1029/2009GL037308.

Bloomer, Bryan J., Konstantin Y. Vinnikov, and Russell R. Dickerson (2010). "Changes in seasonal and diurnal cycles of ozone and temperature in the eastern U.S." In: *Atmospheric Environment* 44.21, pp. 2543–2551. DOI: 10.1016/j.atmosenv.2010.04.031.

Bloomfield, Peter et al. (1996). "Accounting for meteorological effects in measuring urban ozone levels and trends". In: *Atmospheric Environment* 30.17, pp. 3067–3077. DOI: 10.1016/1352-2310(95)00347-9.

Bravo, Mercedes A. et al. (2016). "Racial isolation and exposure to airborne particulate matter and ozone in understudied US populations: Environmental justice applications of downscaled numerical model output". In: *Environment International* 92-93, pp. 247–255. DOI: 10.1016/j.envint.2016.04.008.

Butler, Thomas J. et al. (2011). "Response of ozone and nitrate to stationary source NOx emission reductions in the eastern USA". In: *Atmospheric Environment* 45.5, pp. 1084–1094. DOI: 10.1016/j.atmosenv.2010.11.040.

Bzdok, Danilo, Naomi Altman, and Martin Krzywinski (2018). *Statistics versus machine learning — Nature Methods*. URL: https://www.nature.com/articles/nmeth.4642.

Carnell, E et al. (2019). "Modelling public health improvements as a result of air pollution control policies in the UK over four decades—1970 to 2010". In: *Environmental Research Letters* 14.7, p. 074001. DOI: 10.1088/1748-9326/ab1542.

Carro-Calvo, Leopoldo et al. (2017). "Spatial clustering and meteorological drivers of summer ozone in Europe". In: *Atmospheric Environment* 167, pp. 496–510. DOI: 10.1016/j.atmosenv.2017.08.050.

Carslaw, David C. and Karl Ropkins (2012). "openair — An R package for air quality data analysis". In: https://cran.r-project.org/web/packages/openair/index.html. URL: https://www.sciencedirect.com/science/article/pii/S1364815211002064.

Chaloulakou, Archontoula, Michaela Saisana, and Nikolas Spyrellis (2003). "Comparative assessment of neural networks and regression models for forecasting summertime ozone in Athens". In: *Science of The Total Environment* 313.1, pp. 1–13. DOI: 10.1016/S0048-9697(03)00335-8. (Visited on 04/04/2023).

Chapman, S. C., N. W. Watkins, and D. A. Stainforth (2019). "Warming Trends in Summer Heatwaves". In: *Geophysical Research Letters* 46.3, pp. 1634–1640. DOI: 10.1029/2018GL081004.

Chen, Tianqi and Carlos Guestrin (2016). "XGBoost: A Scalable Tree Boosting System". In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '16: The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco California USA, pp. 785–794. DOI: 10.1145/2939672.2939785.

Chi, Yufeng et al. (2023). "Sequential spatiotemporal distribution of $PM_{2.5}$, SO_2 and Ozone in China from 2015 to 2020". In: *Earth System Science Data Discussions*, pp. 1–38. DOI: 10.5194/essd-2023-76.

Christidis, Nikolaos, Gareth S. Jones, and Peter A. Stott (2015). "Dramatically increasing chance of extremely hot summers since the 2003 European heatwave". In: *Nature Climate Change* 5.1, pp. 46–50. DOI: 10.1038/nclimate2468.

Clifton, Olivia E. et al. (2020). "Dry Deposition of Ozone Over Land: Processes, Measurement, and Modeling". In: *Reviews of Geophysics* 58.1. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2019RG000670, e2019RG000670. DOI: 10.1029/2019RG000670.

Cohen, Aaron J et al. (2017). "Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: an analysis of data from the Global Burden of Diseases Study 2015". In: *The Lancet* 389.10082, pp. 1907–1918. DOI: 10.1016/S0140-6736(17)30505-6.

COMEAP (2015). *COMEAP: quantification of mortality and hospital admissions associated with ground-level ozone*. GOV.UK. URL: https://www.gov.uk/government/publications/comeap-quantification-of-mortality-and-hospital-admissions-associated-with-ground-level-ozone.

Cooper, O. R. et al. (2014). "Global distribution and trends of tropospheric ozone: An observation-based review". In: *Elementa: Science of the Anthropocene* 2, p. 000029. DOI: 10.12952/journal.elementa.000029.

Cooper, Owen R. et al. (2012). "Long-term ozone trends at rural ozone monitoring sites across the United States, 1990–2010". In: *Journal of Geophysical Research: Atmospheres* 117 (D22). DOI: 10.1029/2012JD018261.

Davison, A. C. and R. L. Smith (1990). "Models for Exceedances Over High Thresholds". In: *Journal of the Royal Statistical Society: Series B (Methodological)* 52.3. _eprint: https://rss.onlinelibrary.wiley.com/doi/pdf/10.1111/j.2517-6161.1990.tb01796.x, pp. 393–425. DOI: 10.1111/j.2517-6161.1990.tb01796.x.

Defra (2020). *UK and EU Air Quality Limits- Defra, UK*. Archive Location: UK; United Kingdom Publisher: Department for Environment, Food and Rural Affairs (Defra). URL: https://uk-air.defra.gov.uk/air-pollution/uk-eu-limits.

— (2021a). *Emissions of air pollutants in the UK – Nitrogen oxides (NOx)*. GOV.UK. URL: https://www.gov.uk/government/statistics/emissions-of-air-pollutants/emissions-of-air-pollutants-in-the-uk-nitrogen-oxides-nox.

— (2021b). *Emissions of air pollutants in the UK – Non-methane volatile organic compounds (NMVOCs)*. GOV.UK. URL: https://www.gov.uk/government/statistics/emissions-of-air-pollutants/emissions-of-air-pollutants-in-the-uk-non-methane-volatile-organic-compounds-nmvocs.

— (2021c). *Ozone in the UK Recent Trends and Future Projections*. Archive Location: UK; United Kingdom Publisher: Department for Environment, Food and Rural Affairs (Defra). URL: https://uk-air.defra.gov.uk/library/reports?report_id=1064.

— (2022). *Air pollution in the UK 2021 - Compliance Assessment Summary*. URL: https://uk-air.defra.gov.uk/library/annualreport/assets/documents/annualreport/air_pollution_uk_2021_Compliance_Assessment_Summary_Issue1.pdf.

