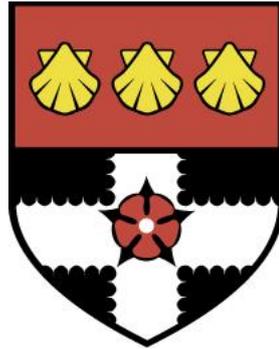


UNIVERSITY OF READING

Department of Meteorology



# Identifying Meteorological Inaccuracies in Air Quality Forecasts

KAJA MILCZEWSKA

*A thesis submitted for the degree of Doctor of Philosophy*

September 2021



# Declaration

I confirm that this is my own work and the use of all material from other sources has been properly and fully acknowledged.

Kaja Milczewska

September 2021



# Abstract

Air quality (AQ) forecasts are important for helping the public mitigate adverse health effects associated with episodes of high pollution. In order to create an accurate AQ forecast, it is important to correctly represent meteorological processes due to their strong influence on pollutant concentrations.

The aim of this work is to determine the relationship between forecast errors in surface pollutant concentrations and meteorological forecast errors within the operational AQ model, Air Quality in the Unified Model (AQUM). This thesis explores three different approaches of evaluation to determine the impact of meteorological forecast errors in 10 m wind speed, 1.5 m temperature and precipitation on pollutant errors ( $O_3$ ,  $NO_2$ ,  $PM_{10}$  and  $PM_{2.5}$ ). These are point-based, neighbourhood and process-based methods.

Point-based metrics evaluate forecasts against observations, paired in space and time. The evaluation reveals negative forecast bias in diurnal cycles of summertime  $NO_2$  and positive bias in  $O_3$ , with a 2-hour lag in timing of the forecast increase of morning concentrations. It is shown that night-time 10 m wind speed over-estimation coincides with the largest  $O_3$  and  $NO_2$  biases. Point-based evaluation identifies a negative bias in PM concentrations which decreases by 10% to 25% after under-estimating precipitation.

Neighbourhood evaluation relaxes the spatial constraint for forecast - observation pairs. It is a novel mechanism of attributing forecast errors in AQ to meteorology. Strongest positive correlations between night-time  $O_3$  and wind speed forecast errors are shown to occur at a neighbourhood of 100 km<sup>2</sup>. Forecast error anti-correlations between  $NO_2$  and wind speed reach a maximum at a smaller neighbourhood.

Finally, process-based evaluation is used to test whether statistical relationships between  $O_3$ ,  $NO_2$  and 10 m wind speed forecast errors are caused by the physical process of entrainment. To quantify the influence of entrainment in the model on forecast total oxidant ( $O_3 + NO_2 = O_x$ ) concentrations, an experiment using the off-line NAME model simulates tracer dispersion under different meteorology configurations. The experiment confirms that lag in the forecast morning  $O_x$  increase is due to delayed boundary layer development. In light of the results, it is recommended that model developers implement a land-surface parametrisation with better urban heat storage to improve modelled surface heat fluxes, nocturnal boundary layer stability and nocturnal winds. Improving these may result in more accurate surface AQ forecasts.



# Acknowledgements

Firstly, special thanks to my main supervisor, Prof. Helen Dacre for her support, encouragement, wisdom and buckets of patience during my PhD. I promise to never write a sentence which spills over five lines ever again. To my external supervisors from the UK Met Office: Dr. Paul Agnew, Lucy Neal and Dr. Marion Mittermaier, who provided the expertise, resources and encouragement I needed to complete this thesis.

Thanks to my external and internal examiners, Prof. Oliver Wild and Prof. Sue Gray, for an engaging viva examination which put me through my paces. Extended thanks to my Monitoring Committee members: Prof. Bryan Lawrence and Prof. Giles Harrison for useful discussions and helping me keep the project on track. Many thanks to the Mesoscale Research group, whose members patiently put up with (probably) the worst of my presentations and showed interest in my work throughout. A special shout-out to Wendy Neale, an absolute Zoom rockstar in the midst of the gloomy pandemic.

Life at Reading Meteorology department would not be the same without the stellar post-docs and PhD students. Thanks to Hannah, Caroline and Natalie for keeping my mental health in check and just being generally lovely. Colleagues who turned into friends: Dom, Beth, Aga, Michael, H el ene, Bethan and Josh. Ex-Reading cheerleaders from afar, Flavia and Rachel whose support was felt from 6000 miles away: obrigada! Friends who kept my head up when life turned sour: Sally and Rachael, you are real-life heroes.

My family - Mama, Agee, Mick and Julie and the rest! - who kept on believing in me and being so proud even when I doubted my purpose. Nikki, Geoff, Lynn and Sam for providing all the necessary vegan chocolate and encouragement to get me through.

Finally, thank you Dan for keeping me stable through this turbulent time, for resolving my greyzone moods and for providing all the baked goods. Thank you for your unwavering support, encouragement, love and understanding. You never doubted me once, and I could not have wished for a better triple-lockdown buddy during the Covid-19 pandemic. Can we get a puppy now, please?

This was a NERC funded PhD project, with a CASE award from the UK Met Office and additional support from SCENARIO DTP. To this day I cannot fathom that I was chosen to complete this project, and I am forever grateful for the opportunity.



# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Motivation . . . . .	1
1.2	Forecasting air quality and meteorology: a background . . . . .	2
1.2.1	Observations: ground-based and remote . . . . .	6
1.2.2	AQ forecasting in the United Kingdom . . . . .	8
1.3	Forecast model verification . . . . .	9
1.4	Thesis aims and structure . . . . .	11
<b>2</b>	<b>Background</b>	<b>13</b>
2.1	Atmospheric trace constituents . . . . .	13
2.1.1	Particulate Matter . . . . .	13
2.1.2	Tropospheric O <sub>3</sub> and NO <sub>2</sub> . . . . .	14
2.2	Atmospheric boundary layer meteorology . . . . .	19
2.2.1	Turbulence in the atmosphere . . . . .	21
2.3	Physical processes and meteorological variables affecting atmospheric species	27
2.3.1	Surface winds . . . . .	27
2.3.2	Dry deposition . . . . .	28
2.3.3	Wet deposition . . . . .	29
2.3.4	Cloud cover . . . . .	31
2.3.5	Summary . . . . .	31
<b>3</b>	<b>Point-based evaluation of AQ and meteorological variables</b>	<b>33</b>
3.1	Introduction . . . . .	33
3.1.1	Research questions . . . . .	35
3.2	Observations . . . . .	37
3.2.1	Automatic Urban and Rural Network (AURN) . . . . .	38
3.2.2	WMO synoptic sites and radar . . . . .	39
3.2.3	Typical errors in observations . . . . .	40
3.3	The Met Office Unified Model . . . . .	41
3.4	Air Quality configuration of the Unified Model (AQUM) . . . . .	42
3.5	Forecast verification of air quality variables . . . . .	43
3.5.1	Diurnal cycles of O <sub>3</sub> and NO <sub>2</sub> . . . . .	44
3.5.2	Summary of O <sub>3</sub> , NO <sub>2</sub> and O <sub>x</sub> evaluation . . . . .	50
3.5.3	Diurnal cycles of PM <sub>2.5</sub> and PM <sub>10</sub> . . . . .	51
3.6	Forecast verification of meteorological variables . . . . .	52

3.6.1	10 m wind speed . . . . .	53
3.6.2	1.5m temperature . . . . .	55
3.6.3	Precipitation . . . . .	56
3.7	Relationships between air quality and meteorological forecast errors . . .	66
3.7.1	10 m wind vs O <sub>3</sub> and NO <sub>2</sub> . . . . .	66
3.7.2	Precipitation vs PM <sub>2.5</sub> and PM <sub>10</sub> . . . . .	69
3.8	Summary, discussion and conclusions . . . . .	75
3.8.1	O <sub>3</sub> , NO <sub>2</sub> and 10 m wind speed . . . . .	76
3.8.2	PM and precipitation . . . . .	77
<b>4</b>	<b>Neighbourhood forecast verification</b>	<b>81</b>
4.1	Introduction . . . . .	81
4.1.1	Chapter aims . . . . .	83
4.1.2	Research questions . . . . .	84
4.2	High Resolution Assessment (HiRA) . . . . .	85
4.3	Probabilistic Verification Scores . . . . .	86
4.3.1	Brier Score . . . . .	86
4.3.2	Continuous Ranked Probability Score . . . . .	88
4.4	Using spatial forecast verification with non-gridded observations . . . . .	89
4.5	Relationship between total oxidant and 10 m wind speed forecast errors .	92
4.5.1	Spatial correlations of O <sub>3</sub> and NO <sub>2</sub> observations . . . . .	93
4.5.2	Neighbourhood verification of surface wind speed . . . . .	95
4.5.3	Forecast error correlations . . . . .	97
4.5.4	Summary and conclusions for this section . . . . .	103
4.6	Relationship between PM and precipitation forecast errors . . . . .	104
4.6.1	Data and methodology . . . . .	104
4.6.2	Relationships between forecasts and observations . . . . .	106
4.6.3	Comparison of errors in PM <sub>2.5</sub> and precipitation . . . . .	107
4.6.4	Precipitation BS and its relationship with PM . . . . .	109
4.6.5	Conclusions: neighbourhood evaluation of relationships in precipi- tation and PM errors . . . . .	114
4.7	Chapter summary and conclusions . . . . .	115
<b>5</b>	<b>Process-based evaluation of boundary layer development and forecast errors in O<sub>3</sub> and NO<sub>2</sub></b>	<b>117</b>
5.1	Introduction . . . . .	117
5.2	Chapter aims . . . . .	120
5.3	Model configurations and key parametrisation schemes . . . . .	122
5.3.1	NAME . . . . .	122
5.3.2	UKV . . . . .	124

5.3.3	Global configuration of the Unified Model . . . . .	125
5.3.4	The Lock boundary layer parametrisation scheme . . . . .	125
5.3.5	Urban land tile schemes: Best and MORUSES . . . . .	128
5.4	Evaluation of physical parameters . . . . .	130
5.4.1	Meteorological conditions . . . . .	131
5.4.2	Boundary layer depth . . . . .	133
5.4.3	Vertical velocity . . . . .	136
5.4.4	Surface sensible heat fluxes . . . . .	138
5.5	Methodology for idealised NAME experiments . . . . .	139
5.5.1	NWP input . . . . .	140
5.5.2	Tracer sources . . . . .	140
5.5.3	Tagging: tracking movement of tracers . . . . .	144
5.5.4	Tracer sinks . . . . .	144
5.5.5	Boundary layer schemes in experiments . . . . .	145
5.6	Results of NAME tracer experiments . . . . .	145
5.6.1	Diurnal variability . . . . .	146
5.6.2	Comparison with Eulerian budget study . . . . .	152
5.6.3	Synoptic variability: diurnal cycles . . . . .	158
5.6.4	Synoptic variability: case study of one week . . . . .	161
5.6.5	Inhomogeneous boundary layer simulations . . . . .	162
5.7	Summary of results, conclusions and discussion . . . . .	163
5.7.1	Summary of results . . . . .	164
5.7.2	Chapter conclusions and discussion . . . . .	165
<b>6</b>	<b>Conclusions</b> . . . . .	<b>169</b>
6.1	Impact of 10 m wind speed errors on O <sub>3</sub> and NO <sub>2</sub> forecast errors . . . . .	170
6.2	Impact of entrainment representation on O <sub>3</sub> and NO <sub>2</sub> forecast errors . . . . .	173
6.3	Impact of precipitation errors on PM <sub>2.5</sub> and PM <sub>10</sub> . . . . .	175
6.4	Summary . . . . .	177
6.5	Further questions and discussion . . . . .	178
	<b>References</b> . . . . .	<b>183</b>



# 1. Introduction

## 1.1 Motivation

Exposure to pollutants in the atmosphere can pose a risk to human health in the form of respiratory and cardiovascular diseases. During prolonged periods of elevated concentration levels, pollutants such as ozone ( $O_3$ ) and particulate matter (PM) can trigger existing respiratory problems such as asthma, reduce lung function and cause lung disease. For example, nitrogen dioxide ( $NO_2$ ) has been found to be associated with symptoms of bronchitis in children already suffering from asthma ([World Health Organization, 2020](#)), while episodes of high  $O_3$  concentrations can trigger pre-existing respiratory and cardiovascular problems, resulting in increased hospital admissions ([World Health Organization, 2006](#); [COMEAP, 2015](#)). As of 2016, an estimated 91% of the world's population lives within regions where ambient AQ standards fall short of the safe exposure limit guidelines, proposed by the World Health Organisation (WHO) ([World Health Organization, 2020](#)). These guidelines are in place for five most harmful pollutants:  $O_3$ ,  $PM_{2.5}$  and  $PM_{10}$ ,  $NO_2$ , carbon monoxide (CO) and sulphur dioxide ( $SO_2$ ).

The problem of pollution is as much an issue worldwide as it is an issue in the United Kingdom. In the UK alone, an estimated 50,000 premature deaths due to poor air quality occur annually [House of Commons Environmental Audit Committee \(HCEA\) \(2010\)](#). Approximately 85% of UK towns had measured levels of  $PM_{2.5}$  (i.e. particles with aerodynamic diameter  $\leq 2.5 \mu m$ ) above the safe annual mean threshold of  $10 \mu g m^{-3}$ , which is a major contributing factor to lowering the residents' life expectancy by an average of 8.6 months ([World Health Organization, 2020](#)).

An accurate forecast of short-term (i.e. hourly, daily) surface pollutant concentrations can inform the public and enable them to try mitigating some of the adverse health risks associated with exposure to elevated concentrations. Producing an accurate pollutant forecast involves careful representation of many different physical and chemical processes. Surface concentrations of the five aforementioned pollutants are variable on

regional scales, as their sources and sinks have anthropogenic (e.g. emissions from vehicles, factories, etc.) and natural (e.g. volcano ash, desert dust) components of temporal and spatial variability. Sources of error within an air quality (AQ) forecast may come from chemical transformations of atmospheric species; or they may be a result of inaccuracies in emissions inventories used within the model; or the errors may be a consequence of errors in modelling the meteorology. It can therefore be a challenge for an AQ model to fully capture all the mechanisms which influence forecast pollutant concentrations. The aim of this thesis is to explore how the accuracy of meteorology in numerical weather prediction (NWP) models affects the forecast errors of surface pollutant concentrations on regional scales in the United Kingdom.

## 1.2 Forecasting air quality and meteorology: a background

Historically, the AQ forecasting and numerical weather prediction (NWP) communities have engaged in little conversation about the possible feedback effects of one model type to another, which has led to separation in respective model development (Zhang, 2008; Brunner et al., 2015). Off-line AQ models rely on meteorological input from a separate meteorological model. However, pollutants and aerosols in the atmosphere can also affect the meteorology, e.g. through acting as cloud condensation nuclei, which, if represented incorrectly, could translate to cloud cover errors in the meteorology forecast and subsequent reduction of downward solar radiation (Bangert et al., 2011). An on-line modelling framework provides this two-way feedback between the AQ and meteorology. It is only within the last 20 years that the AQ and NWP communities have been working together, partly in light of increasing computational power, to develop on-line AQ configurations which support feedback effects (Baklanov et al., 2014). These processes are relevant on the short time-scales of NWP as well as time-scales relevant to climate models and as such, their representation is important to not only the AQ and NWP communities but also for climate prediction. On-line modelling frameworks can therefore support a shorter model time-step, e.g. 5 minutes instead of 1 - 3 hours in an off-line model. As a result of transitioning from off-line to on-line modelling, regional AQ forecasts have experienced vast improvements in accuracy over the past decade (Brunner et al., 2015).

Operational, regional AQ forecasts are generally useful at lead times of anywhere between 24 - 144 hours ahead, where forecast performance tends to degrade with lead time (Honoré et al., 2008). The accuracy of a coupled chemistry-meteorology AQ model to predict pollutants in the troposphere is partly dependent on the accuracy of the me-

teorology fields which influence pollutant dispersion, in addition to their influence on the chemistry of reactive species (Zhang et al., 2013). For instance, a persistent O<sub>3</sub> bias is a challenge faced by many regional and global AQ models (Brown-Steiner et al., 2015; Im et al., 2015b; Travis et al., 2016), and surface O<sub>3</sub> concentrations are sensitive to a number of factors rooted in chemistry and meteorology. Small-scale meteorological features (e.g. local cloudiness, surface heat fluxes and urban heat island effect; Zhang et al., 2012; Eder et al., 2006; Hogrefe et al., 2007) and circulations (e.g. land-sea breezes; Hess et al., 2004) are all important for accurate calculation of pollutant concentrations. A realistic representation of the loss of species from the atmosphere is also important for accurately forecasting their concentrations: namely wet deposition due to rainfall, dry deposition to surfaces and ventilation. Furthermore, horizontal and vertical resolution of the AQ model can result in O<sub>3</sub> bias - for example, Brown-Steiner et al. (2015) demonstrate that global climate - chemistry models have a greater O<sub>3</sub> bias over the eastern US when 26 vertical model levels instead of 56. Higher vertical and horizontal resolution, however, comes at a greater computational cost. On-line regional AQ models thus need to capture all the relevant physical and chemical processes within as high degree of accuracy as possible, while minimising the computational cost of running the model.

Various model-intercomparison studies within the past decade have revealed a large spread in the accuracy of pollution forecasts in Europe and North America, particularly for O<sub>3</sub>. For example, Baklanov et al. (2014) outline various models' performances and the feedback effects between specific meteorological and AQ parameters. The survey of 30 European AQ models reveals that 10-m wind speed, precipitation, temperature and short-wave radiation are some of the most important meteorological factors on the modelled atmospheric composition. Within the Air Quality Model Evaluation International Initiative (AQMEII) framework, Vautard et al. (2012) evaluate a large number of mesoscale meteorological models used to drive off-line AQ simulations and report significant variability in key meteorological parameters for AQ forecasting such as shortwave radiation and a systematic overestimation in wind speeds during stable conditions. Brunner et al. (2015) have expanded this evaluation to on-line coupled chemistry-meteorology models and have found that despite the improvements associated with on-line AQ modelling, wind speed biases continue to persist, in addition to a large model spread in planetary boundary layer (PBL) depths and solar incoming radiation. This could indicate that forecasting wind speed requires more attention within many AQ models.

Since it is not possible to explicitly resolve every chemical and meteorological process

in a forecast model, parametrisation schemes are used to represent various complex and sub-grid scale processes. However, the choice of a physical parametrisation scheme could potentially introduce errors into parameters such as horizontal and vertical wind-speed, due to the scheme's use across different models and configurations, each of which requires fine-tuning and some aspects of a parametrisation scheme may not be compatible with another scheme. It could also introduce errors within modelling of AQ variables: for instance, the choice of boundary layer parametrisation will have an effect on the vertical and horizontal transport of  $O_3$ , as highlighted by [Im et al. \(2015b\)](#), since the boundary layer is where most vertical mixing occurs. Meanwhile, forecast errors pertaining to the choice of cloud scheme have been shown by both [Eder et al. \(2006\)](#) and [Ryu et al. \(2018\)](#) to influence the amount of short-wave radiation at the Earth's surface and the amount of  $NO_2$  photolysis, which impacts  $O_3$  concentrations.

Various other studies have explored the influence of forecast errors in meteorology on AQ and  $O_3$  in particular. [McNider and Pour-Biazar \(2020\)](#) provide a thorough review of recent studies on relationships between AQ and meteorology forecasts in the United States, generally agreeing that modelled wind speed, surface temperature and night-time boundary layer stability are among the most influential meteorological variables to AQ forecasting. [Im et al. \(2015b\)](#) conducted an inter-model comparison on the meteorological impacts on  $O_3$  forecast accuracy of regional European AQ models. They suggest that the vast majority of regional models' inability to correctly diagnose boundary layer height and stability is reflected in the persistent large  $O_3$  biases. It is one of the first studies of its kind, where the rigorous evaluation of regional meteorological forecasts over Europe and United States reveals a systematic winter-time underprediction of  $O_3$ . Another study by [Hanna and Yang \(2001\)](#) evaluated 4 mesoscale meteorological models used within AQ forecasting and confirmed that night-time boundary layer stability is often under-estimated, which can lead to enhanced vertical dispersion of pollutants and too much mixing.

[Savage et al. \(2013\)](#) suggests that since the night-time boundary layer in mid-latitude regions is typically stable, misclassification of its stability in regional models may lead to miscalculation of the overnight concentrations of  $O_3$ . This could affect the concentrations of gaseous pollutant species within the entrained air masses into the boundary layer from aloft, and suggests that due to vertical mixing throughout the boundary layer, entrainment of polluted air masses may influence surface pollutant concentrations.

Surface wind speed is another variable which has systematic bias in models. The widely used Weather Research and Forecasting (WRF) model has been shown to over-

estimate night-time wind speed in both coastal and in-land regions of the United States, as well as coastal parts of Spain. As a result of this, AQ configurations associated with WRF (e.g. WRF-Chem) can experience too much mixing of  $O_3$  into the boundary layer from aloft (Ngan et al., 2013; Lee et al., 2014). Ngan et al. (2012) demonstrated that  $O_3$  biases near metropolitan areas present in Community Multiscale Air Quality (CMAQ) forecasts are minimised when the model is re-configured with improved wind speed fields. Over-estimating night-time winds speed has also been found to be a common problem within European AQ models (Brunner et al., 2015), alongside the night-time stability misclassification.

On the other hand, there exist studies which demonstrate that meteorology is not the main source of AQ forecast error - particularly for daily and 8-hour  $O_3$  maxima - due to e.g. biases in the  $NO_2$  forecast from PREV’AIR (Honoré et al., 2008). However, it appears that whether or not meteorological processes are the leading factor of AQ forecast error is dependent on the location of study. For example, the study of Honoré et al. is a long-term evaluation of the PREV’AIR forecast model over France, which is a continental country. Their result contrasts that for  $O_3$  evaluation in Sydney, Australia (Hess et al., 2004) where sea-breezes and other meteorological processes dominate the  $O_3$  error more than emissions or chemistry. It is therefore a reasonable hypothesis that a regional AQ forecast over an island - such as the UK - may experience more sensitivity to meteorology and lateral boundary conditions (LBCs) than a continental or land-locked region.

LBCs provide pollutant concentrations on the boundaries of limited area models, which, together with the winds at the boundaries, determine the in-flow of pollution into the model domain. These LBCs are derived from a larger-domain or global model. The in-domain concentrations of a given pollutant are dependent on an accurate estimate of transport through the boundaries and hence on the validity of the LBCs. Any errors in these boundary concentrations could propagate into the model interior and introduce errors. Potential errors associated with imposing inaccurate boundary conditions on regional forecast models have been widely studied (Staniforth, 1997; Tang et al., 2009; Davies, 2014) and it has been shown that LBCs are a significant source of pollutant concentration errors over land. In particular, upper-tropospheric winds transport polluted air masses from continental Europe to the UK under certain synoptic conditions, which contain both  $O_3$  and its precursors.

### 1.2.1 Observations: ground-based and remote

Observations of pollutant concentrations on various scales are imperative to evaluating the accuracy of an AQ forecast and providing error information which could be valuable to improving the forecast. Depending on the spatial and temporal scales of the forecast - i.e. whether the forecast covers global or regional scales; whether it is daily or seasonal - different types of measurements are appropriate, such as satellite data, ground-based operational networks or measurement campaigns focusing on a specific place and time. The observation type can also be dependent on the lifetime and abundance of the atmospheric species in question.

#### 1.2.1.1 Satellite observations

Observations of the global distribution of pollutants such as  $O_3$ ,  $NO_x$  ( $NO_2 + NO$ ) and aerosol optical depth (AOD) are measured by remote sensors on satellites. These are generally passive instruments which absorb radiation emitted from atmospheric species in the atmosphere. Satellite measurements are particularly useful in data assimilation and verification of global-scale models. Global models assimilate the observed data and combine them with a knowledge of physics and chemistry to create 3-dimensional profiles of e.g.  $O_3$ . An increase in abundance of global observational data of atmospheric constituents over the past couple of decades has been made possible by international collaborative programmes and satellite missions, e.g. the European Space Agency (ESA) ‘Envisat’ satellite (which ended in 2012); the EU Copernicus Atmosphere Monitoring Service (CAMS; [Schroedter-Homscheidt et al., 2016](#))<sup>1</sup>, which used data collected between 2009 - 2014 by the Monitoring Atmospheric Composition and Climate (MACC) project<sup>2</sup>. Although the use of satellite data for monitoring and evaluation of column  $O_3$  and  $NO_2$  is increasing in popularity, routine monitoring of boundary-layer and surface concentrations is made difficult many factors, e.g. lifetime of the species and vertical profile. For instance, the vertical profile of species such as  $O_3$  is such that concentrations are lower in the troposphere than the stratosphere, which reduces the satellite’s sensitivity to measure concentrations in the lower troposphere ([Kar et al., 2010](#); [Foret et al., 2014](#)). The Ozone Monitoring Instrument (OMI), which operates in the UV / visible range, is an example of an instrument with good sensitivity to column tropospheric  $O_3$  but weak sensitivity to surface concentrations, particularly in cloudy conditions, due to  $O_3$  UV absorption.

---

<sup>1</sup><http://atmosphere.copernicus.eu/>, date accessed 20-01-2021

<sup>2</sup><https://cordis.europa.eu/project/id/218793>, date accessed 20-01-2021

### 1.2.1.2 Flight-based and ground measurement campaigns

Specific campaigns are useful for gaining a thorough insight into the nature of pollution at a precise location. Because they focus on a specific place and time, one-off campaigns typically measure various parameters specific to the research question, such as pollutant concentrations alongside meteorological variables (e.g. fluxes), which are typically not included in routine measurements. One-off campaigns generally span periods of months to years and involve extensive planning. They are particularly useful for studying a region of interest in greater detail than is possible with just routine measurements, e.g. megacities with high pollution levels, often using a combination of ground-based instruments, instruments mounted on top of high-rise buildings and aircraft measurements. Large ground-based campaigns have been conducted in recent years to thoroughly investigate the composition, chemical and meteorological sources and sinks of pollutants such as O<sub>3</sub> and NO<sub>x</sub> in e.g. London (Bohnenstengel et al., 2015) and Beijing (Shi et al., 2019). Successful flight campaigns with e.g. the Facility for Airborne Atmospheric Measurements (FAAM; <http://www.faam.ac.uk>) aircraft have provided airborne measurements of the chemical composition of e.g. Saharan dust (Ryder et al., 2019) and joint modelling studies to investigate CO<sub>2</sub> from volcano eruptions (Ilyinskaya et al., 2018).

### 1.2.1.3 Ground-based routine observation networks

Campaigns are a powerful method of gaining thorough insight into pollution within a specific region and / or time, while satellites are useful to ensure a global coverage of monitoring O<sub>3</sub>, NO<sub>2</sub> and aerosols at low resolution. However, ground-based monitoring networks are still best suited for routine monitoring of regional and near-surface concentrations. In-situ networks provide direct surface measurements, which are more accurate within the region they represent than observations with remote sensing. They also provide data for archives of long-term measurements, from which robust statistical evaluations can be conducted. One disadvantage of ground-based networks on both global and national scales is the sparseness of the monitoring stations, which can introduce error due to poor representativeness of the surrounding area, in addition to uncertainties due to the sensitivity of the instrument itself. For most pollutants, representative errors in measurement are more dominant than instrument errors, due to the high spatial variability of the pollutant fields. The sparse coverage of measurements in remote areas is not an issue for a gridded satellite product or a carefully planned measurement campaign.

Even on national scales, the coverage of ground-based AQ measurement sites is irreg-

ular. For example the UK, the Automatic Urban and Rural Network (AURN) provides routine, quality-controlled measurements of various pollutants relevant to human health. The observations are used to verify the operational Air Quality Configuration in the Unified Model (AQUM; [Savage et al., 2013](#)) i.e. the regional AQ forecast model for the UK. More information about AURN will follow in section [3.2.1](#).

## 1.2.2 AQ forecasting in the United Kingdom

In the UK, the Department of Environmental, Food and Rural Affairs (DEFRA) issues the public AQ forecast. The 5-day forecast is provided by AQUM from the Met Office and is based on a 10-point scale called the Daily Air Quality Index (DAQI), corresponding to the highest concentrations of any one of the following: PM<sub>2.5</sub>, PM<sub>10</sub>, O<sub>3</sub>, NO<sub>2</sub> or SO<sub>2</sub>. The 10-point scale is grouped into four bands of risk: ‘Low’, ‘Moderate’, ‘High’ and ‘Very High’. The bands are based on concentrations shown in figure [1.1](#), which are calculated as running means over the indicated time periods. The DAQI was proposed by the Committee on the Medical Effects of Air Pollution (COMEAP) and was last reviewed in 2011 ([COMEAP, 2011](#)).

The purpose of the DAQI is to enable the public to mitigate adverse health risks associated with short-term episodes of high pollution concentrations. The DAQI levels are accompanied by recommended actions for people at risk (e.g. children and adults who suffer with asthma, lung and / or heart problems) - for instance, if the level is within the ‘Moderate’ band and above, at-risk groups should reduce strenuous outside exercise.

Because AQUM is central to this thesis, model specifications merit their own section

		Ozone	Nitrogen Dioxide	Sulphur Dioxide	PM <sub>2.5</sub>	PM <sub>10</sub>
					Particles (EU Reference Equivalent)	Particles (EU Reference Equivalent)
					24 hour mean	24 hour mean
Band	Index	Running 8 hourly mean	hourly mean	15 minute mean	24 hour mean	24 hour mean
	1	0-33	0-67	0-88	0-11	0-16
Low	2	34-66	68-134	89-177	12-23	17-33
	3	67-100	135-200	178-266	24-35	34-50
Moderate	4	101-120	201-267	267-354	36-41	51-58
	5	121-140	268-334	355-443	42-47	59-66
	6	141-160	335-400	444-532	48-53	67-75
High	7	161-187	401-467	533-710	54-58	76-83
	8	188-213	468-534	711-887	59-64	84-91
Very High	9	214-240	535-600	888-1064	65-70	92-100
	10	241 or more	601 or more	1065 or more	71 or more	101 or more

Figure 1.1: Daily Air Quality Index based on concentrations of 5 pollutants. Figure from [Connolly et al. \(2013\)](#).

and will therefore be outlined in more detail in section 3.4.

### 1.3 Forecast model verification

It is important the public receive an accurate AQ forecast, so that they can act accordingly. However, all types of forecast have a degree on uncertainty - a truth experienced by anyone who, after checking the weather forecast, has ever been caught out in a bout of rain which was not predicted. In order to improve operational forecast accuracy, it should be routinely compared against a reliable ‘truth’. Because of the many potential origins of pollutant forecast errors, a range of analysis methods is needed in order to fully understand the nature of the errors. There are many different types of forecast (e.g. deterministic, ensemble, probabilistic), with different verification methods best suited to each. Forecast verification provides a statistical evaluation of the accuracy of a forecast, and the following section will discuss the importance of this process, outlining the various methods and how some of them will be used in this thesis.

AQ forecast evaluation is largely based on point-based verification methods. These methods refer to those where a single grid-box of a gridded forecast is compared against an observation in the same grid-box - either a point measurement or a gridded field observation (Ebert, 2008). Alternatively, a single observation point is either compared to the nearest forecast grid neighbour, or the forecast from neighbouring grid-points is bi-linearly interpolated onto the observation point. The metrics used to evaluate the forecast comprise e.g. root mean square error (RMSE), frequency bias (FB) and Pearson’s correlation coefficient (PCC). Forecast evaluation based on contingency tables often uses the Equitable Threat Score (ETS; Gandin and Murphy (1992) - also known as the Gilbert Skill Score (GSS; Gilbert, 1884; Hogan and Mason, 2012)). However, each of the above metrics has various issues associated with their use: RMSE is dominated by large error values; PCC shows correlation but not causation; and despite its name, ETS is not strictly equitable (Hogan et al., 2010). While they are informative on whether a forecast is generally accurate at grid-point scale, these traditional methods provide no information about the source or nature of the error and as such, there is little scope for improving forecast accuracy.

Furthermore, it is important that the forecast matches the observation in both space and time for point-based metrics. However there exist situations in which an error is twice-counted for its displacement in both space and time - known as the ‘double-penalty’ problem (Brown et al., 2012b). It is a common phenomenon in point-based verification,

especially of high-resolution forecasts, but is only an issue when the features of interest are a few grid lengths in dimension. Relaxing the spatial or temporal constraints could allow the distinction of an otherwise useful forecast whose skill is not captured by looking at grid-point values only, even if the object in question (e.g. a rain band) is predicted at slightly the wrong time or place.

Methods which relax the spatial constraint are among the neighbourhood-based group, and include the Fractions Skill Score (FSS; [Roberts, 2008](#)), Upscaling ([Zepeda-Arce et al., 2000](#); [Weygandt et al., 2004](#)) and intensity-scale methods ([Casati et al., 2004](#)), all of which have been developed with the evaluation of precipitation forecasts in mind. Neighbourhood methods such as the Brier Score (BS; [Brier, 1950](#)) or the Continuous Ranked Probability Score (CRPS; [Hersbach, 2000](#)) are of the probabilistic nature, and as such they are best suited to evaluating ensemble forecasts. As the BS will be used later in this thesis, more detail about it and CRPS will be provided in section [4.3](#).

In the AQ forecasting community, traditional verification methods (as outlined above) are commonly used. This is largely due to the deterministic nature of most AQ forecasts. For example, in the multi-model evaluation of O<sub>3</sub> and PM as part of AQMEII ([Im et al., 2015a,b](#)), point-based statistics such as RMSE, PCC, normalised mean standard error (NMSE) and normalised mean bias (NMB) are used to find common biases between AQ models. These metrics are adequate to gain general insight into the traditional forecast errors within model inter-comparison studies. For example, in [Im et al. \(2015b\)](#), hourly surface O<sub>3</sub> is found to be underestimated by 18% on average across a large sample of coupled meteorology - chemistry models, and the PCC is generally above 0.8. Some models in the above study over-estimate night-time O<sub>3</sub> concentrations, which is attributed to under-estimating night-time NO<sub>2</sub> concentrations. PM<sub>10</sub> is generally under-estimated in both urban and rural areas in Europe, with under-estimations of 66% and 75% respectively [Im et al., 2015a](#).

However, in order to improve the accuracy of a specific regional pollution forecast, operational forecasting and model development communities require metrics which could reveal more information about the spatial or temporal relationship between the meteorological and AQ errors. Using a neighbourhood verification method to quantify the relationships between these two types of forecast is novel, as they are typically evaluated separately.

Meteorology and AQ forecast errors are not typically evaluated alongside each other, with long-term availability of both model and observation data. An evaluation with both

model and observational data is advantageous over model-only or observation-only studies, because systematic biases and behaviours in the AQ forecast due to the meteorology can be inferred from comparison against a reliable network of quality-controlled observations. This thesis uses three flavours of verification methods to quantify relationships between certain air pollutants and meteorological variables, which are:

1. point-based;
2. neighbourhood-based;
3. process-based methods.

While the first two methods have already been discussed, the last method on the list is different in nature. It is an evaluation of how physical processes in an AQ model could influence the concentrations of a chemical species through releasing tracers in an idealised simulation. For example, one could evaluate how inaccuracies in the representation of surface energy balance could affect the diagnosis of atmospheric stability, which could in turn affect the modelling of turbulence and thus vertical mixing of pollutants and their chemical precursors. All three of the above methods will be used within this thesis, and will be discussed further within the relevant sections.

## 1.4 Thesis aims and structure

This thesis is split into three working chapters, each one addressing a particular set of questions. Due to their differences, each chapter begins with a separate introduction, literature review and methods section. The AQ and meteorology forecasts used are from AQUM, with the exception of using meteorology from the UM Global and UKV configurations of the Met Office Unified Model in chapter 5. The research questions are outlined below:

### **1. How are surface AQ forecast errors related to forecast errors in meteorological variables when evaluated with traditional, point-based metrics?**

The evaluation in chapter 3 is split into two branches: 1) forecast errors in meteorological variables (10 m wind speed and 1.5 m temperature) are compared against forecast errors in O<sub>3</sub> and NO<sub>2</sub> during DJF and JJA 2017; and 2) forecast errors in precipitation are compared against those in PM<sub>2.5</sub> and PM<sub>10</sub> during DJF 2015. The evaluations are done using traditional, point-based metrics (RMSE, PCC), where the AQ variables are

evaluated at a number of rural and urban background sites, while the meteorology is evaluated at the nearest corresponding WMO station.

**2. Can neighbourhood-based verification metrics provide information about the spatial relationship between AQ and meteorological errors?**

Specifically, chapter 4 uses neighbourhood verification to evaluate the skill of both the meteorological and AQ forecast within the vicinity of the point observation. The purpose of this chapter is to find the extent of spatial influence of the 10 m wind speed forecast error on O<sub>3</sub> and NO<sub>2</sub> forecasts, and to test whether considering the precipitation forecast in a neighbourhood of a PM measurement site could provide information about the nature and effect of the precipitation error.

**3. Can process-based evaluation provide information about the cause of relationships between AQ and meteorological errors?**

Chapter 5 builds on the results of the previous chapters. A process-based approach is used to learn about the contribution of forecast errors in 10 m wind speed and boundary layer development on surface O<sub>3</sub> and NO<sub>2</sub> forecast errors. This is done through designing and completing a semi-idealised tracer experiment using the off-line NAME dispersion model, where the focus was on evaluating the influence of entrainment on the surface concentrations of tracers. NWP from the UM Global and UKV configurations of the Met Office Unified Model were used in the experiment, which was conducted for the period of JJA 2017.

There is also a final chapter 6, where the main conclusions from the thesis are presented in the context of existing literature and further discussion.

## 2. Background

This chapter covers the basic background chemistry (section 2.1) and meteorology (sections 2.2 and 2.3) required to understand the research completed in this thesis.

### 2.1 Atmospheric trace constituents

Described in this section are three atmospheric species central to this thesis: PM (section 2.1.1), O<sub>3</sub> and NO<sub>2</sub> (section 2.1.2). Chemistry of O<sub>3</sub> and NO<sub>2</sub> is outlined in terms of clean and polluted environments, followed by their spatial and temporal variability.

#### 2.1.1 Particulate Matter

Particulate matter (PM) is a family of particles and liquid droplets suspended in air (“aerosols”) containing both natural and anthropogenic species: mineral or volcanic dust, sea salt, nitrates, sulfates, black carbon, biomass burning aerosol, organic carbon from fossil fuels and other secondary organic aerosols. A commonly used definition of the size of the particles uses aerodynamic diameter and, in general, aerosol diameter ranges from a few nanometers (nm) to tens of micrometers ( $\mu\text{m}$ ). For instance, PM<sub>2.5</sub> are particles with aerodynamic diameter  $\leq 2.5 \mu\text{m}$ , and PM<sub>10</sub> is  $\leq 10 \mu\text{m}$ . This is based on the diameter of a sphere in order to simplify the additional complexity of irregular particle shapes.

Particles may be emitted directly (‘primary’) or formed in the atmosphere by being converted from gas to particle (‘secondary’). Depending on their size, particles are removed from the atmosphere through various processes. For example, coarse particles (i.e. diameter  $\geq 2.5 \mu\text{m}$ ) can be removed either through depositing to Earth’s surface (‘dry deposition’), or being physically washed out by falling precipitation (‘wet deposition’). These particles can reside in the troposphere on the timescale of a few days to weeks, depending on their size and the meteorological conditions. In the case of particles being mixed into the free troposphere, they can travel distances long enough to travel between continents. For instance, the UK and Europe sometimes experience interconti-

mental transport of dust from the Sahara desert (e.g. [Ansmann et al., 2003](#)), or volcano ash (e.g. [Dacre et al., 2011](#)).

### 2.1.2 Tropospheric O<sub>3</sub> and NO<sub>2</sub>

NO<sub>2</sub> belongs to the family of Nitrogen Oxides (NO<sub>x</sub> = NO + NO<sub>2</sub>), the majority of which is locally emitted through fossil fuel combustion (approximately 34% of which was from transportation in the UK in 2018 ([UK Department for Transport, 2021](#))). Most of the surface - emitted NO<sub>x</sub> is in the form of NO, but is photochemically converted into NO<sub>2</sub> on the timescale of minutes ([Dentener et al., 2001](#)). NO<sub>x</sub> has a relatively short lifetime of  $\leq 24$  hours within a polluted boundary layer, but up to 10 days in the upper troposphere.

O<sub>3</sub> is a secondary pollutant which occurs naturally in both the stratosphere and troposphere. It is naturally produced via photolysis during sunlight hours and by chemical reactions of its precursors: namely NO<sub>x</sub>, CO and volatile organic compounds (VOC) (e.g. [Crutzen, 1971](#); [Wayne, 2000](#)). Approximately 90% of total atmospheric O<sub>3</sub> resides in the stratosphere and absorbs solar UV radiation, preventing it from reaching the Earth's surface. The remaining 10% of O<sub>3</sub> found in the troposphere is widely regarded as a pollutant harmful to human health (e.g. [COMEAP, 1998](#)) and vegetation (e.g. [Karnosky et al., 2007](#)). It is therefore important to understand the underpinning physical and chemical processes which determine tropospheric O<sub>3</sub> concentrations, in order to model the, accurately. Chemical processes such as photolysis of NO<sub>2</sub> and VOC oxidation, deposition to surfaces, horizontal transport and vertical mixing are all important factors for modelling concentrations of tropospheric O<sub>3</sub> in the planetary boundary layer (PBL) ([Lee et al., 2003](#); [McNider and Pour-Biazar, 2020](#)).

Stratospheric O<sub>3</sub> can also be transported down to the troposphere. Studies using global chemistry transport models provide evidence that vertical exchange processes e.g. from stratosphere-troposphere exchange (STE) can contribute to the surface concentration levels in both the Northern and Southern Hemisphere ([Lelieveld and Dentener, 2000](#); [Williams et al., 2019](#)). O<sub>3</sub> originating from the STE process is usually attributed to long-range, intercontinental transport of precursors (or O<sub>3</sub> itself) above the PBL, where it must undergo vertical transport across the boundary layer top in order to reach the surface. This can often occur in the presence of extra-tropical cyclones, where the phenomenon called 'tropopause folds' mixes stratospheric air into the troposphere, usually in the wake of a cold weather front ([Gray, 2003](#)). STE accounts for approximately 10% of

the  $O_3$  found in the troposphere, i.e. an estimated  $500 \text{ Tg } O_3 \text{ a}^{-1}$  from a global budget of  $5000 \text{ Tg } O_3 \text{ a}^{-1}$  (Wu et al., 2007).

The remaining 90% of tropospheric  $O_3$  is produced by photochemical reactions involving the precursor pollutants mentioned earlier. In the absence of VOC emissions, the amount of total oxidant ( $O_x = O_3 + NO_2$ ) remains in photostationary equilibrium within the troposphere during daytime, and local concentration levels of both constituents near the surface are dependent on the local and regional concentration levels of their precursors. Photochemical production of tropospheric  $O_3$  occurs locally during sunlight hours, ceasing upon sunset. When local production stops, concentrations decay throughout the night as molecules are depleted through titration by NO and deposited to surfaces through dry deposition. As convective processes begin around sunrise, the photochemical production cycle begins again. Details of these processes are the subject matter of the following sections.

Tropospheric chemical loss of  $O_3$  is estimated to account for 80% of the global sink processes, while the other 20% is due to dry deposition to surfaces and uptake by vegetation (Brasseur and Jacob, 2017).  $O_3$  lifetime within the boundary layer is generally on the order of a few days.