— (2023). *Brief history- Defra, UK*. Archive Location: UK; United Kingdom Publisher: Department for Environment, Food and Rural Affairs (Defra). URL: https://uk-air.defra.gov.uk/networks/brief-history.

Deng, Tuo et al. (2022). "Clustering-based spatial transfer learning for short-term ozone forecasting". In: *Journal of Hazardous Materials Advances* 8, p. 100168. DOI: 10.1016/j.hazadv.2022.100168.

Derwent, R. G. et al. (1998). "Observation and interpretation of the seasonal cycles in the surface concentrations of ozone and carbon monoxide at mace head, Ireland from 1990 to 1994". In: *Atmospheric Environment* 32.2, pp. 145–157. DOI: 10.1016/S1352-2310(97)00338-5.

Derwent, Richard G. et al. (2018). "Long-term trends in ozone in baseline and European regionally-polluted air at Mace Head, Ireland over a 30-year period". In: *Atmospheric Environment* 179, pp. 279–287. DOI: 10.1016/j.atmosenv.2018.02.024.

Diaz, Florencia M. R. et al. (2020). "Ozone Trends in the United Kingdom over the Last 30 Years". In: *Atmosphere* 11.5. Number: 5 Publisher: Multidisciplinary Digital Publishing Institute, p. 534. DOI: 10.3390/atmos11050534.

Díaz, Julio et al. (2018). "Short-term effect of tropospheric ozone on daily mortality in Spain". In: *Atmospheric Environment* 187, pp. 107–116. DOI: 10.1016/j.atmosenv.2018.05.059. URL: http://www.sciencedirect.com/science/article/pii/S1352231018303698.

Doherty, Ruth M. et al. (2009). "Current and future climate- and air pollution-mediated impacts on human health". In: *Environmental Health* 8.1, S8. DOI: 10.1186/1476-069X-8-S1-S8.

Du, Jianbang et al. (2022). "Forecasting ground-level ozone concentration levels using machine learning". In: *Resources, Conservation and Recycling* 184, p. 106380. DOI: 10.1016/j.resconrec.2022.106380.

Dueñas, C. et al. (2005). "Stochastic model to forecast ground-level ozone concentration at urban and rural areas". In: *Chemosphere* 61.10, pp. 1379–1389. DOI: 10.1016/j.chemosphere.2005.04.079.

Eastoe, Emma F. (2009). "A hierarchical model for non-stationary multivariate extremes: a case study of surface-level ozone and NOx data in the UK". In: *Environmetrics* 20.4. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/env.938, pp. 428–444. DOI: 10.1002/env.938.

Eastoe, Emma F. and Jonathan A. Tawn (2009). "Modelling Non-Stationary Extremes with Application to Surface Level Ozone". In: *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 58.1, pp. 25–45. URL: https://www.jstor.org/stable/25578145 (visited on 04/07/2020).

EIDC, Environmental Information Data Centre (2021). *Air pollution removed by vegetation in the UK, 2015*. URL: https://www.data.gov.uk/dataset/90afc84e-3354-4257-9aa8-ea106c557681/air-pollution-removed-by-vegetation-in-the-uk-2015.

Entwistle, Jon et al. (1997). "The magnitude and extent of elevated ozone concentrations around the coasts of the British Isles". In: *Atmospheric Environment* 31.13. DOI: 10.1016/S1352-2310(97)00022-8.

Environmental Prediction, National Centers for (2000). *NCEP FNL Operational Model Global Tropospheric Analyses, continuing from July 1999*. DOI: 10.5065/D6M043C6.

Eslami, Ebrahim et al. (2020). "A real-time hourly ozone prediction system using deep convolutional neural network". In: *Neural Computing and Applications* 32.13, pp. 8783–8797. DOI: 10.1007/s00521-019-04282-x.

Fernández, Alberto et al. (2018). *Learning from Imbalanced Data Sets*. Cham. DOI: 10.1007/978-3-319-98074-4.

Finch, Douglas P. and Paul I. Palmer (2020). "Increasing ambient surface ozone levels over the UK accompanied by fewer extreme events". In: *Atmospheric Environment* 237, p. 117627. DOI: 10.1016/j.atmosenv.2020.117627.

Fisher, Ronald Aylmer and L. H. C. Tippett (1928). "063: Limiting Forms of the Frequency Distribution of the Largest of Smallest Member of a Sample." In: URL: https://digital.library.adelaide.edu.au/dspace/handle/2440/15198.

Fleming, Zoë L. et al. (2018). "Tropospheric Ozone Assessment Report: Present-day ozone distribution and trends relevant to human health". In: *Elementa: Science of the Anthropocene* 6.12. Ed. by Detlev Helmig and Alastair Lewis. DOI: 10.1525/elementa.273.

Friedman, Jerome H. (2001). "Greedy Function Approximation: A Gradient Boosting Machine". In: *The Annals of Statistics* 29.5. Publisher: Institute of Mathematical Statistics, pp. 1189–1232. URL: https://www.jstor.org/stable/2699986 (visited on 04/29/2022).

Gao, Y. et al. (2013). "The impact of emission and climate change on ozone in the United States under representative concentration pathways (RCPs)". In: *Atmospheric Chemistry and Physics* 13.18, pp. 9607–9621. DOI: 10.5194/acp-13-9607-2013.

Gariazzo, Claudio et al. (2020). "A multi-city air pollution population exposure study: Combined use of chemical-transport and random-Forest models with

dynamic population data". In: *Science of The Total Environment* 724, p. 138102. DOI: 10.1016/j.scitotenv.2020.138102.

Gauthier-Manuel, Honorine et al. (2022). "Improvement of downscaled ozone concentrations from the transnational scale to the kilometric scale: Need, interest and new insights". In: *Environmental Research* 210, p. 112947. DOI: 10.1016/j.envres.2022.112947.

Gégo, Edith et al. (2007). "Observation-Based Assessment of the Impact of Nitrogen Oxides Emissions Reductions on Ozone Air Quality over the Eastern United States". In: *Journal of Applied Meteorology and Climatology* 46.7, pp. 994–1008. DOI: 10.1175/JAM2523.1.

Geiss, Andrew, Sam J. Silva, and Joseph C. Hardin (2022). "Downscaling atmospheric chemistry simulations with physically consistent deep learning". In: *Geoscientific Model Development* 15.17, pp. 6677–6694. DOI: 10.5194/gmd-15-6677-2022.

Gelfand, Alan E. and Erin M. Schliep (2016). "Spatial statistics and Gaussian processes: A beautiful marriage". In: *Spatial Statistics*. Spatial Statistics Avignon: Emerging Patterns 18, pp. 86–104. DOI: 10.1016/j.spasta.2016.03.006.

Ghazali, Nurul Adyani et al. (2010). "Transformation of nitrogen dioxide into ozone and prediction of ozone concentrations using multiple linear regression techniques". In: *Environmental Monitoring and Assessment* 165.1, pp. 475–489. ISSN: 1573-2959. DOI: 10.1007/s10661-009-0960-3.

Ghude, Sachin D. et al. (2014). "Reductions in India's crop yield due to ozone". In: *Geophysical Research Letters* 41.15, pp. 5685–5691. DOI: 10.1002/2014GL060930.

Gong, Bing and Joaquín Ordieres-Meré (2016). "Prediction of daily maximum ozone threshold exceedances by preprocessing and ensemble artificial intelligence techniques: Case study of Hong Kong". In: *Environmental Modelling & Software* 84, pp. 290–303. DOI: 10.1016/j.envsoft.2016.06.020.

Gouldsbrough, Lily et al. (2022). "A temperature dependent extreme value analysis of UK surface ozone, 1980–2019". In: *Atmospheric Environment* 273, p. 118975. DOI: 10.1016/j.atmosenv.2022.118975.

Granier, Claire and Guy P. Brasseur (2003). "The impact of road traffic on global tropospheric ozone". In: *Geophysical Research Letters* 30.2. DOI: 10.1029/2002GL015972.

Grell, Georg A. et al. (2005). "Fully coupled "online" chemistry within the WRF model". In: *Atmospheric Environment* 39.37, pp. 6957–6975. DOI: 10.1016/j.atmosenv.2005.04.027.

Griffiths, Paul T. et al. (2021). "Tropospheric ozone in CMIP6 simulations". In: *Atmospheric Chemistry and Physics* 21.5, pp. 4187–4218. DOI: 10.5194/acp-21-4187-2021.

Gu, Yixuan et al. (2020). "Observed dependence of surface ozone on increasing temperature in Shanghai, China". In: *Atmospheric Environment* 221, p. 117108. DOI: 10.1016/j.atmosenv.2019.117108.

Guillas, Serge et al. (2008). "Statistical correction and downscaling of chemical transport model ozone forecasts over Atlanta". In: *Atmospheric Environment* 42.6, pp. 1338–1348. DOI: 10.1016/j.atmosenv.2007.10.027.

Gyarmati-Szabó, János, Leonid V. Bogachev, and Haibo Chen (2017). "Nonstationary POT modelling of air pollution concentrations: Statistical analysis of the traffic and meteorological impact". In: *Environmetrics* 28.5, e2449. DOI: 10.1002/env.2449.

Halios, Christos H. et al. (2022). "Chemicals in European residences – Part I: A review of emissions, concentrations and health effects of volatile organic compounds (VOCs)". In: *Science of The Total Environment* 839, p. 156201. DOI: 10.1016/j.scitotenv.2022.156201.

Haman, C. L. et al. (2014). "Relationship between boundary layer heights and growth rates with ground-level ozone in Houston, Texas". In: *Journal of*

Geophysical Research: Atmospheres 119.10, pp. 6230–6245. DOI: 10.1002/2013JD020473.

Hanlon, Helen M. et al. (2021). "Future changes to high impact weather in the UK". In: *Climatic Change* 166.3, p. 50. DOI: 10.1007/s10584-021-03100-5.

Harmens, Harry et al. (2015). *Air pollution and vegetation: ICP Vegetation annual report 2014/2015*. URL: http://icpvegetation.ceh.ac.uk/publications/documents/FinalICPVegetationannualreport2014-15.pdf.

Hazarika, Suhasini, Parashmoni Borah, and Amit Prakash (2019). "The assessment of return probability of maximum ozone concentrations in an urban environment of Delhi: A Generalized Extreme Value analysis approach". In: *Atmospheric Environment* 202, pp. 53–63. DOI: 10.1016/j.atmosenv.2019.01.021.

Heagle, A S (1989). "Ozone and Crop Yield*". In: *Annual Review of Phytopathology* 27.1. _eprint: https://doi.org/10.1146/annurev.py.27.090189.002145, pp. 397–423. DOI: 10.1146/annurev.py.27.090189.002145.

Hollis, Dan et al. (2019). "HadUK-Grid—A new UK dataset of gridded climate observations". In: *Geoscience Data Journal* 6.2, pp. 151–159. DOI: 10.1002/gdj3.78.

Hooyberghs, Jef et al. (2006). "Spatial interpolation of ambient ozone concentrations from sparse monitoring points in Belgium". In: *Journal of Environmental Monitoring* 8.11, pp. 1129–1135. DOI: 10.1039/B612607N.

Hu, Xiaomin et al. (2022). "Estimation of the Near-Surface Ozone Concentration with Full Spatiotemporal Coverage across the Beijing-Tianjin-Hebei Region Based on Extreme Gradient Boosting Combined with a WRF-Chem Model". In: *Atmosphere* 13.4, p. 632. DOI: 10.3390/atmos13040632. URL: https://www.mdpi.com/2073-4433/13/4/632 (visited on 10/05/2022).

Hundecha, Y. et al. (2008). "A Nonstationary Extreme Value Analysis for the Assessment of Changes in Extreme Annual Wind Speed over the Gulf of St. Lawrence, Canada". In: *Journal of Applied Meteorology and Climatology* 47.11, pp. 2745–2759. DOI: 10.1175/2008JAMC1665.1.

Jenkin, Michael E. (2008). "Trends in ozone concentration distributions in the UK since 1990: Local, regional and global influences". In: *Atmospheric Environment* 42.21, pp. 5434–5445. ISSN: 1352-2310. DOI: 10.1016/j.atmosenv.2008.02.036.

Jenkinson, A.F. and F.P. Collison (1997). "An initial climatology of gales over the North Sea". In: *Synoptic Climatology Branch Memorandum*. 62. Meteorological Office, Bracknell.

Ji, Meng, Daniel S. Cohan, and Michelle L. Bell (2011). "Meta-analysis of the association between short-term exposure to ambient ozone and respiratory hospital admissions". In: *Environmental Research Letters* 6.2, p. 024006. DOI: 10.1088/1748-9326/6/2/024006.

Jiang, Ningbo and Matthew Riley (2015). "Exploring the Utility of the Random Forest Method for Forecasting Ozone Pollution in SYDNEY". In: *Journal of Environment Protection and Sustainable Development* 1, pp. 245–254.