### 2.1.2.1 Clean troposphere

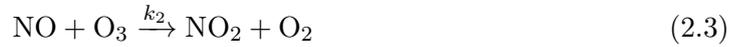
Production of  $O_3$  in clean, tropospheric air most commonly occurs via photolysis of  $NO_2$  and by oxidation of hydrocarbons and CO in the presence of  $NO_x$  (Cocks, 1993; Hobbs, 2000). In the absence of local pollutant emissions,  $NO_2$  absorbs photons above a threshold energy (wavelength  $\lambda \leq 420\text{nm}$ ) which efficiently detach the oxygen (O) atom to form NO and atomic oxygen (Crutzen, 1971):



where  $O(^3P)$  is a ground-state oxygen atom,  $j$  is a reaction rate coefficient on the order of  $10^{-7}$  to  $10^{-5}\text{s}^{-1}$  (depending on  $\lambda$ ),  $h$  is Planck's constant and  $v$  represents a photon, i.e.  $hv$  denotes the energy of the absorbed photon, thus the energy required to disassociate the oxygen atom from  $NO_2$ . Some of the  $O(^3P)$  (ground-state) atoms combine with  $O_2$  (Chapman, 1930; Schiff, 1969):



where  $M$  is another common molecule (such as CO or N<sub>2</sub>). This is the only chemical reaction within the troposphere which results in the creation of O<sub>3</sub>. The newly formed NO (reaction 2.1) is rapidly oxidised by O<sub>3</sub> (from reaction 2.2) to re-form the NO<sub>2</sub> molecule lost in reaction 2.1 and an oxygen atom:



where  $k_1$  and  $k_2$  are additional reaction rate coefficients (cm<sup>-3</sup>s<sup>-1</sup>). Altogether, reactions 2.1 to 2.3 establish a closed system under a photostationary equilibrium of the total oxidant, O<sub>x</sub>. This assumes that the overall rate of change of [NO<sub>2</sub>] is zero and thus we arrive at the description of mixing ratios of O<sub>x</sub>, known as the Leighton relation (Leighton, 1961):

$$\frac{[\text{NO}_2]}{[\text{NO}]} = \frac{k_2}{j} [\text{O}_3] \quad (2.4)$$

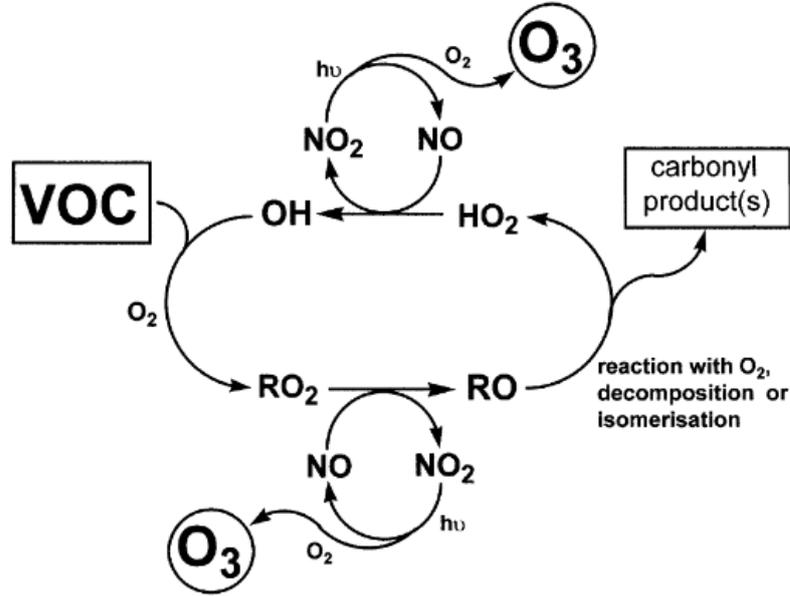
The concentrations of O<sub>3</sub> are therefore sensitive to the NO<sub>x</sub> mixing ratios. The advantage of combining O<sub>3</sub> and NO<sub>2</sub> into a single quantity, O<sub>x</sub>, is that any changes in O<sub>x</sub> are independent of the rapid photochemical reactions that convert O<sub>3</sub> to NO<sub>2</sub>, and vice versa. Thus an increase in O<sub>x</sub> represents the production of O<sub>3</sub> (via transport or photochemistry) or loss of O<sub>3</sub> (via transport or deposition). O<sub>x</sub> is an approximately conserved quantity when the O<sub>3</sub> precursor species (NO<sub>x</sub>, VOCs and the hydroxyl radical, OH) remain in equilibrium, as will be discussed in the following section.

During the night, O<sub>3</sub> is depleted by NO titration, as well as being deposited to surfaces via dry deposition. Since there is no net production of O<sub>3</sub> during the night, this means that generally, O<sub>3</sub> concentrations decrease during the night in both urban and rural areas.

### 2.1.2.2 Polluted regions

The photostationary equilibrium can be disrupted by oxidation of CO, or in the presence of VOCs which can be anthropogenic or biogenic. CO and VOCs act to enhance the conversion of NO → NO<sub>2</sub> without consuming O<sub>3</sub>. Some of the main sources of non-methane VOCs in the most recent report (2015) for the National Atmospheric Emissions Inventory were solvent use (36.9%), production processes (15.4%) and extraction / distribution of fossil fuels (11.4%); road transport only accounts for 2.7%, although it is locally significant (Tsagatakis et al., 2019). Consequently, high concentrations of O<sub>3</sub> may be observed within or downwind of urban regions, where the amount of non-methane VOCs

originating from the above sources is favourable for daytime production of  $O_3$ .



**Figure 2.1:** Schematic of  $O_3$  formation in the presence of VOCs. Figure 1 from [Jenkin and Hayman \(1999\)](#).

Figure 2.1 depicts the by-production of  $O_3$  from VOC degradation by the hydroxyl radical (OH) present in the atmosphere. The cycle involves fast reactions between organic peroxy ( $RO_2$ ) and oxy (RO) radicals. OH required in this process can be formed by photo-disassociation of  $O_3$  during sunlight hours (when  $\lambda \leq 310\text{nm}$ ) ([Biedenkapp et al., 1970](#)):



$HO_x$  and  $NO_x$  radicals catalyse the oxidation process of the VOCs. The alkyl radical R is rapidly oxidised by  $O_2$ , producing the  $RO_2$  radical. If NO is present, as it often is in polluted areas, it reacts with the  $RO_2$  to produce RO and  $NO_2$ , which continues with the photo-disassociative production of  $O_3$  and into the photo-stationary state (see equations 2.1  $\rightarrow$  2.3). Meanwhile the RO conversion into the hydroperoxy radical ( $HO_2$ ) can take on different routes, one of which involves reacting with  $O_2$  and producing carbonyl products alongside  $HO_2$ . (The details of this are outside the scope of this thesis, but more information can be found in e.g. [Jenkin and Hayman, 1999](#)).  $HO_2$  reacts with NO to complete the chain and regenerate OH, subsequently converting more  $NO \rightarrow NO_2$ . Thus, by degrading VOCs in polluted air,  $NO_2$  is produced without the consumption of  $O_3$ .

This NO<sub>2</sub> can then produce more O<sub>3</sub> through photolysis, as per equation 2.1.

### 2.1.2.3 Spatial and temporal variability

Emissions from road transport, solvent use and production processes are generally greater in towns and cities than in rural regions, thus NO<sub>x</sub> concentrations are generally higher in urban areas. Having outlined the importance of surface emissions in the previous section, one can understand that there is an equilibrium shift towards a net production of NO<sub>2</sub> during the day in urban areas (see also Cocks, 1993). Typical NO<sub>x</sub> mixing ratios within the boundary layer are observed to be at least one order of magnitude greater in sub/urban areas than in rural regions, and at least 3 order of magnitude greater than in remote marine locations (Seinfeld and Pandis, 2016). This highlights the influence of anthropogenic emissions as sources of NO<sub>x</sub> and its precursors, thus contributing to O<sub>3</sub> production or loss, depending on the NO : NO<sub>2</sub> mixing ratios (equation 2.4). In particular, decreasing NO results in increased O<sub>3</sub>, and vice versa.

Due to the generally higher NO<sub>x</sub> concentrations in urban than rural regions, O<sub>3</sub> production is usually limited by hydrocarbon availability; whereas low NO<sub>x</sub> concentrations are the major limiting factor for O<sub>3</sub> production in rural areas (Sillman et al., 1990). Therefore O<sub>3</sub> is depleted more rapidly by NO from near emission sources (e.g. from fossil fuel combustion, vehicle exhausts and chemical manufacturing) in urban regions. The scavenging of O<sub>3</sub> by NO generates NO<sub>2</sub> downwind of the local source (equation 2.3), or the primary NO may react with O<sub>2</sub> :



which may then go on to also produce O<sub>3</sub> as part of the photo-stationary cycle. This is not a major O<sub>3</sub>-production pathway, but one which can occur in a NO<sub>x</sub>-limited scenario. Studies such as that by Mauzerall et al. (2005) demonstrate the enhancement in O<sub>3</sub> production in rural regions, downwind of large point-sources of NO<sub>x</sub> in the US.

O<sub>3</sub> and NO<sub>2</sub> have variability on various timescales: daily, weekly and seasonal. For instance, a study by Bower et al. (1989) shows that annually averaged near-surface O<sub>3</sub> concentrations at ground-based UK sites are higher at weekends than weekdays. This effect is pronounced in urban areas, thus a contributing factor could be smaller NO emissions on Sundays, when the typical weekday rush-hour traffic is absent. The study also shows a strong negative correlation between NO<sub>x</sub> and summertime O<sub>3</sub> at 8 rural and

2 urban sites (correlation coefficient  $r = -0.91$ ).

However, the complexity of the relationship between  $O_3$  and  $NO_x$  goes beyond their dependence on sunlight and concentrations of hydrocarbons. The relationship is also influenced by meteorological conditions, whereby seasonal patterns are noticed even within the weekend / weekday split. Brönnimann and Neu (1997) expand on the  $O_x / NO_x$  relation by considering whether meteorological conditions are favourable for  $O_3$  production. The study, conducted over ground-based sites in Switzerland, finds that during conditions of low average wind speeds, high temperatures and solar radiation (i.e. favourable conditions), weekend  $O_3$  peaks are lower than on weekdays by up to 15%. It contrasts the findings of Pryor and Steyn (1995), where the opposite effect occurs in Lower Fraser Valley. Jenkin et al. (2002) model the trajectories linked to photochemical smog events in the UK and find that high  $O_3$  concentrations are, to a large extent, influenced by the UK's proximity to the European continent. The study confirms that its geographical location makes the UK susceptible to elevated  $O_3$  concentrations as a result of anticyclonic conditions and easterly winds, often occurring during the summer season. They also find that high  $O_3$  episodes are more likely to happen at the end of the working week (Saturdays and Fridays). They attribute this to the existence of a timescale associated with chemically transforming the accumulated VOCs and  $NO_x$  throughout the week into  $O_3$ .

Elevated surface  $O_3$  concentration levels tend to be a spring-summer phenomenon in the northern hemisphere largely due to more photochemical production than during the winter (Monks, 2000). For example, Lelieveld and Dentener (2000) shows that the Mace Head measurement site in Ireland ( $53^\circ$  N,  $10^\circ$  W) experiences a maximum in observed background concentrations of surface  $O_3$  between April - June for the 15 year study period of 1979-1993. An additional factor could be frequent spring-summer anticyclonic conditions and therefore high pressure, which could enhance accumulation of ground-emitted  $O_3$  precursors such as oxides of Nitrogen and hydrocarbon precursors emitted by motor vehicles, industrial sources, fuel combustion and solvents near the surface. Accumulation of these precursors near the surface can aid  $O_3$  production.

## 2.2 Atmospheric boundary layer meteorology

All of the meteorology relevant to this thesis occurs within the lowest parts of the atmosphere, which includes the troposphere and, within that, the atmospheric boundary layer (ABL). This is the layer of air nearest to the Earth's surface, its height defined by a

temperature inversion which typically occurs at 1000 - 2000 m during the day and 100 - 300 m at night, as depicted in figure 2.2 from Garratt (1992). The boundary layer height (sometimes also defined as the mixing layer height, specifically when considering aerosols) typically depends on the time of day, the amount of shear production and buoyant suppression of turbulence. Figure 2.2 illustrates the stages of ABL development. It shows that shortly after sunrise, solar heating of the Earth's surface stimulates convection and the subsequent ascent of air masses which rise to the top of the mixed layer. Throughout the day, vertical distributions of pollutants are expected to be well mixed in this layer, therefore almost constant with height. As the sun sets, surface heating ceases and the mixed layer becomes the residual layer, as part of the growth process of the nocturnal boundary-layer (NBL). The layer of air within around 300m to the Earth's surface becomes the NBL, which may be decoupled from the residual layer above by a temperature inversion, i.e. with a layer of air warmer than that below. The inversion is a stable layer which decouples the ABL from the rest of the troposphere, preventing further vertical pollutant transport. It therefore acts as a 'lid' on pollutant mixing in the vertical. The process begins again upon sunrise, when the NBL is eroded by increasing thermally generated turbulence, which eventually becomes larger than the effect of stability. The eroded NBL thus becomes one with the developing mixed layer during the early morning hours.

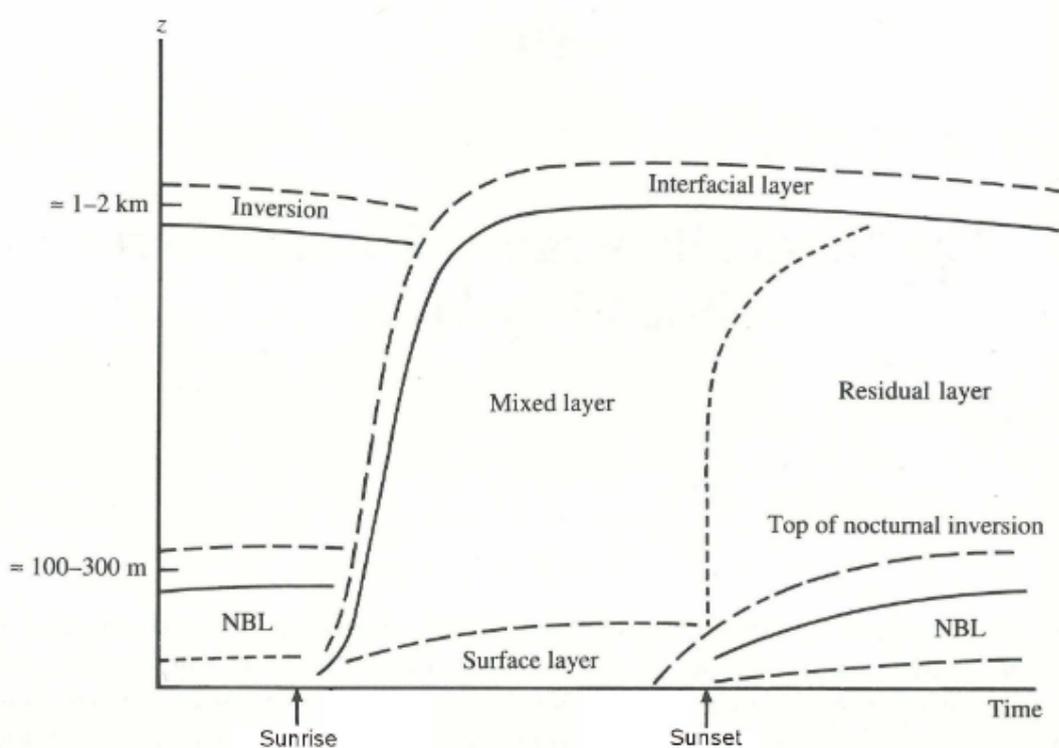


Figure 2.2: From Garratt (1992). Typical boundary layer structure throughout the day.

Harvey et al. (2013) evaluated Doppler lidar observations at Chilbolton (a rural location in the UK) and demonstrated that the depicted kind of ABL development under clear-sky conditions occurs only 0.9% of the 3-year study period, making it the 5th most probable boundary layer evolution. The most common diurnal evolution of the ABL is that of stable boundary layer with no cloud at night, changing to unstable during daylight hours. However, despite being the most probable occurrence, it only accounts for 6.4% of the time, which implies that there is great variability in the types of diurnal boundary layer development. The formation of the ABL is determined by stability, which is a factor of solar heating, surface heat fluxes and mechanical turbulence generated by wind shear.

### 2.2.1 Turbulence in the atmosphere

This section serves as an introduction to turbulence and its role within boundary layer meteorology. It is important to understand some of the theory of this process, as later in the thesis it will be demonstrated how the representation of turbulence within coupled AQ - meteorology models underpins the modelled boundary layer development and behaviour. The first-order closure problem is introduced in section 2.2.1.2, and an explanation of how modelled boundary layer stability depends on the Richardson number (Ri) is given in section 2.2.1.3. A description of the surface sensible heat flux and its role within boundary layer stability is outlined in section 2.2.1.4. Finally, we arrive at the concept of turbulent kinetic energy (TKE) and its relation to wind speed in section 2.2.1.5. Although the above processes may seem far-removed from the overarching theme of AQ forecasting, they underpin the meteorological processes addressed within this thesis and are therefore important to understand.

#### 2.2.1.1 Turbulence

Turbulence is associated with superposition of small-scale eddies in the atmosphere which carry small fluctuations of parameters such as horizontal and vertical momentum, moisture and heat. It is a dissipative process generated by two mechanisms within the ABL: wind shear and thermal convection. When warm thermals rise from the ground, positive turbulent kinetic energy (TKE) is generated, while cold surface temperatures and static stability of the ABL result in TKE consumption by buoyancy. Using Reynolds' averaging, fluctuations of relevant quantities can be separated from the grid volume-

average terms. This means that, e.g. buoyancy,  $B$  can be defined as:

$$B = \frac{g}{\theta} \overline{w'\theta'} \quad (2.8)$$

where  $g = 9.81 \text{ ms}^{-2}$ ,  $\theta$  is the potential temperature, and  $\overline{w'\theta'}$  is the covariance between vertical wind fluctuations and  $\theta$  fluctuations, corresponding to a vertical heat flux. When  $\overline{w'\theta'} > 0$ , the two variables are positively correlated and the environment is statically unstable. When  $\overline{w'\theta'} < 0$ , the correlation is negative and environment is statically stable.

If we consider an air parcel which has been vertically displaced with no heat exchange to its surroundings ('adiabatic'), the relation of its temperature relative to the new surroundings changes. It is considered a 'neutral' environment when the air parcel is of the same temperature as its surroundings. A 'statically stable' environment is one where, depending on the vertical profile of the temperature gradient, a vertically displaced parcel is either (a) colder or (b) warmer than its new surroundings and will be pushed back to its original position by buoyant restoration ([Stull, 1988](#)).

Wind shear generation of turbulence, denoted here as  $S$ , is a product of the momentum flux  $\overline{u'w'}$  and the mean vertical gradient of horizontal transport:

$$S = -\overline{u'w'} \frac{\partial \bar{u}}{\partial z} \quad (2.9)$$

Momentum and heat fluxes are key quantities within turbulence modelling, as they tell us about transport due to deviations from the mean (i.e. turbulence). Working with these leads to closure problems, which can be represented in models by making some assumptions discussed in the following section.

### 2.2.1.2 First-order closure

When considering turbulent flow, one encounters the closure problem: the number of unknown terms in an equation is larger than the number of equations. Equation 2.8 contains the vertical heat flux term, which is not closed because in order to solve it, higher order momentum terms (for which there are not enough equations in the set) need to be solved. In order to get around this, one can make assumptions about the higher-order terms in the set of equations needed to solve the problem. For example, by approximating equations above the zero-order mean variables (e.g. wind, temperature), only the unknown equations for the second-order moments are left, i.e. the turbulent heat

flux,  $\overline{w'\theta'}$ . Any turbulent flux can be approximated using the first-order, local closure:

$$\overline{u'_j \xi'} = -K \frac{\partial \bar{\xi}}{\partial x_j} \quad (2.10)$$

with scalar quantity  $K$  (units of  $\text{m}^2\text{s}^{-1}$ ) and  $\xi$  is any turbulent transport term, e.g. momentum or heat. When  $K$  is positive, the direction of transport of the flux represented by  $\overline{u'_j \xi'}$  is down the local gradient of  $\bar{\xi}$  - hence, called a first order K-closure theory. The theory breaks down for large eddy sizes and is only relevant on small turbulent scales. The scalar  $K$  is often referred to as ‘eddy diffusivity’ or ‘eddy viscosity’ ( $K_m$  for momentum,  $K_H$  for heat), and takes the value of zero in the absence of turbulence. Given the two mechanisms associated with turbulence (wind shear and buoyancy), both heat and momentum fluxes must be parametrised and the local K-closure theory allows the following approximations to be made:

$$\overline{w'\theta'} = -K_H \frac{\partial \bar{\theta}}{\partial z} \quad \overline{w'u'} = -K_m \frac{\partial \bar{u}}{\partial z} \quad (2.11)$$

Both turbulent flux terms are now related to their first-order moments,  $\frac{\partial \bar{\theta}}{\partial z}$  and  $\frac{\partial \bar{u}}{\partial z}$ . This is useful because the turbulent heat flux can now be expressed in terms of the dry-air sensible heat flux  $Q_H$  relation (e.g. [Ambaum, 2020](#)),

$$Q_H = \rho c_p \overline{w'\theta'} \quad (2.12)$$

where  $\rho$  is the air density and  $c_p$  is the specific heat capacity at constant pressure ( $1004 \text{ Jkg}^{-1}\text{K}^{-1}$ ). Finally, equation 2.8 can be expressed in terms of  $Q_H$ :

$$B = \frac{g Q_H}{\theta \rho c_p} \quad (2.13)$$

which will become important in the computation of boundary layer stability in a numerical model, as explained in the following section.

### 2.2.1.3 The Richardson number

The Richardson number (Ri) is a dimensionless measure which is used in models to quantify local stability. Using equations 2.8 and 2.9, the ratio of buoyancy to wind-shear

production of turbulence leads to the flux Richardson number  $Ri_f$ :

$$Ri_f = \frac{-B}{S} = \frac{\frac{g}{\theta} \overline{(w'\theta')}}{\overline{u'w' \frac{\partial \bar{u}}{\partial z}}}$$

These local gradients are difficult to measure, but they may be approximated from observations (e.g. using eddy covariance methods) at discrete model height levels. Substituting the terms for momentum and heat flux from the set of equations in 2.11, results in the bulk ( $Ri_B$ ) number, which evaluates the gradients at discrete model levels:

$$Ri_B = \frac{\frac{g}{\theta} \frac{\Delta \theta}{\Delta z}}{\left(\frac{\Delta u}{\Delta z}\right)^2} \quad (2.14)$$

Note that  $K_H$  and  $K_m$  are of the same order of magnitude. Ultimately,  $Ri_B$  is a measure of local stability in non-local terms. If  $Ri_B$  is larger than some critical value ( $Ri_{crit}$ ), turbulent flow becomes laminar as buoyant suppression in the numerator of equation 2.14 dominates the ratio.  $Ri_{crit}$  is often taken to have the value of 1, but in some model configurations it can be 0.25, depending on whether certain types of stability functions are used (Lock et al., 2017).

#### 2.2.1.4 Surface sensible heat flux

As shown in equation 2.13, buoyant forces are dependent on the magnitude and sign of  $Q_H$ , where a positive value leads to air parcel ascent through convective thermals.  $Q_H$  represents the loss of energy from the ground to surrounding air.