Karmakar, Sumitra et al. (2022). "Effects of Ozone on Plant Health and Environment: A Mini Review". In: *Research Journal of Agricultural Sciences An International Journal* 13, pp. 612–619.

Kavassalis, Sarah C. and Jennifer G. Murphy (2017). "Understanding ozone-meteorology correlations: A role for dry deposition". In: *Geophysical Research Letters* 44.6, pp. 2922–2931. DOI: 10.1002/2016GL071791.

Ke, Guolin et al. (2017). "LightGBM: A Highly Efficient Gradient Boosting Decision Tree". In: *Advances in Neural Information Processing Systems*. Vol. 30. Curran Associates, Inc.

Keller, Christoph A. et al. (2021). "Global impact of COVID-19 restrictions on the surface concentrations of nitrogen dioxide and ozone". In: *Atmospheric Chemistry and Physics* 21.5. Publisher: Copernicus GmbH, pp. 3555–3592. ISSN: 1680-7316. DOI: 10.5194/acp-21-3555-2021.

Kendon, Mike et al. (2019). "State of the UK climate 2018". In: *International Journal of Climatology* 39 (S1), pp. 1–55. DOI: 10.1002/joc.6213.

Kim, Young Min, Stuart Harrad, and Roy M. Harrison (2001). "Concentrations and Sources of VOCs in Urban Domestic and Public Microenvironments". In: *Environmental Science & Technology* 35.6. Publisher: American Chemical Society, pp. 997–1004. DOI: 10.1021/es000192y.

Klein, Petra M., Xiao-Ming Hu, and Ming Xue (2014). "Impacts of Mixing Processes in Nocturnal Atmospheric Boundary Layer on Urban Ozone Concentrations". In: *Boundary-Layer Meteorology* 150.1, pp. 107–130. DOI: 10.1007/s10546-013-9864-4.

Kleinert, Felix, Lukas H. Leufen, and Martin G. Schultz (2021). "IntelliO3-ts v1.0: a neural network approach to predict near-surface ozone concentrations in Germany". In: *Geoscientific Model Development* 14.1, pp. 1–25. DOI: 10.5194/gmd-14-1-2021.

Kumar, Awkash et al. (2017). "Application of AERMOD for short-term air quality prediction with forecasted meteorology using WRF model". In: *Clean Technologies and Environmental Policy* 19.7, pp. 1955–1965. DOI: 10.1007/s10098-017-1379-0.

Kumar, Krishan et al. (2004). "Forecasting Daily Maximum Surface Ozone Concentrations in Brunei Darussalam—An ARIMA Modeling Approach". In: *Journal of the Air & Waste Management Association* 54.7, pp. 809–814. DOI: 10.1080/10473289.2004.10470949.

Kumar, Ujjwal and Koen De Ridder (2010). "GARCH modelling in association with FFT–ARIMA to forecast ozone episodes". In: *Atmospheric Environment* 44.34, pp. 4252–4265. DOI: 10.1016/j.atmosenv.2010.06.055.

Lamb, Hubert Horace (1972). *British Isles Weather Types and a Register of the Daily Sequence of Circulation Patterns, 1861-1971*. Google-Books-ID: pAJsQgAACAAJ. H.M. Stationery Office. 85 pp.

Lang, Polly (2020). "New approaches to the statistical analysis of air quality network data: insights from application to national and regional UK networks". PhD thesis. University of York. URL: https://etheses.whiterose.ac.uk/28164/.

Lauwaet, D. et al. (2013). "Impact of nesting resolution jump on dynamical downscaling ozone concentrations over Belgium". In: *Atmospheric Environment* 67, pp. 46–52. DOI: 10.1016/j.atmosenv.2012.10.034.

Lee, James D. et al. (2006). "Ozone photochemistry and elevated isoprene during the UK heatwave of august 2003". In: *Atmospheric Environment* 40.39, pp. 7598–7613. DOI: 10.1016/j.atmosenv.2006.06.057.

Leeson, Amber A., Emma Eastoe, and Xavier Fettweis (2018). "Extreme temperature events on Greenland in observations and the MAR regional climate model". In: *The Cryosphere* 12.3, pp. 1091–1102. DOI: 10.5194/tc-12-1091-2018.

Lelieveld, Jos and Frank J. Dentener (2000). "What controls tropospheric ozone?" In: *Journal of Geophysical Research: Atmospheres* 105 (D3), pp. 3531–3551. DOI: 10.1029/1999JD901011.

Levine, Joel S. et al. (1984). "Tropospheric sources of NOx: Lightning and biology". In: *Atmospheric Environment (1967)*. CACGP Symposium on Tropospheric Chemistry With Emphasis on Sulphur and Nitrogen Cycles and the Chemistry of Clouds and Precipitation 18.9, pp. 1797–1804. DOI: 10.1016/0004-6981(84)90355-X.

Li, Qinbin et al. (2002). "Transatlantic transport of pollution and its effects on surface ozone in Europe and North America". In: *Journal of Geophysical Research: Atmospheres* 107 (D13), ACH 4–1–ACH 4–21. DOI: 10 . 1029 / 2001JD001422.

Li, Rui et al. (2020). "Satellite-based estimation of full-coverage ozone (O3) concentration and health effect assessment across Hainan Island". In: *Journal of Cleaner Production* 244, p. 118773. DOI: 10.1016/j.jclepro.2019.118773.

Li, Ying et al. (2013). "Importance of NOx control for peak ozone reduction in the Pearl River Delta region". In: *Journal of Geophysical Research: Atmospheres* 118.16, pp. 9428–9443. DOI: 10.1002/jgrd.50659.

Li, Ying-Ruo et al. (2021). "Application of ARIMA Model for Mid- and Long-term Forecasting of Ozone Concentration". In: *Huan jing ke xue= Huanjing kexue* 42.7, pp. 3118–3126. DOI: 10.13227/j.hjkx.202011237.

Lin, Chun et al. (2017). "Spatiotemporal evaluation of EMEP4UK-WRF v4.3 atmospheric chemistry transport simulations of health-related metrics for NO_2, O_3, PM_{10}, and $PM_{2.5}$ for 2001–2010". In: *Geoscientific Model Development* 10.4, pp. 1767–1787. DOI: 10.5194/gmd-10-1767-2017.

Lin, Meiyun et al. (2020). "Vegetation feedbacks during drought exacerbate ozone air pollution extremes in Europe". In: *Nature Climate Change* 10.5, pp. 444–451. DOI: 10.1038/s41558-020-0743-y.

Liška, Tomáš (2021). "Effect of workplace mobility on air pollution exposure and its inequality in the UK". Accepted: 2021. PhD thesis. University of Edinburgh. DOI: 10.7488/era/1654.

Liu, Liangke et al. (2023). *Near Real-Time Distribution of Ozone in China from 2013 to 2020 and Agricultural Impacts*. EGU23-6562. Conference Name: EGU23. Copernicus Meetings. DOI: 10.5194/egusphere-egu23-6562.

Liu, Riyang et al. (2020). "Spatiotemporal distributions of surface ozone levels in China from 2005 to 2017: A machine learning approach". In: *Environment International* 142, p. 105823. DOI: 10.1016/j.envint.2020.105823.

Liu, Zhenze et al. (2022). "Correcting ozone biases in a global chemistry–climate model: implications for future ozone". In: *Atmospheric Chemistry and Physics* 22.18, pp. 12543–12557. DOI: 10.5194/acp-22-12543-2022.

Lowe, Jason et al. (2018). "UKCP18-Overview-report". In: URL: https://www.researchgate.net/profile/Stephen-E-Belcher/publication/345815169_UKCP18-Overview-report/links/5faed14aa6fdcc9ae04dc04e/UKCP18-Overview-report.pdf.

Lu, Haoxian et al. (2019). "Overview on the spatial–temporal characteristics of the ozone formation regime in China". In: *Environmental Science: Processes &*

Impacts 21.6. Publisher: Royal Society of Chemistry, pp. 916–929. DOI: 10.1039/C9EM00098D.

Lundberg, Scott and Su-In Lee (2017). "A Unified Approach to Interpreting Model Predictions". In: *arXiv:1705.07874 [cs, stat]*. arXiv: 1705.07874.

Mallet, Vivien, Gilles Stoltz, and Boris Mauricette (2009). "Ozone ensemble forecast with machine learning algorithms". In: *Journal of Geophysical Research: Atmospheres* 114 (D5). DOI: 10.1029/2008JD009978.

Mann, R C (1997). "A high resolution atmospheric emissions inventory for road vehicle sources and the factors influencing its accuracy and sensitivity". In: *Transactions on Ecology and the Environment* 15. URL: https://www.witpress.com/Secure/elibrary/papers/AIR97/AIR97095FU.pdf.

Masseran, Nurulkamal et al. (2015). "Modeling air quality in main cities of Peninsular Malaysia by using a generalized Pareto model". In: *Environmental Monitoring and Assessment* 188.1, p. 65. DOI: 10.1007/s10661-015-5070-9.

Meijer, Johan R et al. (2018). "Global patterns of current and future road infrastructure". In: *Environmental Research Letters* 13.6, p. 064006. DOI: 10.1088/1748-9326/aabd42.

Milojevic, Ai et al. (2017). "Socioeconomic and urban-rural differentials in exposure to air pollution and mortality burden in England". In: *Environmental Health* 16.1, p. 104. DOI: 10.1186/s12940-017-0314-5.

Mohnen, V.A., W. Goldstein, and W.-C. Wang (1993). "Tropospheric Ozone and Climate Change". In: *Air & Waste* 43.10, pp. 1332–1334. DOI: 10.1080/1073161X.1993.10467207.

Monks, P. S. et al. (2015). "Tropospheric ozone and its precursors from the urban to the global scale from air quality to short-lived climate forcer". In: *Atmospheric Chemistry and Physics* 15.15. Publisher: European Geosciences Union, pp. 8889–8973. DOI: 10.5194/acp-15-8889-2015.

Monks, Paul S (2000). "A review of the observations and origins of the spring ozone maximum". In: *Atmospheric Environment* 34.21, pp. 3545–3561. ISSN: 1352-2310. DOI: 10.1016/S1352-2310(00)00129-1.

Moustris, K. P. et al. (2012). "Application of Multiple Linear Regression Models and Artificial Neural Networks on the Surface Ozone Forecast in the Greater Athens Area, Greece". In: *Advances in Meteorology* 2012, e894714. DOI: 10.1155/2012/894714.

Munir, Said, Haibo Chen, and Karl Ropkins (2012). "Modelling the impact of road traffic on ground level ozone concentration using a quantile regression approach". In: *Atmospheric Environment* 60, pp. 283–291. DOI: 10.1016/j.atmosenv.2012.06.043.

Napi, Nur Nazmi Liyana Mohd et al. (2020). "Multiple Linear Regression (MLR) and Principal Component Regression (PCR) for Ozone (O3) Concentrations Prediction". In: *IOP Conference Series: Earth and Environmental Science* 616.1, p. 012004. DOI: 10.1088/1755-1315/616/1/012004.

National Statistics, Office for (2018). *NUTS Level 1 (January 2018) Names and Codes in the United Kingdom*. URL: https://www.data.gov.uk/dataset/00d0c48d-afab-405e-98eb-5083aaae5fe8/nuts-level-1-january-2018-names-and-codes-in-the-united-kingdom.

Neal, L. S. et al. (2014). "Application of a statistical post-processing technique to a gridded, operational, air quality forecast". In: *Atmospheric Environment* 98, pp. 385–393. DOI: 10.1016/j.atmosenv.2014.09.004.

Nolte, Christopher G. et al. (2021). "Regional temperature-ozone relationships across the U.S. under multiple climate and emissions scenarios". In: *Journal of the Air & Waste Management Association* 71.10, pp. 1251–1264. DOI: 10.1080/10962247.2021.1970048.

Nuvolone, Daniela, Davide Petri, and Fabio Voller (2018). "The effects of ozone on human health". In: *Environmental Science and Pollution Research* 25.9, pp. 8074–8088. DOI: 10.1007/s11356-017-9239-3.

Office, Met et al. (2020). *HadUK-Grid Gridded Climate Observations on a 1km grid over the UK, v1.0.2.1 (1862-2019)*. In collab. with Centre For Environmental Data Analysis (CEDA) et al. Medium: application/xml Type: dataset. DOI: 10.5285/89908DFCB97B4A28976DF806B4818639.

Ogur, Eric and Sam Kariuki (2014). "Effect of Car Emissions on Human Health and the Environment". In: *International Journal of Applied Engineering Research* 9, pp. 11121–11128.