$Q_H$  is an important component in determining convective instability, and it also appears in the surface energy balance equation (Oke, 1987; Best et al., 2006):

$$C \frac{dT_s}{dt} = Q^* - Q_H - Q_E - Q_G \quad (2.15)$$

where  $C$  is the ground heat capacity,  $T_s$  is the surface temperature,  $Q^*$  is the net radiation at all wavelengths,  $Q_H$  and  $Q_E$  are the sensible and latent heat fluxes respectively, and  $Q_G$  is the ground energy exchange between the urban canopy and soil. When the surface temperature is at equilibrium, i.e.  $\frac{dT_s}{dt} = 0$ , the energy is conserved. Where  $Q_H > 0$ , the layer of air above the ground is colder than the surface, resulting in the flow of heat from the ground to the air. This scenario enables convection, and therefore vertical mixing of air. Where  $Q_H < 0$ , the layer of air above the surface is warmer than the ground, so

energy flows from the atmosphere into the ground.

A realistic approach to modelling  $Q_H$  throughout the day is necessary for an accurate representation of the effects of buoyant forces on turbulence and the representation of stability of the boundary layer. Correctly representing the nature of stability is crucial for determining the development of the PBL. Consider the case of  $Q_H$  having the wrong sign in the model upon sunrise, e.g. if it is negative where it would otherwise be observed to be positive. No convection would take place and the depth of the mixed layer (figure 2.2) would deviate from reality, as its growth would be hindered by only relying on growth through wind shear.

During night-time, and especially in the presence of clear-sky conditions,  $Q_H$  is usually negative, as the main exchange of energy is in the form of radiative cooling of the Earth's surface. This is often the case in rural regions. However, according to measurements in larger cities such as London,  $Q_H$  can remain positive throughout the night (Bohnenstengel and Hendry, 2016). The difference in energy balance between an urban region and its surroundings contributes to the 'urban heat island effect', named thus because of the characteristic higher temperatures over urban regions than surrounding rural regions (Oke, 1982). This difference can be on the order of 5 K in London, especially during clear-sky conditions as shown by Bohnenstengel et al. (2011). Therefore, if the land-surface scheme used in an AQ model does not resolve the complexity of the surface energy balance, any errors in modelled  $Q_H$  are likely to propagate into the calculation of buoyant production of turbulence, which may result in errors relating to the morning development of the mixed convective boundary layer.

### 2.2.1.5 Turbulent kinetic energy, wind speed and entrainment

Brasseur and Jacob (2017) state that concentrations of secondary reactive atmospheric species such as  $O_3$  are likely to be more abundant downwind of an emissions source than at its location. It is therefore intuitive that wind speed and wind direction near the surface is important for horizontal dispersion and vertical mixing of pollutants within the boundary layer. Wind speed can therefore be used as a proxy for the TKE within the boundary layer.  $V_{TKE}$  is a turbulence velocity scale whose definition comprises the turbulent wind vector components:

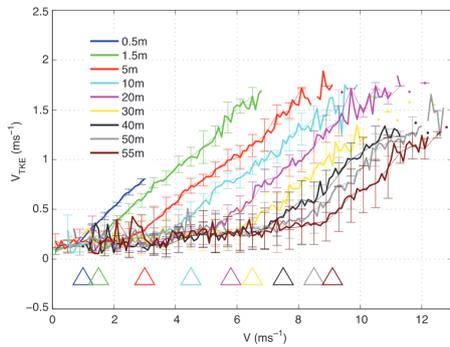
$$V_{TKE} = \sqrt{\frac{1}{2} (\sigma_u^2 + \sigma_v^2 + \sigma_w^2)} = \sqrt{TKE} \quad (2.16)$$

where  $\sigma_{u,v,w}$  are the standard deviation of the wind vector components. TKE is a measure of stability, which in turn determines the mixing ratios of pollutants throughout the boundary layer, as well as determining the size and strength of eddies which entrain air masses from the residual layer down into the boundary layer. The residual layer may also act as a night-time ‘reservoir’ of  $O_3$  which had been trapped in previous day’s mixed layer, before decoupling from the developing NBL by the nocturnal inversion (see the ‘sunset’ part of figure 2.2). It is therefore important to acknowledge that pollutant concentrations within the boundary layer could be affected by the amount of TKE, especially via the process of entrainment across the boundary layer top. Using a conservation equation for the budget of a chemical species, the entrainment flux of air from the residual layer as the mixed layer grows can be determined (Brasseur and Jacob, 2017):

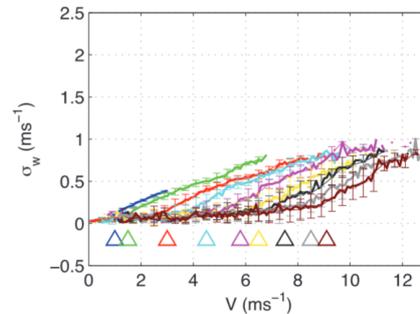
$$F(h) = (\rho_i - \rho_e) \frac{dh}{dt} \quad (2.17)$$

where  $F(h)$  is the entrainment flux at the boundary layer top (defined as positive upwards),  $\rho_i$  and  $\rho_e$  are the mass concentration of the chemical species within the mixed layer and the residual layer respectively, and  $\frac{dh}{dt}$  is the boundary layer growth rate. When  $\frac{dh}{dt} > 0$  and  $\rho_e > \rho_i$ , i.e. concentrations above the mixed layer are greater than concentrations below the boundary layer top while the boundary layer deepens, the entrainment flux  $F(h)$  acts to increase the concentrations within the mixed layer. Entrainment and TKE are therefore of particular interest in an evaluation of boundary layer concentrations of pollutant forecasts.

Sun et al. (2012) provide empirical evidence that in neutral and stable boundary layer regimes, there exists a relationship between TKE and wind speed ( $v$ ) when the



**Figure 2.3:** From Sun et al. (2012). Relation between wind speed and turbulence strength ( $V_{TKE}$ ) during stable conditions.



**Figure 2.4:** From Sun et al. (2012). Relation between wind speed and  $\sigma_w$  during stable conditions.

wind speed exceeds critical values ( $v_c$ ) at a range of altitudes. Above the critical wind speed  $v_c$ , turbulence generation is mechanical, i.e.  $R_i$  is small due to large vertical shear in the horizontal wind speed and air flow is turbulent. Whereas below the threshold, any mechanical turbulence is weak and generated through local fluctuations in wind shear. This means that the relationship between TKE and  $v < v_c$  is weak. Figure 2.3 shows that, e.g. for winds measured at a height of 10 m,  $v_c = 4.5\text{ms}^{-1}$ . Furthermore, turbulence is suppressed by buoyancy when mechanical generation is small.

The vertical velocity variance ( $\sigma_w$ ) component has been singled out in figure 2.4 to demonstrate its relation with wind speed,  $v$ . The relationship is such that at low heights (e.g. 5 m agl), smaller values of wind speed are associated with a specific  $\sigma_w$  than at higher altitudes (e.g. 50 m agl).  $\sigma_w$  is one of the required parameters within NAME's random walk parametrisation scheme for transport associated with unresolved mesoscale motion. Furthermore,  $\sigma_w$  is required by NAME's entrainment parametrisation scheme in order to calculate the flux of particles across the modelled boundary layer top (discussed in section 5.3.1).

## 2.3 Physical processes and meteorological variables affecting atmospheric species

Vertical and horizontal transport of gases or particles (hereafter collectively referred to as 'atmospheric species') ensures that concentrations are generally equally-distributed throughout the depth of the mixed daytime layer. However, the concentrations of atmospheric species within the boundary layer can be affected by different processes, depending on the species' physical properties. For instance, aerosols such as PM are efficiently removed from the atmosphere by washout (wet deposition), whereas gases such as  $\text{O}_3$  or  $\text{NO}_2$  rely on dry deposition as the main physical sink process, with further dependence on cloud cover which will be discussed in the following sections.

### 2.3.1 Surface winds

Winds at low altitudes are variable in direction and speed (i.e. shear production of turbulence), which means that pollutants are well-mixed but are unlikely to travel far because of small-scale eddy circulations near the ground. Within the troposphere, atmospheric species undergo vertical and horizontal mixing, as well as accumulation and dispersion, due to a turbulent atmospheric boundary layer. This is particularly impor-

tant for primary pollutants (e.g.  $\text{NO}_x$  and PM), whose horizontal distribution relative to their emissions sources is affected by the wind. The nature of the overarching synoptic pressure systems, and magnitude and direction of the wind speeds, is relevant for dispersion away from sources. These processes are also important for the mixing of secondary pollutants (e.g.  $\text{O}_3$  and also  $\text{NO}_x$ ), but other meteorological variables such as cloud cover are relevant to chemical transformations: for instance,  $\text{O}_3$  production is favoured during sunny, clear-sky days over days which are overcast. For example, [Cox et al. \(1975\)](#) found that during anticyclonic conditions (characterised by low surface winds, clear skies and elevated temperatures), observed levels of daytime  $\text{O}_3$  at three southern-UK sites were approximately doubled. Prevailing easterly winds during the anticyclonic period of study provided favourable synoptic conditions for tropospheric, long-range transport of polluted,  $\text{O}_3$  rich continental air masses into the UK - a result supported in [Jenkin et al. \(2002\)](#). In terms of physical sinks, [Wu et al. \(2007\)](#) analyse  $\text{O}_3$  concentrations within a number of global models and estimate that although chemical reactions amount to around 80% of the tropospheric  $\text{O}_3$  losses, the remaining major sink processes are due to deposition to surfaces, which is described in the following section.

### 2.3.2 Dry deposition

Dry deposition is a physical process which needs to be accounted for in an AQ model. It is a major sink of tropospheric gaseous species and accounts for about 25% of the total global  $\text{O}_3$  removal from the troposphere ([Lelieveld and Dentener, 2000](#)). The role of dry deposition is particularly important for  $\text{O}_3$  over vegetated areas during sunlight hours, when plant stomata open and  $\text{O}_3$  deposits to their interior ([O'Connor et al., 2014](#)). It is therefore important to understand this process in order to accurately model daily surface  $\text{O}_3$  concentrations in both global and regional chemistry transport models.

Dry deposition involves a direct transfer of chemical species from the atmosphere onto surfaces on the Earth. Chemical species such as  $\text{O}_3$  are transported from source regions to a region within proximity to the Earth's surface, where they encounter and are deposited on surfaces such as buildings, vegetation and terrestrial surfaces. This deposition is enhanced in a turbulent atmosphere, whereby air parcels are repeatedly brought into contact with surfaces ([Wayne, 2000](#)). Therefore it is required for turbulent mixing to continue near the surface in order for dry deposition to take place during night-time ([Derwent et al., 1998](#)).

Over continents, an average deposition velocity of  $v_d = 0.4 \text{ cm s}^{-1}$  ([Seinfeld and](#)

Pandis, 2016) and a night-time boundary layer depth of around 100m means that air masses nearest to the mixing height level may be brought into contact with a surface on the time-scale of 6 hours, i.e. the time-scale of this process is comparable to the duration of one night and cannot be ignored when evaluating diurnal cycles.

The efficiency of diffusion of the chemical species through the surface depends on many factors, such as the material, shape, roughness and moisture of the surface; environmental conditions such as temperature, humidity and thus season; as well as the nature of the species in question. Chemistry parametrisation schemes in models therefore have to account for all of these parameters. Many chemistry schemes, such as United Kingdom Chemistry and Aerosols (UKCA) (O'Connor et al., 2014) use the dry deposition approach of Wesely (1989), similar to that of electrical resistances in series, where  $v_d$  is calculated as:

$$v_d = \frac{1}{r_t} = \frac{1}{r_a + r_b + r_c} \quad (2.18)$$

where  $r_t$  is the total resistance and  $r_a$ ,  $r_b$ ,  $r_c$  are the aerodynamic, quasi-laminar layer (i.e. the layer of air closest to the surface), and surface (or canopy) resistances, respectively.  $r_a$  is most dependent on the meteorological conditions such as 10m wind-speed,  $r_b$  is a molecular diffusion term and  $r_c$  depends on surface type, thus making it directly dependent on the land surface scheme used in the model (C. Hardacre, *personal communication*). For example, Hardacre et al. (2015) conducted an evaluation of global chemistry transport models to find substantial differences across the ensemble of 15 models of O<sub>3</sub> dry deposition, driven by differences in land cover. They found  $v_d$  can range between  $v_d = 0.1\text{cm s}^{-1}$  to  $v_d = 2\text{cm s}^{-1}$ , with the largest inter-model differences in O<sub>3</sub> dry deposition over oceans and tropical forests.

### 2.3.3 Wet deposition

The process involving scavenging of atmospheric species by hydrometeors – such as rain, snow, cloud or fog droplets – is encompassed by the general term ‘wet deposition’. The umbrella term refers to a composite of numerous micro- and macro- processes, of relevant length scales ranging from  $10^{-6}$  m (e.g. cloud-physics processes) to  $10^6$  m (e.g. synoptic scale fronts). Cloud droplets and raindrops ( $10^{-5}$  to  $10^{-3}$  m) are involved with two scavenging processes, through which atmospheric species are brought into contact with condensed water and removed from the atmosphere by downward transport towards

the Earth’s surface. These processes are:

- **rainout** - uptake of species by droplets within the cloud and subsequently precipitating out the scavenged material;
- **washout** - scavenging of species by collision of falling rain or snow droplets.

Both processes happen in the presence of cloud, while washout occurs specifically in the presence of falling rain or snow. Physical and chemical micro-processes (e.g. aqueous chemistry) within the cloud are vastly important for the fate of the scavenged species, as such both washout and rainout are included in the AQUM model. However, this thesis considers only the sensitivity of washout on pollutants in the atmosphere; in particular, what happens to particulate matter in the presence of rain and snow.

A falling raindrop may or may not successfully collide with an aerosol particle, depending on their relative sizes, velocities and their collision efficiency. The efficiency of removal of species through wet scavenging is therefore dependent on the type and solubility of the species. Collision efficiency ( $E$ ) is defined as “the ratio of the total number of collisions occurring between droplets and particles to the total number of particles in an area equal to the droplet’s effective cross-sectional area” (Seinfeld and Pandis, 2016). Where  $E = 1$ , all aerosol particles within the geometric volume that a falling raindrop sweeps out are collected, although typically  $E \ll 1$ . Finding analytical solutions for estimating  $E$  becomes a complex fluid mechanics problem when flow patterns around the falling raindrops are considered, and though it is an interesting route to explore the numerical solutions within the model’s treatment of wet deposition, it is outside the scope of this thesis.

In AQUM, wet deposition is parametrised as a first-order loss rate (Savage et al., 2013), which is a function of a species-specific scavenging coefficient  $\lambda$ , precipitation rate  $R$  ( $\text{mm hr}^{-1}$ ) and particulate concentration,  $C$  ( $\mu\text{gm}^{-3}$ ) (Jones et al., 2019a). For example, a first-order removal rate is assumed for mineral dust particles:

$$\frac{\delta C}{\delta t} = -\lambda R^a C \quad (2.19)$$

where  $\lambda$  is of the order  $10^{-5}$  to  $10^{-4}$  (experimentally derived),  $a = 1$  for mineral dust but  $a = \frac{2}{3}$  for other particulates such as nitrate or sulphate aerosols. This means that the rate of PM removal will be greater when concentrations are high than when concentrations are low (although not proportionally when  $a < 1$ ). The fact that wet deposition is directly dependent on the precipitation rate means that there should exist a relationship between

PM concentrations and rain rate,  $R$ .

#### 2.3.4 Cloud cover

The amount of cloud cover present influences daytime and nighttime temperatures, as well as relative humidity and short-wave solar radiation. With cloud present, the diurnal range of surface temperatures has been found by Dai et al. (1999) to reduce by 25 - 50% relative to clear-sky days, as daytime temperatures are suppressed by reduced short-wave solar radiation and night-time temperatures are enhanced by increased infrared radiation back to the surface. Temperature and insolation can affect photolysis rates and emissions of biogenic hydrocarbons (e.g. isoprene and monoterpene), where emissions increase with temperature and reach a maximum at approximately 40°C, then decrease rapidly (Guenther et al., 1993; Zhang et al., 2018). Due to scattering of UV radiation, high cloud cover can therefore decrease photochemical O<sub>3</sub> production as described in section 2.1.2. Aqueous chemistry within clouds can also reduce tropospheric O<sub>3</sub> (especially if the regime is NO<sub>x</sub>-rich) as well as other atmospheric constituents (e.g. formaldehyde, hydroxyl and NO) (Lelieveld and Crutzen, 1990). High cloud cover can also increase relative humidity, which can enhance dry deposition of O<sub>3</sub> by stomatal uptake, thus reducing O<sub>3</sub> concentrations (Kavassalis and Murphy, 2017). In terms of transport, a thick cloud cover can inhibit ventilation of pollutants out of the boundary layer.

Forecasting cloud cover remains a challenging parameter within NWP due to many different factors, e.g. the representation of boundary layer turbulence within models, and the non-linear effect on cloud formation due to anthropogenic aerosols which act as cloud condensation nuclei (Dabberdt et al., 2004). Some studies exist to quantify O<sub>3</sub> forecast error attributable to forecast errors in cloud cover. For example, it has been shown that the WRF-Chem model only predicts 55% of clouds in the correct location and under-predicts optical depth, resulting in over-estimation of summertime hourly surface O<sub>3</sub> by up to 60 ppb (approximately 120 μgm<sup>-3</sup>) (Ryu et al., 2018).

#### 2.3.5 Summary

Providing an accurate AQ forecast can be difficult because not only do models have to account for the vast quantity of chemical reactions for each species, but also the underlying meteorology and physical processes need to be represented accurately, for reasons outlined in this chapter. Here, some of the main physical (e.g. boundary layer development, transport and mixing within the boundary layer, dry and wet deposition, cloud cover)

and chemical (photochemistry of  $O_3$  and relationship with  $NO_2$ ) processes have been outlined. In addition, modelling of secondary pollutants can also be sensitive to the choice of an emissions inventory for the precursor species (Coates et al., 2016; Travis et al., 2016), which adds another element of uncertainty to the forecast. It is important to acknowledge that forecasts always have a degree of uncertainty, and biases may present themselves upon verification.

## 3. Point-based evaluation of AQ and meteorological variables

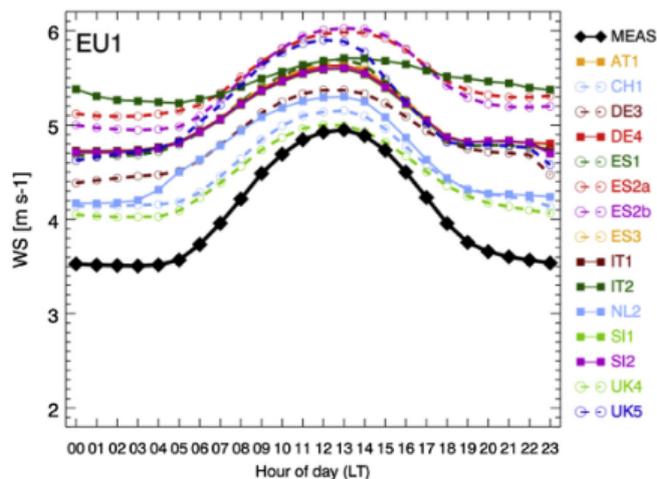
### 3.1 Introduction

AQ models produce gridded forecast fields of pollutant concentrations, which need to be compared against observations in order to ensure good model accuracy, as well as to identify areas for model improvement. Forecast verification is a tool used to answer a variety of questions about the forecast, in order to assess its ‘skill’ or ‘value’, either relative to another forecast or in an absolute sense (Joliffe and Stephenson, 2012). A related question is one which asks, e.g. is the current model upgrade better than its predecessor? In this chapter, an absolute evaluation of the forecast model is conducted, in the sense that we wish to find out how close is the model representation of the desired variable to observed values. This question is just as applicable to forecasting pollutant concentrations as it is to meteorological variables, such as wind speed or precipitation. Various verification methods serve to answer this question, and the simplest methods provide the most basic - but still useful - answers.

Some of the point-based summary scores frequently used in both air quality forecasting and NWP are: bias, root-mean-square-error (RMSE), mean error, Pearson-r correlation coefficient (PCC) between predicted and observed values, and contingency tables (with respect to defined thresholds) (e.g. Wilks, 2020). There is merit in using these kind of scores, as they provide a simple to understand overview of the model’s performance based on selected parameters<sup>1</sup>. For example, mean error is a measure of the difference between the predicted and observed variable and can be of varying sign and magnitude. Bias tells us about the systematic error of the forecast model when compared to observations, averaged over the whole dataset. RMSE is used to check the average magnitude

---

<sup>1</sup>A useful source for familiarisation with forecast verification metrics is the website of the Joint Working Group on Forecast Verification Research (JWGFVR), with contributions from established researchers in the field of verification: <https://www.cawcr.gov.au/projects/verification/>



**Figure 3.1:** From Brunner et al. (2015). A comparison of mean diurnal cycles of 10 m wind speed from coupled chemistry - meteorology models over a European domain.

of the forecast error, without gaining information about the direction of error deviations. PCC tells us whether there is a linear relation between two sets of data, but does not identify whether one set depends on the other, merely shows whether they are correlated.

A first-order answer from these traditional metrics can be sufficient to describe how well the forecast matches the observations, but it does not necessarily explain the errors without further investigation. For example, Brunner et al. (2015) successfully compare meteorological variables from a number of coupled meteorology – chemistry transport models (CTMs) using, among others, mean error and PCC, and find that many European models over-estimate surface wind speed (figure 3.1). By examining the diurnal cycle, they found that the positive bias in surface wind speed is most pronounced at night, when the boundary layer is most likely to be stable. Because wind speed (and, to some degree wind direction) is an important parameter for modelling pollutant dispersion and mixing, it is valuable to know whether it generally has a positive or negative bias. Although more sophisticated methods are needed to establish the source of the wind speed bias (which could be related to the use of long-tailed stability functions within operational models’ boundary layer schemes), it is important to acknowledge the bias’ existence (especially at a particular time of the day) because it may lead to an error in the forecast of other variables dependent on it, such as pollutant species.