Olivier, J. G. J. et al. (1998). "Global air emission inventories for anthropogenic sources of NOx, NH3 and N2O in 1990". In: *Environmental Pollution*. Nitrogen, the Confer-N-s First International Nitrogen Conference 1998 102.1, pp. 135–148. DOI: 10.1016/S0269-7491(98)80026-2.

Onwukwe, Chibuike and Peter L. Jackson (2021). "Gridded bias correction of modeled PM2.5 for exposure assessment, and estimation of background concentrations over a coastal valley region of northwestern British Columbia, Canada". In: *Journal of the Air & Waste Management Association* 71.2, pp. 156–169. DOI: 10.1080/10962247.2020.1844342.

Orru, Hans et al. (2013). "Impact of climate change on ozone-related mortality and morbidity in Europe". In: *European Respiratory Journal* 41.2, pp. 285–294. DOI: 10.1183/09031936.00210411.

Ortiz, Edison Y. et al. (2021). "Combination of GEOS-CF model with Machine Learning as a tool for forecasting regional pollution in Bogotá". In: *2021 Congreso Colombiano y Conferencia Internacional de Calidad de Aire y Salud Pública (CASAP)*. 2021 Congreso Colombiano y Conferencia Internacional de Calidad de Aire y Salud Pública (CASAP), pp. 1–4. DOI: 10.1109/CASAP54985.2021.9703381.

Otero, N et al. (2016). "Synoptic and meteorological drivers of extreme ozone concentrations over Europe". In: *Environmental Research Letters* 11.2, p. 024005. DOI: 10.1088/1748-9326/11/2/024005.

Otero, Noelia, Henning W. Rust, and Tim Butler (2021). "Temperature dependence of tropospheric ozone under NOx reductions over Germany". In: *Atmospheric Environment* 253, p. 118334. DOI: 10.1016/j.atmosenv.2021.118334.

Otero, Noelia, Jana Sillmann, et al. (2018). "A multi-model comparison of meteorological drivers of surface ozone over Europe". In: *Atmospheric Chemistry and Physics* 18.16, pp. 12269–12288. DOI: 10.5194/acp-18-12269-2018.

Pannullo, Francesca et al. (2017). "Quantifying the impact of current and future concentrations of air pollutants on respiratory disease risk in England". In: *Environmental Health* 16.1, p. 29. DOI: 10.1186/s12940-017-0237-1.

Phalitnonkiat, Pakawat et al. (2016). "Extreme ozone events: Tail behavior of the surface ozone distribution over the U.S." In: *Atmospheric Environment* 128, pp. 134–146. DOI: 10.1016/j.atmosenv.2015.12.047.

Pickands, James (1975). "Statistical Inference Using Extreme Order Statistics". In: *Annals of Statistics* 3.1, pp. 119–131. DOI: 10.1214/aos/1176343003.

Ponomarev, Nikolai, Vladislav Yushkov, and Nikolai Elansky (2021). "Air Pollution in Moscow Megacity: Data Fusion of the Chemical Transport Model and Observational Network". In: *Atmosphere* 12.3, p. 374. DOI: 10.3390/atmos12030374.

Pope, R J et al. (2016). "The impact of synoptic weather on UK surface ozone and implications for premature mortality". In: *Environmental Research Letters* 11.12, p. 124004. DOI: 10.1088/1748-9326/11/12/124004.

Porter, William C. and Colette L. Heald (2019). "The mechanisms and meteorological drivers of the summertime ozone–temperature relationship". In: *Atmospheric Chemistry and Physics* 19.21. Publisher: Copernicus GmbH, pp. 13367–13381. ISSN: 1680-7316. DOI: 10.5194/acp-19-13367-2019.

Powers, Jordan G. et al. (2017). "The Weather Research and Forecasting Model: Overview, System Efforts, and Future Directions". In: *Bulletin of the American Meteorological Society* 98.8, pp. 1717–1737. DOI: 10.1175/BAMS-D-15-00308.1. URL: https://journals.ametsoc.org/view/journals/bams/98/8/bams-d-15-00308.1.xml.

Ramli, Nor Azam, Nurul Adyani Ghazali, and Ahmad Shukri Yahaya (2010). "Diurnal Fluctuations of Ozone Concentrations and its Precursors and Prediction of Ozone Using Multiple Linear Regressions". In: *Malaysian Journal of Environmental Management*. URL: http://journalarticle.ukm.my/2359/1/MJEM_2010_2__5_Nor_Azam_Ramli__F_.pdf.

Reddy, K. K. et al. (2012). "Influences of the boundary layer evolution on surface ozone variations at a tropical rural site in India". In: *Journal of Earth System Science* 121.4, pp. 911–922. DOI: 10.1007/s12040-012-0200-z.

Reich, Brian J., Howard H. Chang, and Kristen M. Foley (2014). "A spectral method for spatial downscaling". In: *Biometrics* 70.4, pp. 932–942. DOI: 10.1111/biom.12196.

Reich, Brian J., Montserrat Fuentes, and David B. Dunson (2011). "Bayesian Spatial Quantile Regression". In: *Journal of the American Statistical Association* 106.493, pp. 6–20. DOI: 10.1198/jasa.2010.ap09237.

Ren, Xiang et al. (2022). "Flexible Bayesian Ensemble Machine Learning Framework for Predicting Local Ozone Concentrations". In: *Environmental Science & Technology* 56.7, pp. 3871–3883. DOI: 10.1021/acs.est.1c04076.

Reyes, Jeanette and Marc Serre (2014). "Combining Observed PM2.5 with Regionalized Bias- Corrected Chemical Transport Models". In: *ISEE Conference Abstracts*. DOI: 10.1289/isee.2014.O-175.

Ribeiro, Marco Tulio, Sameer Singh, and Carlos Guestrin (2016). ""Why Should I Trust You?": Explaining the Predictions of Any Classifier". In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '16. New York, NY, USA, pp. 1135–1144. DOI: 10.1145/2939672.2939778.

Rieder, H E et al. (2013). "Changes in the frequency and return level of high ozone pollution events over the eastern United States following emission controls". In: *Environmental Research Letters* 8.1, p. 014012. DOI: 10.1088/1748-9326/8/1/014012.

Robeson, S. M. and D. G. Steyn (1990). "Evaluation and comparison of statistical forecast models for daily maximum ozone concentrations". In: *Atmospheric Environment. Part B. Urban Atmosphere* 24.2, pp. 303–312. DOI: 10.1016/0957-1272(90)90036-T.

Romer, Paul S. et al. (2018). "Effects of temperature-dependent NO_x emissions on continental ozone production". In: *Atmospheric Chemistry and Physics* 18.4, pp. 2601–2614. DOI: 10.5194/acp-18-2601-2018.

Romero-Alvarez, Johana et al. (2022). "Sources of surface O3 in the UK: tagging O3 within WRF-Chem". In: *Atmospheric Chemistry and Physics* 22.20, pp. 13797–13815. DOI: 10.5194/acp-22-13797-2022.

Sandermann Jr, Heinrich (1996). "Ozone and Plant Health". In: *Annual Review of Phytopathology* 34.1. _eprint: https://doi.org/10.1146/annurev.phyto.34.1.347, pp. 347–366. DOI: 10.1146/annurev.phyto.34.1.347.

Savage, N. H. et al. (2013). "Air quality modelling using the Met Office Unified Model (AQUM OS24-26): model description and initial evaluation". In: *Geoscientific Model Development* 6.2, pp. 353–372. DOI: 10.5194/gmd-6-353-2013.

Saygın, Hasan, Özge Eren, and Hasan Volkan Oral (2018). "Peaks Over Threshold Method Application on Airborne Particulate Matter (PM10) and Sulphur Dioxide (SO2) Pollution Detection in Specified Regions of İstanbul". In: *European Journal of Science and Technology*. URL: https://dergipark.org.tr/en/pub/ejosat/issue/40225/420317.

Scheel, H E et al. (1997). "On the Spatial Distribution and Seasonal Variation of Lower-Troposphere Ozone over Europe". In: *Journal of Atmospheric Chemistry*, p. 18. URL: https://link.springer.com/article/10.1023/A:1005882922435.

Shen, L., L. J. Mickley, and E. Gilleland (2016). "Impact of increasing heat waves on U.S. ozone episodes in the 2050s: Results from a multimodel analysis using extreme value theory". In: *Geophysical Research Letters* 43.8, pp. 4017–4025. DOI: 10.1002/2016GL068432.

Silibello, Camillo et al. (2021). "Spatial-temporal prediction of ambient nitrogen dioxide and ozone levels over Italy using a Random Forest model for population exposure assessment". In: *Air Quality, Atmosphere & Health* 14.6, pp. 817–829. DOI: 10.1007/s11869-021-00981-4.

Simon, Heather et al. (2015). "Ozone Trends Across the United States over a Period of Decreasing NOx and VOC Emissions". In: *Environmental Science & Technology* 49.1, pp. 186–195. DOI: 10.1021/es504514z.

Simpson, D. et al. (2012). "The EMEP MSC-W chemical transport model – technical description". In: *Atmospheric Chemistry and Physics* 12.16, pp. 7825–7865. DOI: 10.5194/acp-12-7825-2012.

Skamarock, C. et al. (2008). "A Description of the Advanced Research WRF Version 3". In: DOI: 10.5065/D68S4MVH. URL: https://opensky.ucar.edu/islandora/object/technotes%3A500/.

Skamarock, William C et al. (2019). "A Description of the Advanced Research WRF Model Version 4". In.

Song, Ge et al. (2022). "Surface UV-assisted retrieval of spatially continuous surface ozone with high spatial transferability". In: *Remote Sensing of Environment* 274, p. 112996. DOI: 10.1016/j.rse.2022.112996.

Spiridonov, Vlado et al. (2019). "Development of air quality forecasting system in Macedonia, based on WRF-Chem model". In: *Air Quality, Atmosphere & Health* 12.7, pp. 825–836. DOI: 10.1007/s11869-019-00698-5.

Srebot, Vera et al. (2009). "Ozone and cardiovascular injury". In: *Cardiovascular Ultrasound* 7.1, p. 30. DOI: 10.1186/1476-7120-7-30.

SRUC (2017). *Climate change: evidence review in Agriculture, Forestry, Land Use, Waste*. URL: http://www.gov.scot/publications/evidence-review-potential-wider-impacts-climate-change-mitigation-options-agriculture/.

Stähle, Christoph (2019). "Future changes in surface ozone air quality in Europe: insights from chemistry-climate model simulations". PhD thesis. Karl-Franzens-Universität Graz. URL: http://unipub.uni-graz.at/obvugrhs/3657221.

Stevenson, Ken et al. (2009). *QA/QC Procedures for the UK Automatic Urban and Rural Air Quality Monitoring Network (AURN) QA QC Manual*. AEA. URL: https://uk-air.defra.gov.uk/assets/documents/reports/cat13/0910081142_AURN_QA_QC_Manual_Sep_09_FINAL.pdf.

Stortini, Michele, Barbara Arvani, and Marco Deserti (2020). "Operational Forecast and Daily Assessment of the Air Quality in Italy: A Copernicus-CAMS Downstream Service". In: *Atmosphere* 11.5, p. 447. DOI: 10.3390/atmos11050447.

Su, Tianning, Zhanqing Li, and Ralph Kahn (2018). "Relationships between the planetary boundary layer height and surface pollutants derived from lidar observations over China: regional pattern and influencing factors". In: *Atmospheric Chemistry and Physics* 18.21. Publisher: Copernicus GmbH, pp. 15921–15935. DOI: 10.5194/acp-18-15921-2018.

Sun, Haitong et al. (2022). "Spatial Resolved Surface Ozone with Urban and Rural Differentiation during 1990–2019: A Space–Time Bayesian Neural Network Downscaler". In: *Environmental Science & Technology* 56.11. Publisher: American Chemical Society, pp. 7337–7349. DOI: 10.1021/acs.est.1c04797.

Sun, Jian et al. (2015). "Estimation of future PM2.5- and ozone-related mortality over the continental United States in a changing climate: An application of high-resolution dynamical downscaling technique". In: *Journal of the Air & Waste Management Association* 65.5, pp. 611–623. DOI: 10.1080/10962247.2015.1033068.

Sun, Wenxiu, Peter Hess, and Chengji Liu (2017). "The impact of meteorological persistence on the distribution and extremes of ozone". In: *Geophysical Research Letters* 44.3, pp. 1545–1553. DOI: 10.1002/2016GL071731.

Tan, Zhaofeng et al. (2018). "Explicit diagnosis of the local ozone production rate and the ozone-NOx-VOC sensitivities". In: *Science Bulletin* 63.16, pp. 1067–1076. DOI: 10.1016/j.scib.2018.07.001.

Tobías, Aurelio and Manuel G. Scotto (2005). "Prediction of extreme ozone levels in Barcelona, Spain". In: *Environmental Monitoring and Assessment* 100.1, pp. 23–32. DOI: 10.1007/s10661-005-7057-4.