A point-based evaluation of gridded forecast fields using traditional summary metrics has its drawbacks. For example, area-averaged forecast values will suffer from inability to represent the sub-grid heterogeneity of the forecast variable (Gilleland et al., 2009). Thus when compared against a point measurement, the grid-box average is likely to provide

a general closeness to the isolated measurement but not represent it fully - hence, this is an issue of ‘representativeness’ (e.g. [Haiden et al. 2012](#)). Another way of thinking about this is to ask: “how well does a point measurement represent the concentrations within its surroundings?”. It is especially relevant for the more inhomogeneous chemical species such as NO<sub>2</sub>, which could have very different local measurements at sites within a radius as close as 5 km ([McNair et al., 1996](#)). If the grid-box spacing is larger than this, it means that there is sub-grid variability in the pollutant concentrations which, by definition, is not captured in the model. Furthermore, orography and variable terrain types surrounding a point measurement can influence the pollutant concentrations at the measurement site location because of local-scale flow patterns, e.g. within valleys, or in urban environments. This could lead to an error in the prediction of a pollutant species at the location, due to unresolved terrain and land-use within the grid-box and its vicinity.

To overcome issues in comparing a gridded forecast to point observations, one could turn to spatial interpolation of observations onto a regular grid - such as ordinary kriging, which is already implemented in the AQUM verification suite ([Neal et al., 2014](#)). While this provides some idea of what the observation may be in a region to which a real measurement was interpolated, the resulting gridded fields are smoothed out and can under- or over-estimate the values at the interpolated locations. Another way to overcome the representativeness problem is to use a more sophisticated spatial verification method. A plethora of such methods has emerged in recent years (described in e.g. [Ebert et al., 2013](#)), some of which will be discussed in chapter 4. These are particularly useful when evaluating gridded forecast fields against point measurements, as they relax some of the spatial constraints and consider a neighbourhood around the point observation, thus minimising (or eliminating) the effect of the “double penalty” error on the perceived forecast skill ([Gilleland et al., 2009](#)).

For the purpose of this chapter, traditional point-based methods only are used to summarise the overall accuracy of the AQUM forecast for some air quality and meteorological variables. The chosen meteorological variables are 10 m wind speed, 1.5 m temperature and precipitation, while the pollutant species potentially affected by errors in the above variables are O<sub>3</sub>, NO<sub>2</sub>, PM<sub>2.5</sub> and PM<sub>10</sub>.

### 3.1.1 Research questions

As noted in section 3.1, many coupled meteorology - chemistry models over-estimate night-time surface wind speed, which may directly affect the modelled transport of pol-

lutants away from sources. Wind shear can also generate turbulence, which contributes to boundary layer development and entrainment of air masses from above the boundary layer. The entrained air masses may have higher or lower concentrations of a particular pollutant than air within the boundary layer, thus acting to either dilute or enhance concentrations within the boundary layer. Thus surface wind speed can also have an indirect effect on pollutant concentrations. This process is particularly important for  $O_3$ , whose concentrations can persist as the daytime boundary layer de-couples from the surface when night approaches. By chemical associations discussed in previous sections,  $NO_2$  is closely related to  $O_3$  and will thus also be affected by the processes which affect  $O_3$ , i.e. surface wind speed. There is potential for forecast errors in surface wind speed to directly affect the forecast atmospheric species through dispersion, thus the first research question to be addressed in the present chapter is:

**1) How are 10 m wind speed and 1.5 m temperature forecast errors related to  $O_3$  and  $NO_2$  forecast errors in AQUM?**

Surface temperature is included because, as pointed out in [Brunner et al. \(2015\)](#), daytime 1.5 m temperature is a parameter often under-predicted, with a bias of around 1 K across the surveyed models. Temperature is influential to chemical species and biogenic VOCs, whose reactions may be temperature-dependent, thus errors in temperature may propagate to the prediction of concentrations for certain chemical species via incorrectly predicting the chemical transformations (as described in section [2.1.2](#)).

Secondly, the effect of precipitation forecast errors on the PM forecast are of particular interest, as it is well known that precipitation is a major sink of PM from the atmosphere. Verification of precipitation is an interesting topic in itself, aspects of which will be explored further in chapter [4](#). Sometimes, one is just interested in whether it rains at all, or not - irrespective of how much. This can also be true for the PM forecast, whereby even a small amount of rain can act to wash out the particulates (see section [2.3.3](#)). The second question addressed in this chapter is therefore:

**2) How are forecast errors in precipitation related to AQUM forecast errors in particulate matter?**

An elaboration on the above research questions could explore potential cross-correlations between the meteorological variables: for example, precipitation is not always independent of wind speed due to e.g. the passage of a large-scale precipitating front, where the conditions at ground-level can be windy. Therefore it could be beneficial to analyse surface wind speed and precipitation together, both in terms of their effect on

$O_3$  and  $NO_2$ , as well as  $PM_{10}$  and  $PM_{2.5}$ . In addition to wind speed, the direction is also important for pollutant dispersion, both at the kilometer-scale (e.g. for dispersion from strong pollutant sources) and for large-scale transport (e.g. continental transport of desert dust). Two other meteorological variables worth investigating are relative humidity and cloud cover, both of which are closely related to precipitation and can exhibit a strong influence on the chemistry of atmospheric species (as introduced in section 2.3.4). Due to lack of time, relative humidity and cloud cover were not evaluated in this thesis. It was decided that priority should be given to precipitation and wind speed, predominantly because observations are readily available for both in order to evaluate forecast errors, and are both continuous in nature which simplifies the evaluation. Wind direction observations are often registered as ordinal names ('north', 'east', etc), which could introduce verification error through comparison with forecast wind direction in continuous polar co-ordinates, therefore it was decided not to evaluate this parameter.

The remainder of the present chapter is structured as follows. First, section 3.2 provides details of the observational datasets used in the evaluations, including a discussion on typical types of error present in measurements. Introduced are the AURN network, the World Meteorological Organisation (WMO) synoptic sites, the Met Office Integrated Data Archive System (MIDAS) database for the meteorological variables, and the NIMROD radar measurements for the precipitation. Then, the Met Office Unified Model will be introduced in section 3.3, followed by an introduction to AQUM in section 3.4. Methodology and results of the air quality forecast evaluation will be presented in section 3.5, where mean forecast error is calculated for the pollutants aggregated over all observation sites. Then, wind speed and temperature are evaluated on a diurnal-cycle basis, while precipitation is compared against radar and rain-gauge observations using categorical statistics from a contingency table in section 3.6. The results of sections 3.5 and 3.6 will enable the answering of the two main research questions asked in this section, which will be presented in sections 3.7.1 and 3.7.2, which will investigate direct relationships between the two types of forecast using correlation coefficients: 10 m wind speed and 1.5 m temperature with  $O_3$  and  $NO_2$ ; then precipitation against  $PM_{10}$  and  $PM_{2.5}$  using error composites.

## 3.2 Observations

This section will introduce the observations used throughout this thesis (AQ and meteorological variables), followed by an overview of the typical errors these may be

susceptible to.

### 3.2.1 Automatic Urban and Rural Network (AURN)

Routine surface observations of pollutants in the UK are provided by the Automatic Urban and Rural Network (AURN). Funded by DEFRA, AURN forms the largest network of automatic pollution measurement stations in the UK comprising 228 sites, 150 of which are active at the time of writing<sup>2</sup>. AURN is a network of irregularly spaced, surface-measurement sites found at various types of location: rural, urban, suburban; and are further split into categories: background, traffic and industrial sites. Because of the variability in the surroundings of measurement sites, those classified as ‘background’ are located at a distance away from local sources (e.g. roads, production plants and busy town centres) and thus give an overview of the ambient concentrations of a pollutant, without being largely affected by strong local sources. The bulk of the stations are located in urban areas (139), 60 of which are urban background sites. Those classified as ‘urban’ are located in towns and cities, while ‘rural’ sites are generally located in remote areas, such as the countryside. ‘Suburban’ sites are usually located within the outskirts of a city or town.

The stations measure concentrations of pollutants deemed most harmful to human health: O<sub>3</sub>, NO<sub>2</sub>, SO<sub>2</sub>, CO, PM<sub>2.5</sub> and PM<sub>10</sub> (COMEAP, 2011). Although the AURN network provides mean hourly measurements of the 6 major pollutants listed above, some stations only measure a selection of them. Data is available from 1973 to present day.

#### 3.2.1.1 Instruments at AURN sites

O<sub>3</sub>, NO, NO<sub>2</sub> and PM are measured at ground level, through the methods of: absorption of ultraviolet (UV) radiation (O<sub>3</sub>); chemiluminescence (NO and NO<sub>2</sub>); and for PM, a combination of: the Filter Dynamic Measurement System, Beta-Attenuation Monitor and Partisol (DEFRA, 2019). Measurement uncertainty of instruments must fall within the European standard compliance target of  $\pm 15\%$  for gaseous analysers and  $\pm 25\%$  for PM instruments (Eaton, 2016) for measurements to be valid and accepted to the database. As of 2015, there exist 81 O<sub>3</sub> and 137 NO<sub>x</sub> active measuring instruments in the AURN network, with each device undergoing bi-annual quality assurance and quality control (QA/QC) and inter-calibration. During this process, the response of each analyser to a

---

<sup>2</sup>For AURN sites, see <https://uk-air.defra.gov.uk/networks/network-info?view=aurun>, date accessed 25-09-2020)

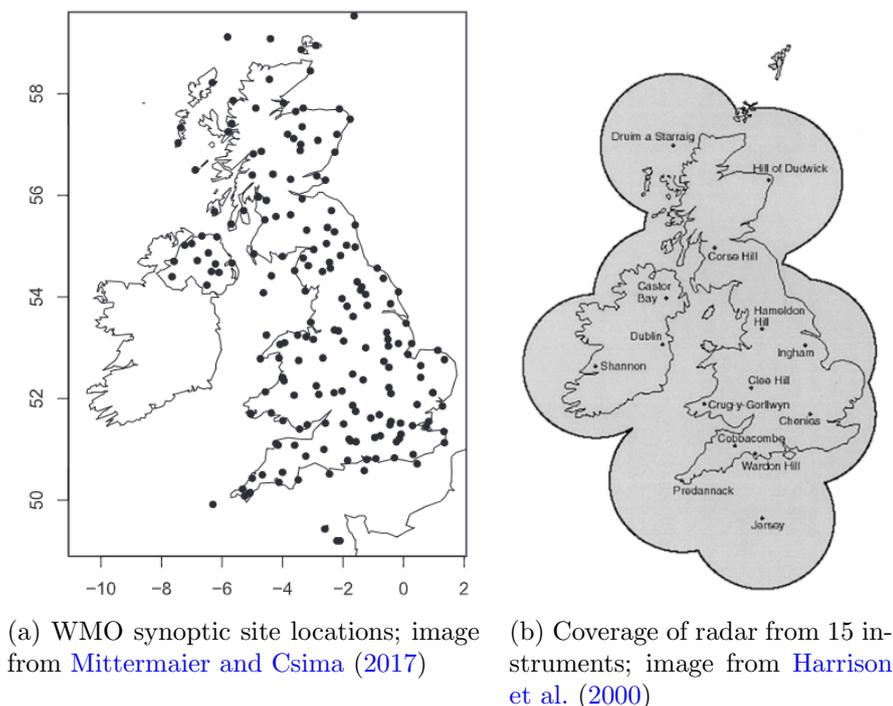
specified amount of gas is measured, the results of the whole network are compiled and problem sites are re-calibrated if their result falls beyond a defined uncertainty threshold of the reference standard (e.g.  $\pm 5\%$  for  $O_3$  photometers,  $\pm 10\%$  for  $NO_x$  analysers). Outlier percentages can vary seasonally and annually. For example, out of all the tests done on  $O_3$  analysers in the summers of 2015 and 2016, outlier percentages were 14% and 21.3% respectively, whereas for the winter seasons they were 16% and 33%. Annually, the overall data capture of  $O_3$  for 2015 and 2016 was 93.85% and 94.21% respectively, which is above the EU Air Quality Directive target of 90% (Eaton, 2016). In conclusion, the observational data provided by AURN is of high standard and can be reliably used for the purpose of station-based forecast model verification.

### 3.2.2 WMO synoptic sites and radar

Ground-based observations of meteorological variables (10 m wind speed, 1.5 m temperature and precipitation) used in this thesis are from the World Meteorological Organisation (WMO) network of land-surface synoptic observation sites (LND SYN). There are currently 166 such sites around the UK (figure 3.2a), taking hourly measurements and undergoing regular quality control checks. Hourly rainfall accumulations are observed with rain gauge instruments. The gauge measurements are stored in the Met Office Integrated Data Archive System (MIDAS) database, accessible from the CEDA archive, Met Office (2012). Synoptic sites are located in open areas such that local features (e.g. trees, buildings, steep ground etc) do not interfere with the measurements. As well as rain gauge observations, precipitation data used in routine verification of the Met Office forecast also come from the fully automated Nimrod system (Golding, 1998) which produces the radar analyses. These comprise hourly rainfall rates gridded to either 1, 2 or 5km resolution. Although throughout this thesis the radar data is simply referred to as “radar”, it is in fact a composite of raw observations from satellite, radar and surface synoptic reports, alongside prediction from NWP. The Nimrod radar product also undergoes a de-cluttering process at each of the 15 radar instrument sites, in order to remove noise and spurious echo. Images are subsequently re-mapped to a 5 km grid, spanning the UK as shown in figure 3.2b. A general overview<sup>3</sup>, detailed information about the calibration and error correction processes in the radar product can be found in Golding (1998) and Harrison et al. (2000). Data is accessible from Met Office (2003).

---

<sup>3</sup>A factsheet and brief history can also be found at [http://artefacts.ceda.ac.uk/badc\\_datadocs/nimrod/factsheet15.pdf](http://artefacts.ceda.ac.uk/badc_datadocs/nimrod/factsheet15.pdf)



**Figure 3.2:** WMO synoptic sites and UK radar coverage.

### 3.2.3 Typical errors in observations

Both AQ and meteorological measurements are susceptible to different types of error. The errors can usually be classified into one of the following categories: instrumental, representivity and sampling error.

**Instrumental error** is one where the measurement taken by the instrument is not close to the truth. For example, a tipping bucket raingauge can register an inaccurate amount of rain by not capturing any raindrops during the process of tipping. A PM filter could measure an inaccurate amount of particles due to blockage. The measurement uncertainty, as described in section [3.2.1.1](#), is typically on the order of  $\pm 20\%$ , depending on the measured species.

**Representivity error** is one where the spatial resolution of the measurement and the model do not match, i.e. the measurement resolves a higher spatial scale than the model is capable of. For example, convective precipitation can occur on the spatial scales of meters to kilometers, which a radar system will be able to resolve better than a coarse forecast model. In the case of atmospheric species, such as PM,  $O_3$  or  $NO_2$ , representative errors will generally have a larger influence on the perception of model bias than instrument error, because spatial variability in the pollutant fields is larger

than the instrument variability.

**Sampling error** is related to biases in the sample against the overall population. For example, wrong statistics could be inferred about the NO<sub>2</sub> concentrations in a given area from the measurement taken at one sampling site, if its proximity to e.g. a large road is not taken into account. This is a reason why the AURN sites are split into types, amongst which are urban background, urban roadside and urban industrial. Wind measurement biases can occur if the instrument is located in an area which is sheltered by nearby objects, e.g. trees or buildings. This could lead to a systematic sampling error in the measurement of e.g. wind direction, which is why the location and surroundings of WMO observation sites are carefully considered.

### 3.3 The Met Office Unified Model

This section provides a general overview of the physical and dynamic structure of the Met Office Unified Model (MetUM). The MetUM uses a ‘seamless’ approach to modelling, which means that all of its various configurations share the same dynamical core and many of the parameterisation schemes to represent sub-gridscale processes (Brown et al., 2012a). As of 2014, the current version of the dynamical core is ENDGame (Wood et al., 2014), which uses non-hydrostatic and fully compressible governing Euler equations of momentum, conservation, thermodynamics and the equation of state (Staniforth and Wood, 2003). The advection is semi-Lagrangian on an Arakawa C-grid (Arakawa and Lamb, 1977) and the time-stepping is semi-implicit.

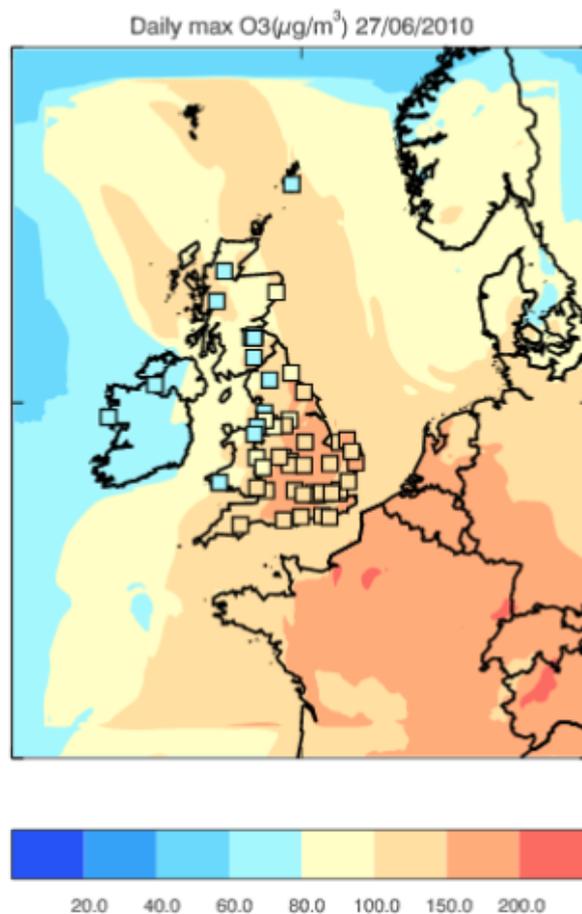
MetUM can be used as a climate model at spatial resolutions of 120 km globally or 4 - 1.5 km regionally; it is also used for numerical weather prediction (NWP) at spatial resolutions of 10 km globally or approximately 1.5 km within a limited-area regional configuration such as the UKV. It can be used for research purposes at scales of 1 km - 100 m due to the unified nature. Barriers exist to the capability of the current dynamical core: namely, increases in resolution and forecast lead-time require ever more computational power. Singularity complications also exist for a lat-lon mesh at the poles. Future developments of the dynamical core involve using a uniform mesh under the project name Gung Ho and a new dynamical core called LFRic, which continues to be developed at the time of writing (see e.g. Adams et al. (2019)).

More information about the parameterisation schemes used within the MetUM which

are relevant to this thesis will be provided in section 5.3 and sub-sections within.

### 3.4 Air Quality configuration of the Unified Model (AQUM)

AQUM is an on-line model operating in a limited area spanning the UK and a small part of continental Europe. The AQUM domain is on a “rotated pole” grid, spanning 45 - 60°N and 12°W - 12°E as shown in figure 3.3. The horizontal grid spacing is 0.11°x 0.11° (~12 km) and there are 63 vertical model levels up to 39km. The chemistry scheme in AQUM is the Regional Air Quality (RAQ) scheme, which uses the tropospheric chemistry sub-routines from the United Kingdom Chemistry and Aerosols (UKCA) project (O’Connor et al., 2014). Chemical reactions for 58 atmospheric species are represented in the RAQ scheme used within AQUM, and photolysis rates are calculated hourly with the Fast-J photolysis scheme of Wild et al. (2000). In operational mode, AQUM is initialised



**Figure 3.3:** AQUM domain, reproduced from Savage et al. (2013). The figure shows the forecast for maximum daily O<sub>3</sub> concentrations ( $\mu\text{g m}^{-3}$ ) for 27th June 2010, overplotted with squares representing observations.

once a day at 18 UTC with meteorological lateral boundary conditions (LBCs) provided by the Met Office Unified Model (MetUM) global forecast. At the domain boundaries, atmospheric pollutant species are provided by the CAMS global reanalysis from the European Centre for Medium-range Weather Forecasting Chemistry (ECMWF) at 3-hourly resolution. Within AQUM's domain, initial conditions for pollutant concentrations are taken from its preceding day's forecast at  $T+24$ , where  $T+X$  here is defined as  $X$  hours after model initialisation. Forecast output is hourly, with the  $(T+6)$  forecast valid at 00 UTC. No assimilation of observational data takes place throughout the model run, but the chemistry and meteorology are a coupled system, i.e they have a two-way feedback with a time-step of 5 minutes (Savage et al., 2013; Neal et al., 2017). This means that the chemistry and aerosols within AQUM also feedback on the meteorology, e.g with cloud condensation nuclei (CCN) which influence the forecast cloud cover. For its primary emissions of atmospheric species, AQUM uses a blend of 3 datasets. The highest resolution dataset used is the 1km UK National Atmospheric Emissions Inventory (NAEI, Jones et al., 2019b), which is an annual inventory, separated into 11 source sectors over land in the UK. NAEI also includes air-traffic emissions, but only domestic take-off and landing emissions. Also used in AQUM are the 50 km resolution emissions data from the European Monitoring and Evaluation Programme (EMEP), and for waters around the UK, the 5 km resolution shipping emissions (EntecUK; Whall et al., 2010) are implemented. Emissions are spread out vertically across the first 4 model levels (20 m, 80 m, 180 m, 320 m).

### 3.5 Forecast verification of air quality variables

In this section, standard point-based error metrics such as RMSE, forecast bias and PCC are used to evaluate the AQUM forecast accuracy for  $O_3$ ,  $NO_2$ ,  $PM_{10}$  and  $PM_{2.5}$ . Hourly forecast data for  $O_3$  and  $NO_2$  is verified against surface observations taken at urban background AURN sites. Urban background sites are of particular interest, because there exists a known negative bias in  $NO$  which contributes to a positive  $O_3$  bias at urban - but not rural - (Savage et al., 2013). There are also intricacies associated with representing areas of less homogeneity, in terms of e.g. land use and boundary layer development, which will be discussed later in this chapter. Due to the pollutants' seasonal variability, the study periods are summer (June, July, August; JJA) 2017 and winter (December, January, February; DJF) 2017. For  $PM_{2.5}$  and  $PM_{10}$ , the study period is JJA and DJF of 2015 over both urban and rural background sites - this is a different study period due

to data availability. Forecast lead time ranges between T+6 to T+30, as the forecast is initialised once a day at 18 UTC, with the first valid forecast time at midnight (i.e. T+6), running freely for 24 hours.

### 3.5.1 Diurnal cycles of O<sub>3</sub> and NO<sub>2</sub>

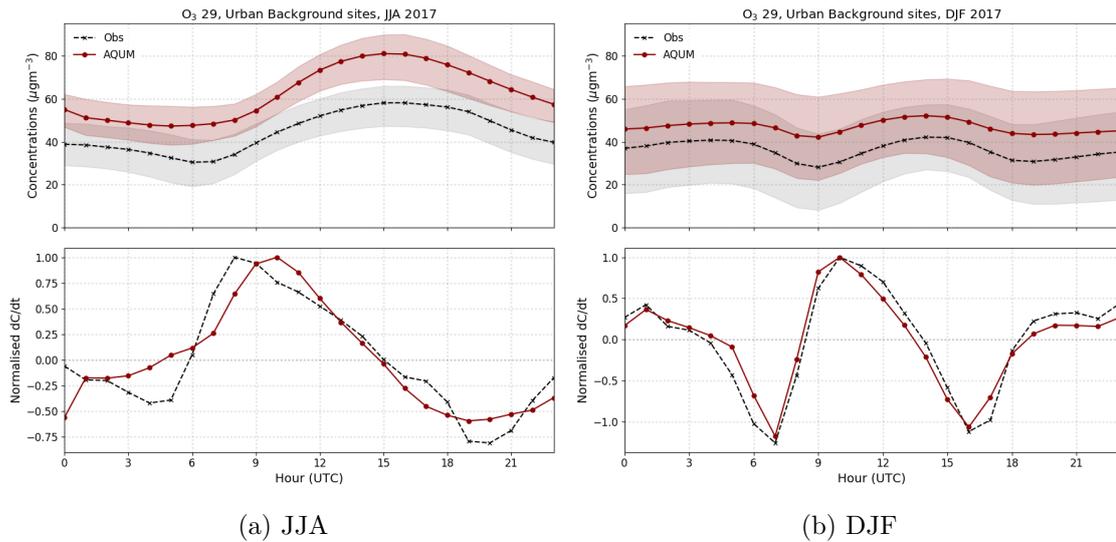
Surface concentrations of O<sub>3</sub> and NO<sub>2</sub> vary throughout the day, forming distinct diurnal cycle profiles. There exists a seasonal variability in the diurnal cycle of observed surface O<sub>3</sub> and NO<sub>2</sub> concentrations, and as such it is worthwhile to examine seasonal differences in AQUM's forecast skill.

#### 3.5.1.1 O<sub>3</sub>

Top panels of both figures 3.4 and 3.5 illustrate the diurnal profile of surface concentrations averaged over 29 urban background sites, meanwhile the bottom panels depict the normalised rate of change of the mean concentrations. The rates of change have been normalised by the maximum rate, in order to make a fair comparison between the observations and forecast.

In figure 3.4a, the observed summertime O<sub>3</sub> mean profile has a minimum at 6 - 7 UTC (30.46  $\mu\text{gm}^{-3}$ ) and a maximum at 15 - 16 UTC (58.12  $\mu\text{gm}^{-3}$ ). Concentrations are expected to be higher during the day than during the night, as photochemical production is the main source of O<sub>3</sub> in daylight hours. Throughout the night, O<sub>3</sub> is lost through titration by NO and deposition to surfaces, hence the observed profile has a downward trend between 0 - 6 UTC. The minimum (30.46  $\mu\text{gm}^{-3}$ ) occurs upon sunrise, when O<sub>3</sub> begins to be replenished by the commencement of photochemical production, as well as entrainment from the residual layer. In the winter (figure 3.4b), the morning minimum occurs later than in the summer, at 9 UTC due to a later time of sunrise. The winter morning minimum (36.21  $\mu\text{gm}^{-3}$ ) is slightly larger than in the summer, but afternoon maximum (48.29  $\mu\text{gm}^{-3}$ ) at 14 UTC is smaller than the summertime maximum at 16 UTC. Winter day-time maximum concentrations are similar to those at 5 UTC (both are  $\sim 40 \mu\text{gm}^{-3}$ ), and the overall profile shape is significantly different from the summer with a smaller diurnal range of mean observed concentrations (12.1  $\mu\text{gm}^{-3}$ ) than the summer (27.7  $\mu\text{gm}^{-3}$ ).

The regularity of the daily profiles happens because of e.g. regular emissions of NO<sub>2</sub> from road transport during the rush-hour (where differences in road vehicle traffic result in different emissions patterns between urban and rural regions, thus diurnal cycles



**Figure 3.4:** Diurnal profiles of forecast (red) and measured (black) hourly  $O_3$  concentrations, averaged over (a) 39 and (b) 41 urban background sites during JJA and DJF respectively. Shaded regions are 25th and 75th percentiles. The bottom panels show normalised rate of change of the mean concentrations.

of different amplitude); photochemical production of  $O_3$  during daylight hours, and its deposition to the ground at night. Concentrations of  $O_3$  and  $NO_2$  are co-dependent due to chemical reactions which involve both species, in addition to  $NO$  and  $VOCs$  (as discussed in section 2.1.2), thus it is appropriate to compare their accuracy side by side.

There is a systematic positive forecast bias in  $O_3$  concentrations, which ranges between 10 - 20  $\mu\text{gm}^{-3}$  in JJA (figure 3.4a) and 8 - 12  $\mu\text{gm}^{-3}$  in DJF (figure 3.4b). These results are consistent with the first AQUM verification of Savage et al. (2013), where the  $O_3$  overestimation averaged over 55 sites for the year-long period of April 2010 - April 2011, was 8.38  $\mu\text{gm}^{-3}$  with the highest seasonal bias (of around 20  $\mu\text{gm}^{-3}$ ) occurring during the summer. This bias is already known to the community and hypothesised to be caused by emissions and model lateral boundary conditions. Existing bias is therefore minimised during post-processing stage of the forecast (Neal et al., 2014) before it is published. This section will not further explore the causes of this systematic overestimation, but will explore the potential causes of the missing morning minimum and delayed increase in forecast  $O_3$  during the summer.

Let us focus on the bottom panels of figure 3.4. In the winter, forecast bias is relatively constant throughout the day because the diurnal cycle in the mean observed and forecast  $O_3$  concentrations is very similar. In the summer, the diurnal cycles in observed and forecast  $O_3$  are slightly different, and thus result in a bias with a temporal component which is not present in the winter. In the bottom panel of figure 3.4a, the

modelled maximum increase in average concentration levels during the morning lags the observations by 1 - 2 hours. This temporal offset is also present for JJA and DJF in 2015 (not shown), thus making it not an anomalous result. Furthermore, the observed minimum in  $O_3$  concentrations at 6 UTC is missing from the forecast, as indicated by the absence of a negative slope in observed concentrations between 4 - 5 UTC (bottom panel figure 3.4a). Because this discrepancy between observed and forecast concentrations is not present in the winter (figure 3.4b), it is possible that some process is missing which is more prevalent during the summer than the winter time.

The time window between 4 - 9 UTC is an important one for the changes in both  $O_3$  and  $NO_2$  concentrations, in terms of the physical and chemical changes within the surface boundary layer happening at this time. Sunrise during the summer in London is between 3:42 UTC at the earliest, and 5:10 UTC at the latest<sup>4</sup>. During the winter, sunrise occurs between 6:46 UTC and 8:47 UTC. Due to the onset of solar radiation, photochemical  $O_3$  production begins around this time and increases with the amount of sunlight. However, mixing of air from the residual layer as the boundary layer grows also contributes to this increase - indeed, figure 3.4a confirms that the observed rate of change becomes positive between 5 and 6 UTC, and the forecast is already positive at 5 UTC. The slope of both the observed and forecast increase in concentrations rise fast from 6 UTC. But while observed rates of increase reach a maximum by 8 UTC, the forecast rate of increase has a maximum only at 10 UTC. This two-hour difference in maximum rates of increase is not present in the DJF comparison.

During the winter, both forecast and observed concentrations begin to rise between 8 and 9 UTC. Assuming that there is no seasonally varying bias in the prediction of incoming solar radiation, the fact that AQUM predicts the correct  $O_3$  increase in DJF after sunrise suggests that the representation of  $O_3$  photochemical production is correct in AQUM. A possible explanation for the seasonal difference in the  $O_3$  morning increase is that the representation of entrainment of  $O_3$ -rich air into the boundary layer is inaccurate.

There are a few possibilities why the temporal offset is only present in one of the seasons. Firstly, boundary layer stability differs between the seasons. For example, a study into lidar and sonic anemometer classifications of boundary layer types at a rural location (Chilbolton) in the UK (Harvey et al., 2013), shows a clear diurnal and seasonal pattern in boundary layer types for the period of June 2008 - June 2011. They find that the fraction of stable-type boundary layers decreases from 0.96 in the night to 0.70

---

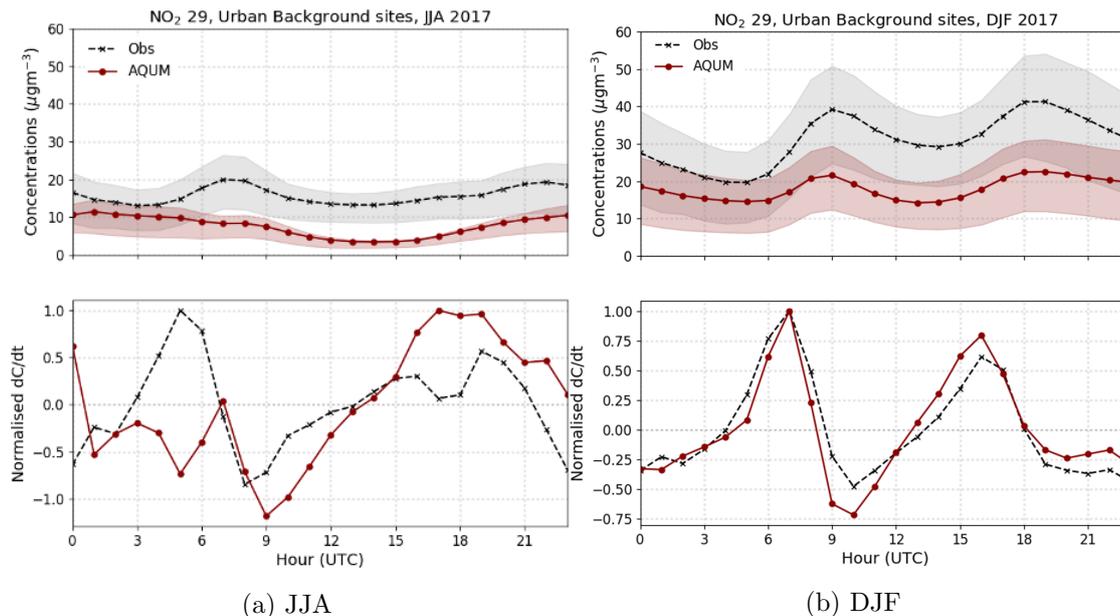
<sup>4</sup>Sunset and sunrise information from <https://www.timeanddate.com/sun/uk/london>, note use of BST

upon sunrise (DJF), and respectively from 0.90 to 0.28 in JJA. Summertime months are therefore more likely to have a neutral or unstable boundary layer shortly after sunrise than winter months. It is important for the model to accurately predict the stability of the boundary layer, as the rate of growth depends on its stability type and as such, it is a key parameter for turbulent mixing of pollutants within the boundary layer (as discussed in section 2.2). Secondly, within the boundary layer type classifications of Harvey et al. (2013), which are based on the UK Met Office scheme outlined in Lock et al. (2000), there are sub-categories based on cloud cover and type. These are important for the amount of shortwave solar radiation reaching the Earth's surface, and thus the amount of sunlight - the key ingredient for photochemical production of  $O_3$ . For example, by examining the influence of cloud error on surface  $O_3$  in the WRF-Chem model, Ryu et al. (2018) find that the cloud cover forecast accuracy is particularly important for forecasting  $O_3$  in VOC-limited regions and under-estimating cloud can contribute 40% of the total maximum daily 8-hour average bias. As such, any mis-categorisation of cloud cover and type may affect  $O_3$  production. Lastly,  $O_3$  precursors such as VOCs, CO and  $NO_x$  are necessary for the  $O_3$  chemical production cycle to occur, as discussed in section 2.1.2. It is therefore worth considering the accuracy of the  $NO_2$  forecast alongside the  $O_3$  forecast.

### 3.5.1.2 $NO_2$

Figure 3.5 is an evaluation of the same sites and study period as figure 3.4, but for  $NO_2$ . The diurnal cycle of  $NO_2$  at urban sites (predominantly traffic sites, but also background sites) is strongly influenced by morning and evening rush-hour emissions from road transport (e.g. Lee et al. (2020)). The morning and evening peaks are stronger during the winter than summer because of e.g. people's choice of transport to work and more domestic heating.

$NO_2$  concentrations generally tend to be lower than  $O_3$ , with the lowest average concentrations during the summer (figure 3.5a). In both seasons,  $NO_2$  is under-predicted in AQUM, with the largest mean bias during daylight hours. This is consistent with the evaluation of Savage et al. (2013), where it is shown that there exists a small, positive  $NO_2$  bias for rural sites, but it is not as large and less significant than the negative bias at urban sites. This could be because predicted  $NO_x$  is instantaneously spread-out over the model grid-box, meaning that predicted concentrations are dispersed away from their ground-based sources (which, due to road vehicle emissions, are generally higher in urban regions) and therefore under-predicted. Savage et al. (2013) thus expect the negative



**Figure 3.5:** Diurnal profiles of forecast (red) and measured (black) hourly  $\text{NO}_2$  concentrations, averaged over (a) 39 and (b) 40 urban background sites during JJA and DJF respectively. Shaded regions are 25th and 75th percentiles. The bottom panels show rate of change of the mean concentrations, normalised by the maximum rate occurring during the day.

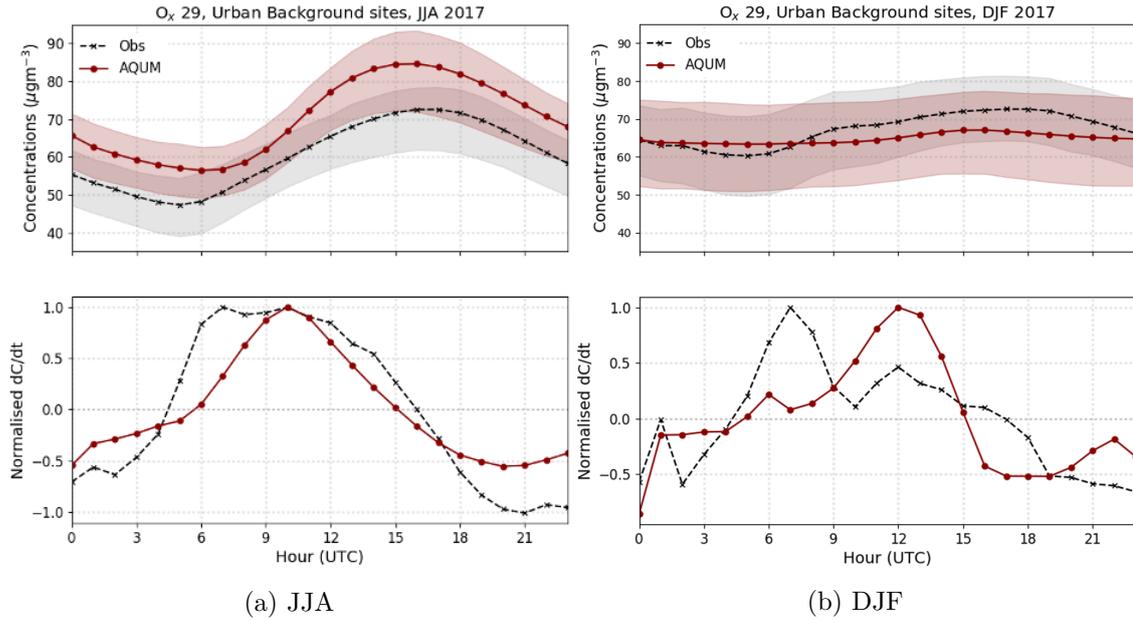
bias in NO (not shown) to contribute to the positive  $\text{O}_3$  bias at the urban background sites through under-estimating the process of titration. Moreover, studies evaluating the NAEI inventory in London, UK using an eddy-covariance technique with an instrument mounted on the BT Tower (Lee et al. 2015) and a research aircraft (Vaughan et al. 2016) both reveal that road transport  $\text{NO}_x$  emissions in NAEI are under-estimated by up to 50%. This error will propagate into the modelled  $\text{NO}_x$  concentrations and thus contribute to the  $\text{O}_3$  overestimation.

### 3.5.1.3 Total oxidant, $\text{O}_x$

Assuming that the fast reactions between  $\text{O}_3$ ,  $\text{NO}_2$  and NO are the leading mechanism for determining the night-time concentrations, then one would expect the diurnal rates of change of  $\text{O}_3$  and  $\text{NO}_2$  to be anti-correlated.

During the winter,  $r = -0.93$  for observed and  $r = -0.97$  for modelled quantities ( $p < 0.01$ ). During the summer, observed rates of change are still anti-correlated, though with a weaker relation ( $r = -0.57$ ,  $p < 0.01$ ), whereas the modelled relation is much stronger than observed ( $r = -0.83$ ,  $p < 0.01$ ).

The similarity between the rate of change of observed and modelled  $\text{O}_3$  and  $\text{NO}_2$  concentrations during the winter suggests that chemistry between the two quantities is



**Figure 3.6:** Diurnal profiles of forecast (red) and measured (black) hourly  $O_x$  ( $O_3 + NO_2$ ) concentrations, averaged over 29 urban background sites during JJA and DJF respectively. Shaded regions are 25th and 75th percentiles. The bottom panels show rate of change of the mean concentrations, normalised by the maximum rate occurring during the day.

represented well in AQUM. However, given the summertime discrepancy between modelled and observed rates of change of both  $O_3$  and  $NO_2$ , there is likely to be another mechanism contributing to the summertime concentrations which is separate from the representation of chemistry.

In order to eliminate the fast chemical reactions between  $O_3$  and  $NO_2$ , the sum of the two species is considered as the total oxidant,  $O_x$ .  $O_x$  is a conserved quantity in the photo-stationary state, i.e. in the absence of VOCs and other hydrocarbons. This is a fine assumption to make when considering urban background sites, as the measurements are representative of ambient urban concentrations (i.e. as opposed to road-side). Since the molecular masses of  $O_3$  and  $NO_2$  are similar (48 g/mol and 46 g/mol respectively), for convenience their concentrations in  $\mu\text{gm}^{-3}$  are added together without prior conversion to ppb<sup>5</sup>, as this should not lead to a gross error. A correct alternative would be to scale the  $NO_2$  concentrations by 48 / 46. Let us assume that the atmosphere is clean. Then, the observed and modelled  $O_x$  are shown in figure 3.6.

In the summer (figure 3.6a), the lag in the morning increase of  $O_x$  is evident in the bottom panel, where observed  $O_x$  concentrations start to rise at 5 UTC, followed by modelled concentrations rising 2 hours later, from 7 UTC. Mean  $O_x$  increases continuously

<sup>5</sup>Conversion from ppb to  $\mu\text{gm}^{-3}$  at 20° and 1013 mb:  $O_3$ : 1ppb = 1.996  $\mu\text{gm}^{-3}$ ,  $NO_2$ : 1ppb = 1.913  $\mu\text{gm}^{-3}$  (<http://www.apis.ac.uk/unit-conversion>)

between 6 - 10 UTC, whereas the modelled quantity reaches a maximum increase only at 10 UTC. This is consistent with the  $O_3$  evaluation in section 3.5.1.1. During the winter (figure 3.6b), the diurnal profile has less diurnal variability than during the summer, which could be explained by the stronger anti-correlation of both modelled and observed  $O_3$  and  $NO_2$ , whereby they act to ‘cancel out’ under the assumption of a photo-stationary state where the main driver of winter diurnal variability is the chemistry.

### 3.5.2 Summary of $O_3$ , $NO_2$ and $O_x$ evaluation

Using point-based metrics, it was established how well AQUM represents  $O_3$  and  $NO_2$  at their observed locations, without any additional spatial or temporal information about e.g. the model’s accuracy within a region surrounding the point observations. It was shown that  $O_x$  has a positive systematic bias and larger diurnal variability in the summer than winter, meanwhile  $NO_2$  has a negative bias and a larger diurnal variability in the winter than the summer. Strong observed and modelled anti-correlations exist between the diurnal rates of change in  $NO_2$  and  $O_3$  during the winter but not during the summer, suggesting that there is likely a physical mechanism during the summertime which contributes to the concentrations of  $O_3$  and  $NO_2$ . Lee et al. (2020) recently investigated the change in observed  $O_3$  and  $O_x$  concentrations at AURN urban background sites due to  $NO_x$  reduction during the 2020 Covid-19 lockdown period. They found that daily mean  $O_3$  and  $O_x$  concentrations increased by 11% and 3.2% respectively from the 2015 – 2019 daily mean reference, attributing this increase to the statistically significant reduction in  $NO_2$  concentrations at 89% of the urban sites (compared to the base reference period). Although Lee et al. (2020) considered the daily mean and median values, their results support the anti-correlation between  $O_3$  and  $NO_2$  found in this chapter. However, their results are slightly removed from mine, because the focus of this section was on the diurnal rates of change of the pollutants, rather than daily mean values.

In order to eliminate arguments stemming from the representation of chemistry between the two quantities, one could consider the diurnal cycle of  $O_x$  under the assumption of a photo-stationary state. As with  $O_3$ , considering the diurnal cycle of  $O_x$  confirms a lag in the modelled morning increase of concentrations by around 2 hours.

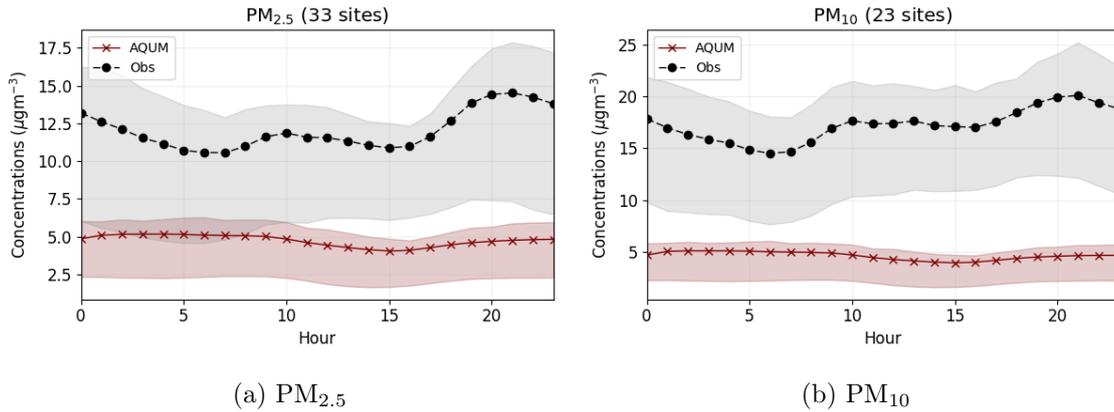
It would be useful to eliminate the chemical contributions to the surface concentrations of  $O_x$ , and explore only the physical mechanisms such as boundary layer depth, its rate of growth, entrainment, wind speed, cloud cover, and so forth. Although some of the highlighted studies in this section use a “bottom-up” approach to e.g. reveal a discrepancy

in  $\text{NO}_x$  emissions used in AQUM, a development of a “top-down” method of exploring the physical processes within the boundary layer could show whether entrainment of free tropospheric air also contributes to the surface concentrations of  $\text{O}_3$  and  $\text{NO}_2$ .