Torgo, Luís et al. (2015). "Resampling strategies for regression". In: *Expert Systems* 32.3, pp. 465–476. DOI: 10.1111/exsy.12081.

Towler, Erin et al. (2020). "Extreme-value analysis for the characterization of extremes in water resources: A generalized workflow and case study on New Mexico monsoon precipitation". In: *Weather and Climate Extremes* 29, p. 100260. DOI: 10.1016/j.wace.2020.100260.

Trail, M. et al. (2013). "Downscaling a Global Climate Model to Simulate Climate Change Impacts on U.S. Regional and Urban Air Quality". In: *Geoscientific Model Development DIscussions* 6. NTRS Author Affiliations: Georgia Inst. of Tech., Columbia Univ. NTRS Report/Patent Number: GSFC-E-DAA-TN8471 NTRS Document ID: 20140011350 NTRS Research Center: Goddard Space Flight Center (GSFC). DOI: 10.5194/gmdd-6-2517-2013.

Tudor, Cristiana (2022). "Ozone pollution in London and Edinburgh: spatiotemporal characteristics, trends, transport and the impact of COVID-19 control measures". In: *Heliyon* 8.11. Publisher: Elsevier. DOI: 10.1016/j.heliyon.2022.e11384.

Van Dingenen, Rita et al. (2009). "The global impact of ozone on agricultural crop yields under current and future air quality legislation". In: *Atmospheric Environment* 43.3, pp. 604–618. DOI: 10.1016/j.atmosenv.2008.10.033.

Vardoulakis, Sotiris and Clare Heaviside (2012). "Health Effects of Climate Change in the UK 2012". In: *London: Health Protection Agency*, p. 242. URL: https://www.p-plus.nl/resources/articlefiles/HealthEffectsofClimateChangeintheUK20123.pdf.

Varotsos, Konstantinos V., Christos Giannakopoulos, and Maria Tombrou (2019). "Ozone-temperature relationship during the 2003 and 2014 heatwaves in Europe". In: *Regional Environmental Change* 19.6, pp. 1653–1665. DOI: 10.1007/s10113-019-01498-4.

Velthoen, Jasper et al. (2022). "Gradient boosting for extreme quantile regression". In: DOI: https://doi.org/10.48550/arXiv.2103.00808.

Vieno, M., A. J. Dore, et al. (2010). "Modelling surface ozone during the 2003 heat-wave in the UK". In: *Atmospheric Chemistry and Physics* 10.16, pp. 7963–7978. DOI: 10.5194/acp-10-7963-2010.

Vieno, M., M. R. Heal, et al. (2016). "The sensitivities of emissions reductions for the mitigation of UK PM2.5". In: *Atmospheric Chemistry and Physics* 16.1, pp. 265–276. DOI: 10.5194/acp-16-265-2016.

Wang, HaiLin et al. (2013). "Characterization and assessment of volatile organic compounds (VOCs) emissions from typical industries". In: *Chinese Science Bulletin* 58.7, pp. 724–730. DOI: 10.1007/s11434-012-5345-2.

Wang, Yuan, Qiangqiang Yuan, Tongwen Li, et al. (2021). "Estimating daily full-coverage near surface O3, CO, and NO2 concentrations at a high spatial resolution over China based on S5P-TROPOMI and GEOS-FP". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 175, pp. 311–325. DOI: 10.1016/j.isprsjprs.2021.03.018.

Wang, Yuan, Qiangqiang Yuan, Liye Zhu, et al. (2022). "Spatiotemporal estimation of hourly 2-km ground-level ozone over China based on Himawari-8 using a self-adaptive geospatially local model". In: *Geoscience Frontiers* 13.1, p. 101286. DOI: 10.1016/j.gsf.2021.101286.

Watson, Gregory L. et al. (2019). "Machine learning models accurately predict ozone exposure during wildfire events". In: *Environmental Pollution* 254, p. 112792. DOI: 10.1016/j.envpol.2019.06.088.

Wilczak, James M. et al. (2009). "Analysis of regional meteorology and surface ozone during the TexAQS II field program and an evaluation of the NMM-CMAQ and WRF-Chem air quality models". In: *Journal of Geophysical Research: Atmospheres* 114 (D7). DOI: 10.1029/2008JD011675.

Wilson, Bradley et al. (2022). "Characterizing changes in extreme ozone levels under 2050s climate conditions: An extreme-value analysis in California". In: *Atmospheric Environment: X* 16, p. 100195. DOI: 10.1016/j.aeaoa.2022.100195.

Wong, David W., Lester Yuan, and Susan A. Perlin (2004). "Comparison of spatial interpolation methods for the estimation of air quality data". In: *Journal of Exposure Science & Environmental Epidemiology* 14.5, pp. 404–415. DOI: 10.1038/sj.jea.7500338.

Yin, Hao et al. (2021). "Unprecedented decline in summertime surface ozone over eastern China in 2020 comparably attributable to anthropogenic emission reductions and meteorology". In: *Environmental Research Letters* 16.12, p. 124069. DOI: 10.1088/1748-9326/ac3e22.

Zhang, Junfeng, Yongjie Wei, and Zhangfu Fang (2019). "Ozone Pollution: A Major Health Hazard Worldwide". In: *Frontiers in Immunology* 10. URL: https://www.frontiersin.org/article/10.3389/fimmu.2019.02518 (visited on 05/24/2022).

Zhang, Yuzhong and Yuhang Wang (2016). "Climate-driven ground-level ozone extreme in the fall over the Southeast United States". In: *Proceedings of the National Academy of Sciences* 113.36, pp. 10025–10030. DOI: 10.1073/pnas.1602563113.

Zhao, Yusheng et al. (2022). "Separately resolving NOx and VOC contributions to ozone formation". In: *Atmospheric Environment* 285, p. 119224. DOI: 10.1016/j.atmosenv.2022.119224.

Zheng, Huang et al. (2023). "Enhanced ozone pollution in the summer of 2022 in China: The roles of meteorology and emission variations". In: *Atmospheric Environment* 301, p. 119701. DOI: 10.1016/j.atmosenv.2023.119701.

Zidek, James V., Weimin Sun, and Nhu D. Le (2000). "Designing and Integrating Composite Networks for Monitoring Multivariate Gaussian Pollution Fields". In: *Journal of the Royal Statistical Society Series C: Applied Statistics* 49.1, pp. 63–79. DOI: 10.1111/1467-9876.00179.