### 3.5.3 Diurnal cycles of $\text{PM}_{2.5}$ and $\text{PM}_{10}$ .

AQUM uses the CLASSIC scheme (Jones et al., 2019a) to represent all the particles which contribute to  $\text{PM}_{2.5}$  and  $\text{PM}_{10}$ . These include nitrates, sulfates, black carbon, biomass burning aerosol, organic carbon from fossil fuel (OCFF), secondary organic aerosol (SOA) and nitrate aerosols. Sea salt is not included in AQUM, as it is only present over the sea within CLASSIC and is not an advected species. The coarse component,  $\text{PM}_{10}$ , is notoriously difficult to forecast accurately, as each of the above components adds an element of possible error to the overall forecast concentrations. For example, Marécal et al. (2015) demonstrates that most (6 out of 8) of the MACC-II (Monitoring Atmospheric Composition and Climate: Interim Implementation) ensemble of models under-predict  $\text{PM}_{10}$  during DJF 2013, with a mean bias of  $-4.5\mu\text{gm}^{-3}$ . There is also a larger variability in forecast - observations correlations for  $\text{PM}_{10}$  than  $\text{O}_3$  across the models included in the comparison, due to the variable complexities in aerosol representation. Largest  $\text{PM}_{10}$  under-estimations in the models were due to missing sea-salt emissions, lack of SOA or coding errors. Diurnal cycles across the models were all similar, which was attributed to the use of a common emissions inventory containing anthropogenic black and organic carbon emissions, i.e. dominant sources of  $\text{PM}_{10}$ . However, mean bias was consistently worse during the night-time than any other time of day, which the authors link to uncertainties of the boundary layer depth. Many models in general struggle with representation of night-time boundary layers, which are usually stable (Harvey et al., 2013).

In the initial evaluation of AQUM, Savage et al. (2013) find that the mean  $\text{PM}_{10}$  forecast concentrations are as little as 50% of the observed concentrations. Figure 3.7 demonstrates a similar mean negative bias as in Savage et al. (2013) but for DJF 2015, evaluated over 33 and 23 sites for  $\text{PM}_{2.5}$  and  $\text{PM}_{10}$  respectively. The magnitude of the overall average negative bias in the  $\text{PM}_{2.5}$  diurnal profile ( $-7.26\mu\text{gm}^{-3}$ , figure 3.7a) is smaller than that of  $\text{PM}_{10}$  ( $-12.48\mu\text{gm}^{-3}$ , figure 3.7b). This is because, as discussed in Marécal et al. (2015), the coarse PM component (2.5 to  $10\mu\text{gm}^{-3}$ ) also encompasses primary particulates in the form of sea salt, transported dust and road vehicle emissions, whose exclusion from the forecast likely contributes to the negative bias. A case study



**Figure 3.7:** Mean diurnal profiles of forecast (red) and measured (black) hourly  $\text{PM}_{10}$  (left) and  $\text{PM}_{2.5}$  concentrations at urban background sites, for DJF 2015. Shaded regions are 25th and 75th percentiles.

into a high  $\text{PM}_{10}$  event which occurred on April 2011 demonstrated that when  $\text{PM}_{10}$  is dominated by secondary aerosols, the forecast bias is improved (Savage et al., 2013).

Notably, both PM components in figure 3.7 show that the mean diurnal profile of concentrations is close to the 75th percentile, i.e. the hourly distributions are heavily skewed by extreme values. For the summertime months (not shown), forecast bias for  $\text{PM}_{10}$  ( $3.83 \mu\text{gm}^{-3}$ ) is still negative but smaller than in winter, while the  $\text{PM}_{2.5}$  bias is slightly positive ( $1.13 \mu\text{gm}^{-3}$ ). As for the shape of the diurnal profile of  $\text{PM}_{10}$  and  $\text{PM}_{2.5}$ , unlike for  $\text{O}_3$  and  $\text{NO}_2$ , there is no distinct diurnal shape - neither in the winter, nor in the summer. Given that boundary layer depth dictates the vertical extent of mixing, it is worth investigating the night-time representation of boundary layer depth in AQUM. This is something we will return to in section 5.4.2.

Despite having mainly discussed the influence of emissions inventories on the modelling of both PM components, physical processes such as precipitation are efficient at removing PM from the atmosphere - this was discussed in section 2.3.3. The following section therefore evaluates AQUM's performance in forecasting precipitation, alongside other meteorological variables like 10 m wind speed and temperature.

### 3.6 Forecast verification of meteorological variables

This section uses point-based verification to assess the AQUM forecast accuracy of some of the meteorological variables key to air quality forecasting. Selected for analysis are 10 m wind speed, 1.5 m temperature and precipitation. These variables were chosen on the importance of their specific physical and / or chemical influence on the fate of

atmospheric pollutants.

Firstly, 10 m wind speed is a key parameter for transport and dispersion of pollutants, both horizontally and vertically (McNider and Pour-Biazar, 2020). Horizontal transport determines the heterogeneity of pollutant concentrations near the Earth's surface by transporting particles away from sources of emissions, meanwhile vertical transport dictates how well-mixed the pollutant concentrations are due to turbulent eddies in the atmosphere.

Secondly, 1.5 m temperature is key for chemically reactive species such as  $\text{NO}_x$  and  $\text{O}_3$  precursors, whose rates of reaction depend on the ambient temperature. The modelling study of Sillman and Samson (1995) demonstrated that a temperature difference of 5 K throughout the troposphere can reduce  $\text{O}_3$  concentrations by 6%, due to increased sinks associated with precursor chemistry. It is therefore important to evaluate the accuracy of forecast temperature in AQUM.

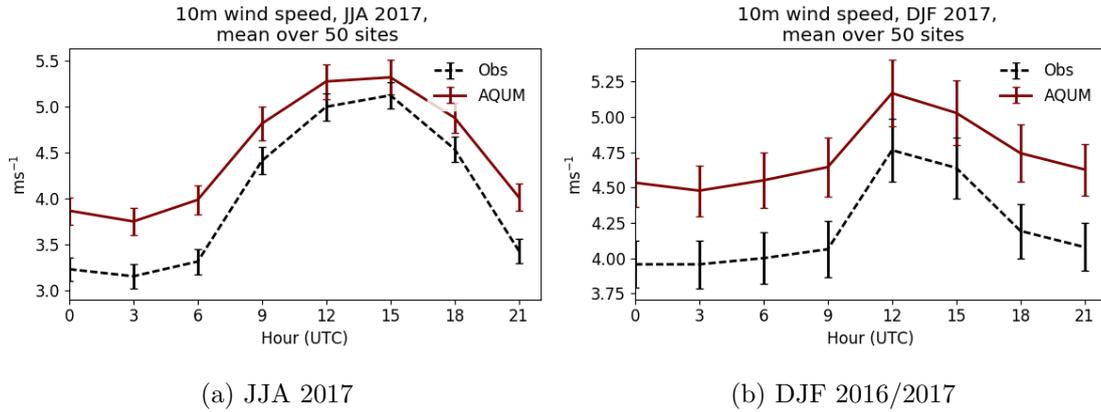
Finally, precipitation provides an effective mechanism for physically clearing the air of aerosols such as PM, via the process of washout - see section 2.3.3. Precipitation can also act to dissolve some species, although this thesis does not deal with any aqueous chemistry.

The following sections will evaluate the surface wind speed and temperature forecasts from AQUM against ground-based observations from WMO sites. Due to annual variability in both the 10 m wind speed and 1.5 m temperature, AQUM forecast of both variables is analysed over the period of 01-12-2016 to 01-03-2017 (DJF) and 01-06-2017 to 01-09-2017 (JJA). Then, section 3.6.3 will introduce and compare two different sources of rainfall observations: radar and rain gauge, followed by discussion on their suitability for verifying the precipitation forecast.

### 3.6.1 10 m wind speed

Forecast data is 3-hourly, and is verified against ground-based observations from 50 WMO sites in various locations around the UK. The sites have been selected according to their location near to an AURN site measuring  $\text{O}_3$  and  $\text{NO}_2$ , where “near” means 24 km or less, i.e. two model grid-boxes. This ensures that the number of such available pairs is large enough to gauge meaningful statistics, whilst still being representative of the general meteorological conditions.

First, let us consider the diurnal cycle of 10 m wind speed, shown in figure 3.8. The two seasons have in common larger magnitude of wind speed during daytime hours (e.g.

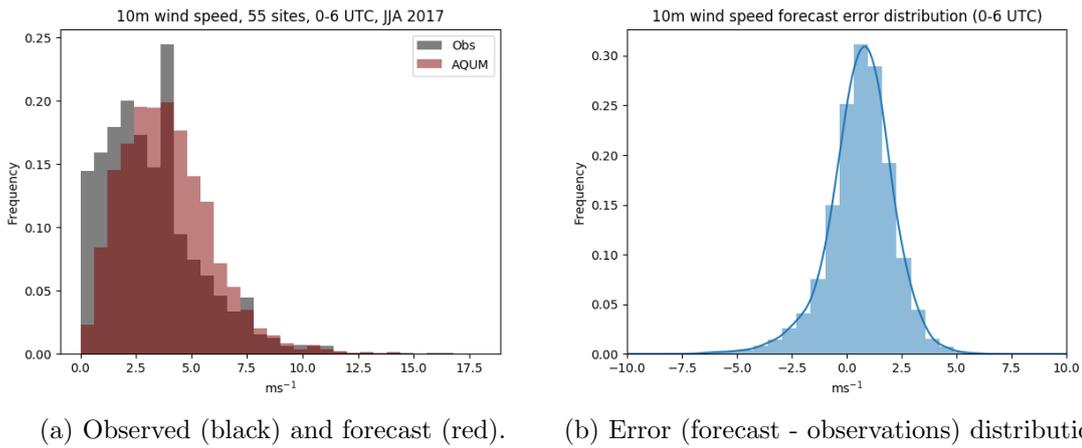


**Figure 3.8:** Mean diurnal cycle of 10 m wind speed at 50 WMO sites for JJA and DJF 2017. Error bars are 1 standard deviation at each 3-hour interval.

12 - 15 UTC) than during the night (21 - 06 UTC) at the selected sites. However, while there is a positive bias present in both seasons throughout the day, JJA has larger mean error during the night ( $0.63 \text{ ms}^{-1}$ ) than DJF ( $0.55 \text{ ms}^{-1}$ ), and the JJA bias is more pronounced during the night than during the day (figure 3.8a). This effect is enhanced when one considers the entire set of 158 WMO sites, in which the daytime error reduces to practically zero (not shown).

Further concentrating on the night-time errors in ground-level wind speed, shown in figure 3.9a is the fact that the forecast distribution is shifted towards higher values than those observed. In particular, the forecast gives fewer instances of “low” wind speeds ( $< 2.5 \text{ ms}^{-1}$ ) than observed; conversely, there are more instances of forecast wind speeds between  $2.5 - 7 \text{ ms}^{-1}$  than observed. It is also important to note that the cup anemometer sensitivity to the lowest measurable value is  $0.51 \text{ ms}^{-1}$  (1 knot). There is good overlap between observed and forecast wind speeds above  $7.5 \text{ ms}^{-1}$ . Figure 3.9b shows that the positive or negative error magnitude can be as large as  $5 \text{ ms}^{-1}$ , with most errors falling within the value range of  $0 - 1 \text{ ms}^{-1}$  during the night-time.

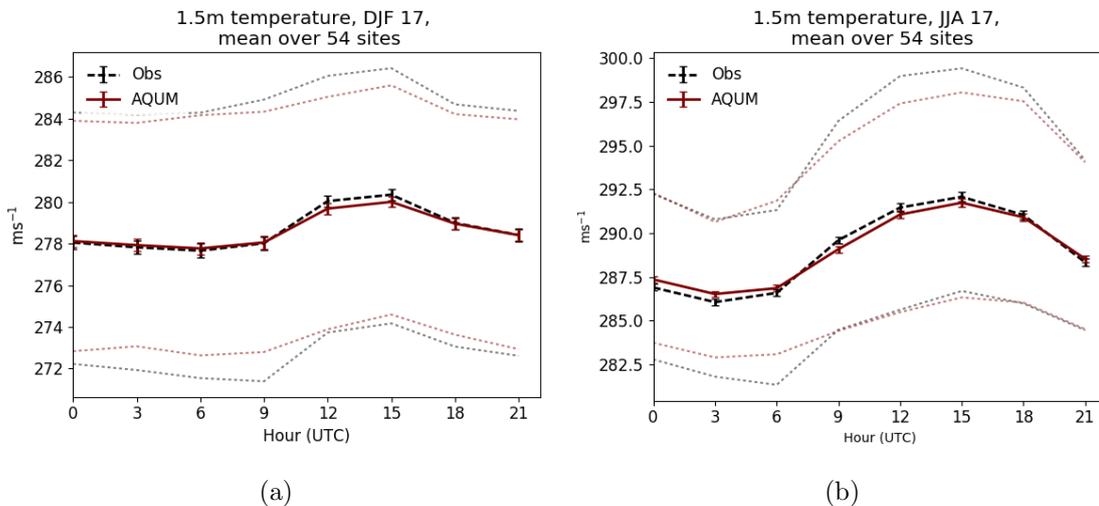
The first-day wind speed forecast was considered, i.e. T+6 to T+27. It would not be expected that the forecast accuracy would diverge significantly over this short lead-time, therefore errors are not attributable to forecast divergence with lead time. The positive night-time bias seen in AQUM is also consistent with other coupled air quality - meteorology regional models, demonstrated by an inter-comparison study of such models by Brunner et al. (2015). Over-estimating night-time wind speed could lead to dilution of pollutants near their emissions source, which will be explored in a later section.



**Figure 3.9:** 10 m wind speed observations, forecast and error distribution at 55 WMO sites, sub-sampled by night-time (0-6 UTC) in JJA 2017.

### 3.6.2 1.5m temperature

As expected, figure 3.10 shows a strong diurnal cycle in surface temperature observations at the selected urban sites, as well as the forecast. There is little discrepancy between observed and forecast values in both the winter (3.10a) and summer (3.10b). The forecast diurnal cycle profiles have marginally reduced variability than the observations in both seasons. Although there exists a negative mean forecast bias between 12 - 15 UTC in both seasons, the difference is within 1 standard deviation of the observations. The difference between extreme values is also small: there is a maximum difference of 1.5 K at 12 UTC between the mean maximum observed and forecast temperatures in



**Figure 3.10:** Mean diurnal cycle of 1.5 m temperature at 50 WMO sites for (a) DJF and (b) JJA 2017. Error bars are 1 standard deviation at each 3-hour interval, and dashed lines are min / max values.

JJA, and 1 K for DJF also at 12 UTC. Because the mean error and differences in extreme values are small, they are unlikely to have any significant influence on forecasting the air pollutants of interest to this study.

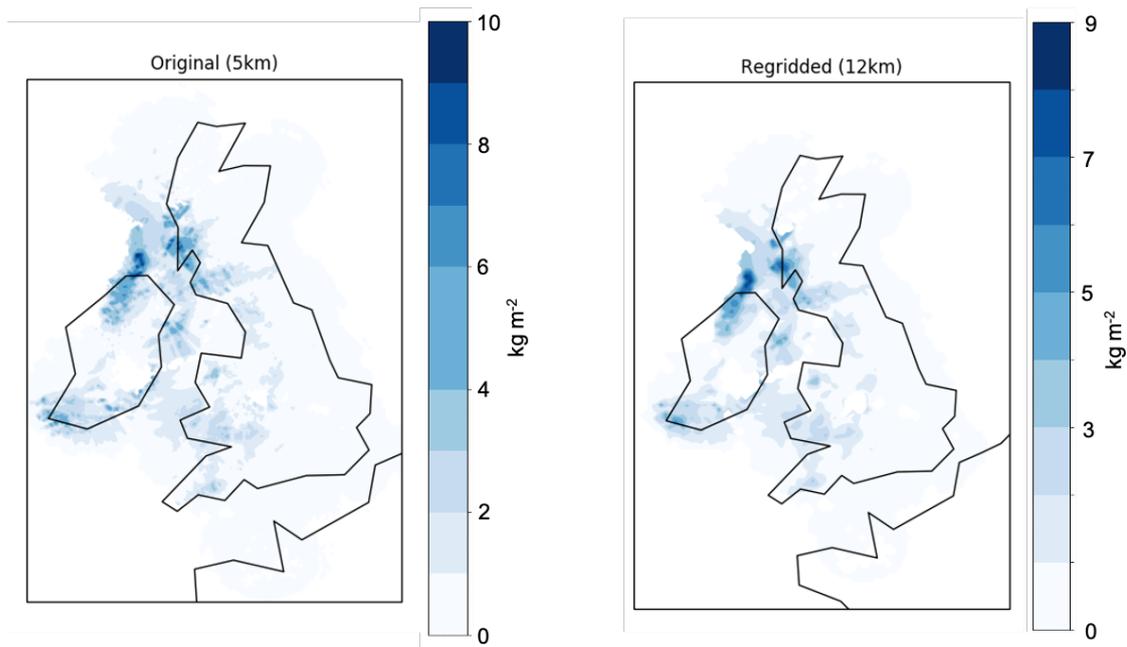
An alternative analysis could be to evaluate temperature anomalies relative to a background reference, e.g. the ERA interim reanalysis. This would enable better understanding of how the forecast performs relative to a background ‘truth’ as opposed to a point measurement, which would also eliminate any errors from extreme values measured at the point location.

### 3.6.3 Precipitation

The following section presents a verification of the hourly AQUM precipitation forecast against radar and rain gauge observations, for the same study period as in the previous section. The forecast is compared first against ground-based observations from MIDAS, then against gridded rainfall observations from the radar analyses, as described in section 3.2.2. Radar data was coarse-grained from its native 5 km to AQUM’s 12 km resolution by using a simple linear interpolation, in order for it to be spatially representative of the forecast. The resultant smoothing of the radar data diminishes its peak values, as presented in figure 3.11.

Furthermore, in order to ensure that the precipitation forecast is verified against a suitable set of observations, radar and rain gauge data are compared with one another in section 3.6.3.3. Both radar and rain gauge observations come with their own set of advantages and disadvantages when considered as “truth” for precipitation forecast verification. For example, radar analyses have been shown to have uncertainty of as high as a factor of two (Joss and Waldvogel, 1990), and instruments are susceptible to ground clutter, low-level orography and calibration issues, among other factors which may influence the recorded measurement. Even though the Met Office Nimrod radar composite system applies quality control and error correction techniques to the soundings, there will nonetheless exist measurement uncertainty and unfiltered noise.

Rain gauge observations provide estimations of ground-level precipitation at point locations and are often used alongside radar soundings to produce quality-controlled radar analyses. Rain gauge observations are also susceptible to measurement uncertainty, e.g. errors in the timing of observation, missing rain during the period of a rain bucket tipping, unfavourable exposure to strong wind (Met Office, 2020b). Because the rain gauge sites are spatially sparse and distributed unevenly, there are some areas where the distance



**Figure 3.11:** A snapshot of the gridded radar at 01/01/2015 15:00 UTC. (a) is the original 5 km resolution; (b) is the coarse-grained 12 km data. Note change in scale of colourbar.

between sites is larger than the grid-box length of some mesoscale models. This means that rain gauges will not always be representative of an entire model grid-box and as such, a “truth” analysis with a reliable spatial coverage – such as a radar-rainfall estimate – is more suited for model verification, accounting for measurement uncertainties (Mittermaier, 2008). Radar soundings are more likely to be accurate when the observed rainfall rate is high, as the signal-to-noise ratio is higher than when precipitation is near to zero and the radar pulse is too small to be reflected.

Hourly precipitation forecast ( $\text{mmh}^{-1}$ ) from AQUM distinguishes convective and large-scale precipitation, alongside information on whether it fell in the form of rain or snow. For the following analysis, the four sub-categories of precipitation were concatenated together as total precipitation. The forecast lead time is T+6 to T+30, i.e. the first “full” forecast day from 00:00 to 00:00 the next day, due to the model’s initialisation time at 18:00 UTC the day before. The forecast data is in the form of grid-box averages, where a bi-linear interpolation can be performed onto a chosen point location within a grid-box (from the 4 neighbouring grid-box values). These locations are co-located with the ground-based rain gauge observations.

### 3.6.3.1 Description of datasets

The MIDAS rain gauge dataset comprises 127 locations. The radar analyses have been linearly interpolated to the set of MIDAS rain gauge locations, in order to directly compare both observations' estimates. In addition, the data has been sub-sampled by close proximity to AQ sites measuring PM concentrations. This is because we are interested in how the forecast verifies against both radar and rain gauge observations at locations of interest involving further work with PM. The maximum distance threshold was set to 24 km, i.e two grid-boxes of the AQUM forecast. Any smaller than that and the sample becomes very small; any larger than this can compromise the realistic representativeness of the rainfall estimate at the AQ site. This gives us a data sample size of 20 sites (JF 2015)<sup>6</sup> and 23 sites (JJA 2015) out of 127 gauge locations. The year 2015 was selected to match the PM study period.

Mean error (forecast - observation) when the winter forecast is verified against rain gauge and radar data is  $-0.20 \text{ mmh}^{-1}$  and  $-0.19 \text{ mmh}^{-1}$  respectively. During the summer, rain gauge verification gives mean forecast error of  $-0.06 \text{ mmh}^{-1}$ , while radar verification gives  $-0.09 \text{ mmh}^{-1}$ . Although these values do all indicate a negative forecast bias, the standard deviation is larger than the magnitude of the mean error. This makes the traditional calculation of the mean error unsuitable.

It is important to note that the precipitation rates are very low, and both the observations and forecast are dominated by zero values (table 3.1): 74.4% of the hourly forecast at 127 sites is zero-valued. During the summer, 67.7% of the forecast data is zero-valued. For rain gauge observations, the amount of zero-valued data is also high: 77.8% in winter, 85.3% in the summer. Radar observations register more rain than the gauges, as the proportion of no-rain events is 56.7% in winter, 65.5% in the summer. At the sub-sampled sites, percentages of no-rain events are even higher. The high frequency of non-events renders the calculation of standard descriptive statistics (i.e mean, standard deviation, variance etc) unsuitable.

Often, precipitation statistics are calculated over a coarser temporal frequency, i.e. 6h or daily accumulations. For example, the Met Office provide publicly available 'climate summaries' and historical time-series, where consulting a daily or monthly precipitation accumulation is appropriate. This kind of analysis avoids the high number non-events present in hourly datasets, making the statistics more robust. On the other hand, this project deals with detailed evaluations of the hourly precipitation forecast, with the inten-

---

<sup>6</sup>December 2014 data was not available for inclusion in the the winter 2015 season dataset.

**Table 3.1:** Percentage of hourly “no-rain” events across 127 MIDAS sites (values in brackets show the percentages at sub-sampled sites), i.e. precipitation rate = 0.0 mmh<sup>-1</sup>

%	Gauge	Radar	Forecast
JF	77.8 (81.8)	56.7 (59.9)	74.4 (78.4)
JJA	85.3 (87.7)	65.5 (68.6)	67.7 (70.7)

tion of linking it to immediate effects on particulate matter. It is therefore not appropriate to substantially coarse-grain the data in time and lose the details<sup>7</sup>. Instead, it is helpful to treat precipitation as a discrete or even binary quantity, for which categorical statistics can be used. These are described in the following section.

### 3.6.3.2 Categorical statistics

Although continuous in nature, precipitation is often treated as a discrete field within forecast verification. This means that thresholds are applied to evaluate whether the predicted precipitation amount falls within the observed category, e.g. < 1 mmh<sup>-1</sup>. As well as examining the difference in observed and forecast rainfall rate per unit time, it can be useful to ask whether the observed rainfall was at all forecast, regardless of its rate. For example, the general public is more likely to consult the forecast in order to decide whether they will need an umbrella today, than to know how much rain exactly it might protect them from. Therefore, forecasters often turn to a contingency table (e.g. table 3.2) to glean more insight about the nature of the precipitation forecast other than just the magnitude of forecast error. Various statistics using any combination of the possible forecast and event pairs can be calculated, such as forecast frequency bias and accuracy.

Frequency bias is a widely used error metric for categorical forecasts, directly comparing the frequency of forecast “events” to observed ones:

$$\text{Frequency bias} = \frac{A + B}{A + C} = \frac{\text{hits} + \text{false alarms}}{\text{hits} + \text{misses}} \quad (3.1)$$

where 1 is the perfect score, while values > 1 and < 1 signify over- and under-forecasting respectively. The accuracy score answers what proportion of forecasts were correct, but is skewed by rare events. The values range from 0 to 1, where 1 is the perfect score. It is calculated as:

$$\text{Accuracy} = \frac{A + D}{A + B + C + D} = \frac{\text{hits} + \text{correct negatives}}{\text{total}} \quad (3.2)$$

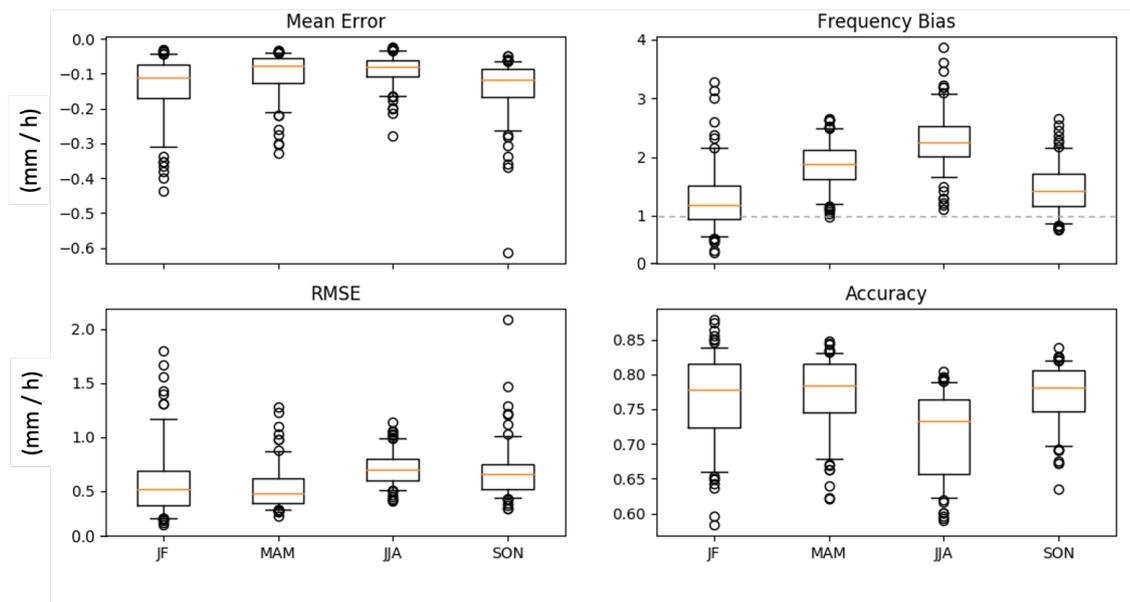
<sup>7</sup>However, in section 3.6.3.3, precipitation forecast and observations are accumulated into 6h windows, which allows some flexibility in timing offsets at point locations.

**Table 3.2:** Contingency table used in verification of categorical variables, such as non-rain events. Statistics are calculated from combinations of: A, B, C, D where A = hits, B = false alarms, C = misses and D = correct negatives.

	Observed	
	yes	no
Forecast	yes	A B
	no	C D

Figure 3.12 presents seasonal partitions of standard error metrics for AQUM forecast of precipitation compared against rain gauge observations. Mean error and RMSE are continuous statistics. A value of 0 represents a perfect forecast in both the mean error and RMSE, but the nature of RMSE is that it is skewed by outliers. In figure 3.12, one can infer that there is little difference between the seasons in the model’s performance when rain gauge observations are used for verification. The mean error is marginally smaller for JJA, though the RMSE is highest at this time of year. This can happen because outliers of a large magnitude skew the distribution due to the squared term.

For the categorical statistics in figure 3.12, precipitation was treated as a binary quantity, i.e. it was partitioned into 2 groups: “no rain” and “rain”. One can infer from the frequency bias that forecast precipitation in JJA is, on average, almost twice as likely to be a ‘false alarm’ (i.e. predicted to happen when it is dry) than in JF. The accuracy statistic shows that around 72% of the JJA and 77% of the JF forecasts were correct, although the extremely high count of “non-events” skews this value, as discussed



**Figure 3.12:** Continuous (mean error, root mean square error) and categorical (frequency bias, accuracy) statistics for hourly AQUM rainfall forecast verified against rain gauge data at 127 sites, for seasonal partitions in 2015.

in section 3.6.3.1 and shown in table 3.1. Since the hourly forecast percentage of no-rain events (74.4%) was closer to the rain gauge observations (77.8%) than radar (56.7%), a naive conclusion might be that verifying the winter forecast against rain gauge would yield a higher accuracy rate than verifying against radar. In contrast, verifying the summer forecast (68% of no-rain events) against radar (66%) would appear more accurate than verifying against rain gauge (85%).

However, it is more useful to also calculate how many non-zero events were correctly forecast when verified against rain gauge and radar. For the winter, verifying against rain gauge observations yields 81.7% correctly forecast no-rain hours, and against radar this is 89.9%. For the summer, those numbers are 72.7% and 82.1% for rain gauge and radar observations respectively. Firstly, this suggests that the model is better at identifying no-rain events during the winter than the summer. Secondly, verifying no-rain events against radar gives the forecast a better ‘hit’ rate, than if it were verified against rain gauge observations. This could be explained by measurement uncertainties related to raindrop size, e.g if the radar pulse is not well reflected by very small drops.

A reason for the seasonal differences seen in the above error metrics could be due to the distribution of observed rain type: convective precipitation dominates over large-scale fronts during the summer in the UK (e.g de Leeuw et al., 2015), whereas winter months will experience more large-scale synoptic fronts. This means that precipitation during the summer is likely to be more localised than in winter, posing representativeness issues for rain gauges which may not capture the localised convective rainfall.

It can be concluded that in a continuous sense, precipitation is usually under-forecast in AQUM, with the largest mean error values during the autumn / winter seasons. On a categorical basis however, there are twice as many forecast rain events than observed during the summer, with the forecast event frequency closest to the truth during winter.

### 3.6.3.3 Differences between radar and rain gauge precipitation

The expectation is that a gridded precipitation forecast is better represented by radar analyses than rain gauge data, in the sense that both the forecast and radar represent a grid-area average rather than a point measurement. In order to test this, radar and rain gauge observations are compared first against each other, and then against AQUM during both JF and JJA 2015. This is done in two ways:

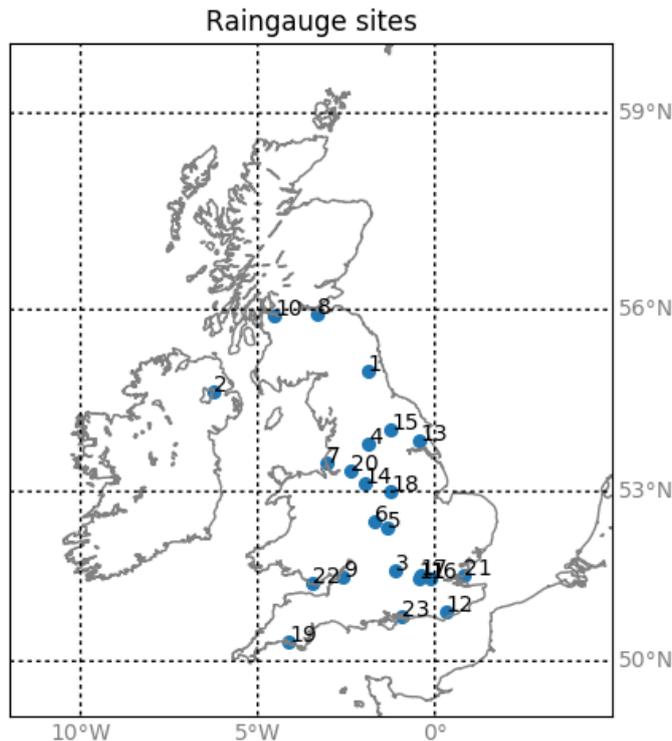
- Calculate the correlation coefficient ( $r$ ) between radar and rain gauge observations at each site;

- Bin the precipitation data into 6h windows and calculate the fraction of “zero-rain” events across the sites.

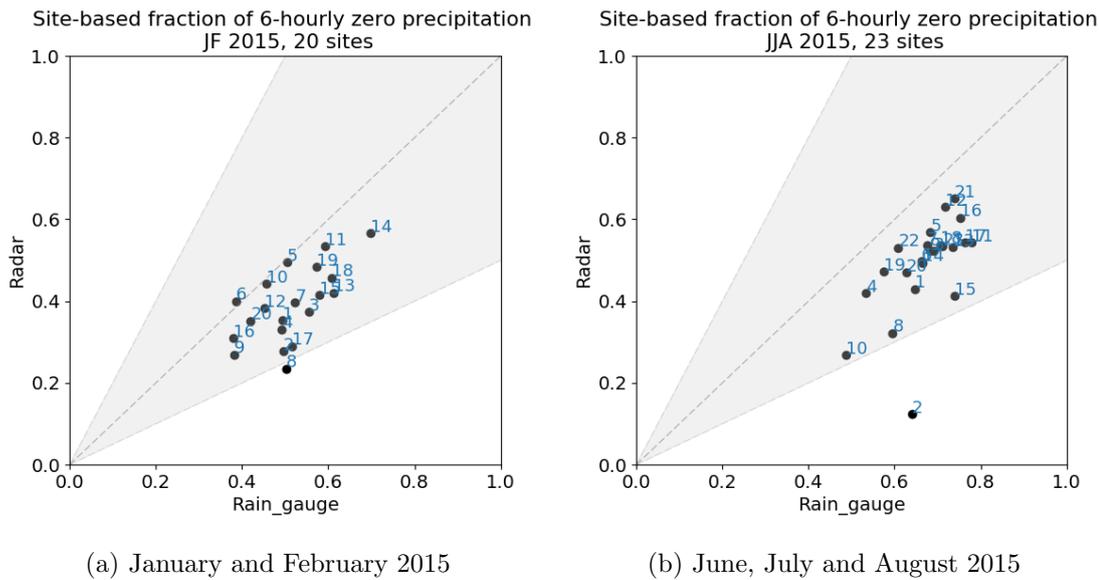
Upon calculating the site-based correlations between radar and rain gauge hourly precipitation at the 23 sites shown in figure 3.13, it was found that the average correlation value  $r = 0.617$  during the winter, and  $r = 0.586$  during the summer. This suggests that during the summer, there is a larger discrepancy between the two measurement types than during the winter, although the seasonal difference is not large.

Because it can be difficult to forecast precipitation at exactly the right location at the right time, data is binned into 6h accumulations in order to avoid the double-penalty error associated with point-based evaluation metrics. This also ensures an equal comparison between the radar and rain gauge measurements, as it is normalised to the amount of data available and is not penalised by missing data.

Figure 3.14 shows the site-based fraction of zero-valued 6h windows in the rain gauge and radar data, for winter (3.14a) and summer (3.14b). Each scatter point represents the observed fraction of zero events at an individual location, as observed by rain gauge and radar. A value of 0.3 implies that at the given site, rain was recorded over 70% of the season; a value of 0.5 means that it rained 50% of the time. 20 locations were evaluated



**Figure 3.13:** Locations of raingauge sites and IDs matched to nearest PM measurement site, for JF 2015.

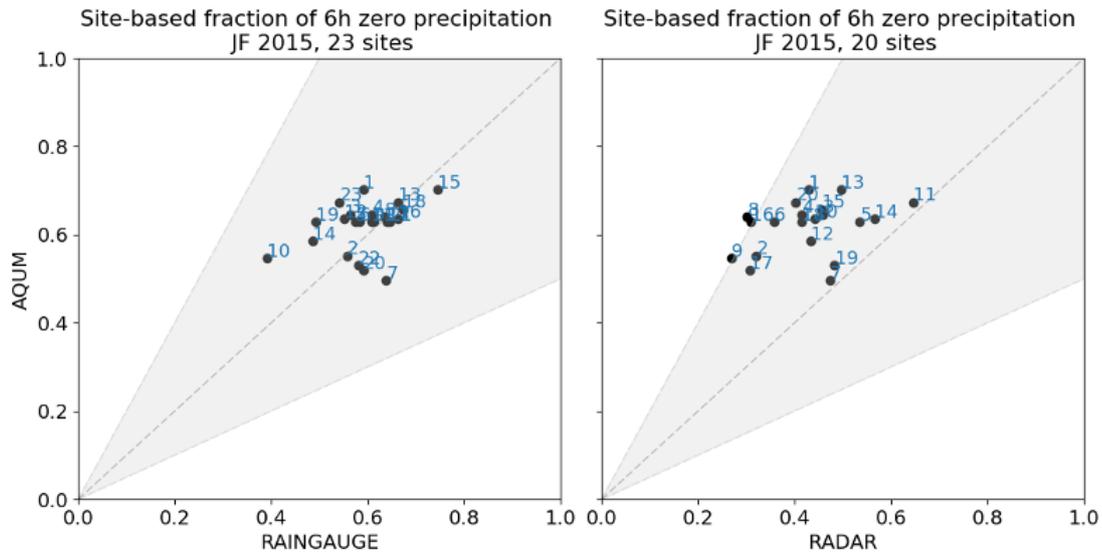


**Figure 3.14:** Site-based comparison of proportion of non-precipitating 6-hour windows within rain gauge and radar observations. Shading represents a factor of 2 out, while numbers on scatter plots are site IDs (alphabetical).

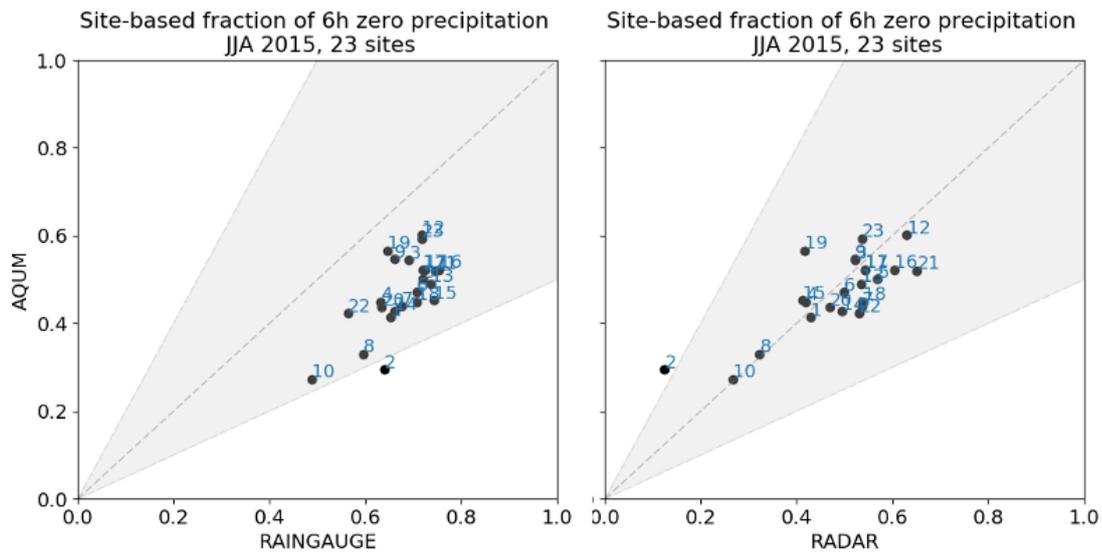
in the winter, 23 in the summer. Numbers next to the scatter points are the site IDs, as shown in figure 3.13. Note that site IDs are different between the winter and summer due to data availability.

In the winter, the fraction of zero events observed by radar and rain gauge instruments at all sites (but one, #8 - Edinburgh, located at  $55.933^\circ$  latitude,  $-3.33^\circ$  longitude) are within a factor of 2, but in general, more no-rain events are recorded by the rain gauges than with radar. The same can be said about the summer in figure 3.14b, with one site registering significantly more no-rain events with rain gauges than radar (#2, Aldergrove, located at  $54.667^\circ$  latitude,  $-6.233^\circ$  longitude). This suggests that in order to compare the rain gauge and radar precipitation, a minimum threshold may be needed. The minimum amount of rain registered by the gauges is 0.2 mm/h, therefore it would be appropriate to apply this threshold to radar. It also means that verifying the precipitation forecast against radar or rain gauge observations may give a different frequency of zero-precipitation events, which may have an effect on the perceived relationship with PM forecast error.

A similar comparison can be made of the two observation types against the forecast data. Figure 3.15 shows the observed fraction of no-rain events on the x-axis, lined up against the forecast fraction on the y-axis for winter (figure 3.15a) and for summer (3.15b). Left panels of both sub-figures are the rain gauge estimates, while the right panels show the radar.



(a) January and February 2015



(b) June, July and August 2015

**Figure 3.15:** Site-based comparison of proportion of non-precipitating 6-hour windows within observed (rain-gauge and radar) and forecast fields. Shading represents a factor of 2 out, while numbers on scatter plots are site IDs (alphabetical).

In figure 3.15a, the forecast fraction of “no rain” for all the sites was within a factor of 2 when compared against the rain gauge fraction, both spanning 0.4 to 0.8 (figure 3.15a, left). When the forecast is verified against radar estimates, figure 3.15a implies that the forecast would appear less accurate because fewer “no rain” events are recorded with radar soundings than are forecast. The fraction of “no rain” events from the radar analyses is between 0.25 to 0.7. Some sites (sites 8, 9 and 16) are hovering around the 2x factor line, which means that rain events were forecast half as much as they were

observed. This is important, because if there was a particularly large (or small) bias in the PM forecast at those sites, one may firstly begin making links with the accuracy of the precipitation forecast at those sites, and secondly start to think about which “truth” was used for the verification.

In the summer time, the overall collection of scatter points shifts along the x-axis for both radar and rain gauge precipitation, relative to winter. This means that radar is now the better representation of “no rain” events, with most (but one) sites’ forecast falling within a factor of 2 of the radar observations. The anomalous site is ID #2, which also happens to be the anomalous site for rain gauge observation, but forecasting not enough “no rain” events in relation to the gauge observations. The anomalous readings could be influenced by the site’s location near to both the Belfast International Airport and Lough Neagh: a lake with surface area measuring over 300 km<sup>2</sup> which could influence the formation of local convection.

### 3.6.3.4 Conclusions

Although precipitation does not have a diurnal cycle, it does vary seasonally and locally. Its sensitivity to orography can be problematic for relatively coarse models – such as AQUM – which are prone to under-resolve orographic features. This can result in forecasts which are not accurate when compared against “truth”. What data is used as “truth” and the evaluation statistics used can also influence the perception of how close the forecast was to reality. It was found that rain gauge and radar readings can sometimes provide similar versions of “truth”; other times they differ greatly. Previous work of [Song et al. \(2017\)](#) indicates that differences in accuracy of radar and gauges could be a result of relative humidity (RH): when RH is low, radar has been shown to over-estimate rainfall. RH is generally low when no rain is present, therefore the conclusion drawn from figure 3.15 - that radar registers fewer non-precipitating instances than rain-gauges - are consistent with [Song et al. \(2017\)](#). Another study of [Biggs and Atkinson \(2011\)](#) demonstrates that gauge and gauge-corrected radar provided a better representation of an extreme hydrological event than radar, although the latter is better at representing rainfall in mountainous terrain. Bearing in mind that the radar analyses used in this evaluation do also encompass rain gauge observations during quality control, they are most likely to be useful for verification methods where spatial fields are concerned, e.g. the Fractions Skill Score (FSS; [Roberts and Lean, 2008](#)) or “Structure, Amplitude, Location” (SAL; [Wernli et al., 2008](#)). For point-based verification, rain gauge readings

provide the most accurate truth at the ground level, although the network is sparse and the instruments can provide unrealistic result if e.g. the tipping bucket gauge is blocked by snow. For summer months, the distinction between the two observation types matters more than during winter, likely due to the prevalence of localised, convective-scale precipitation which does not occur as often during winter. For the rest of this chapter where precipitation is concerned, both measurement types will be considered.

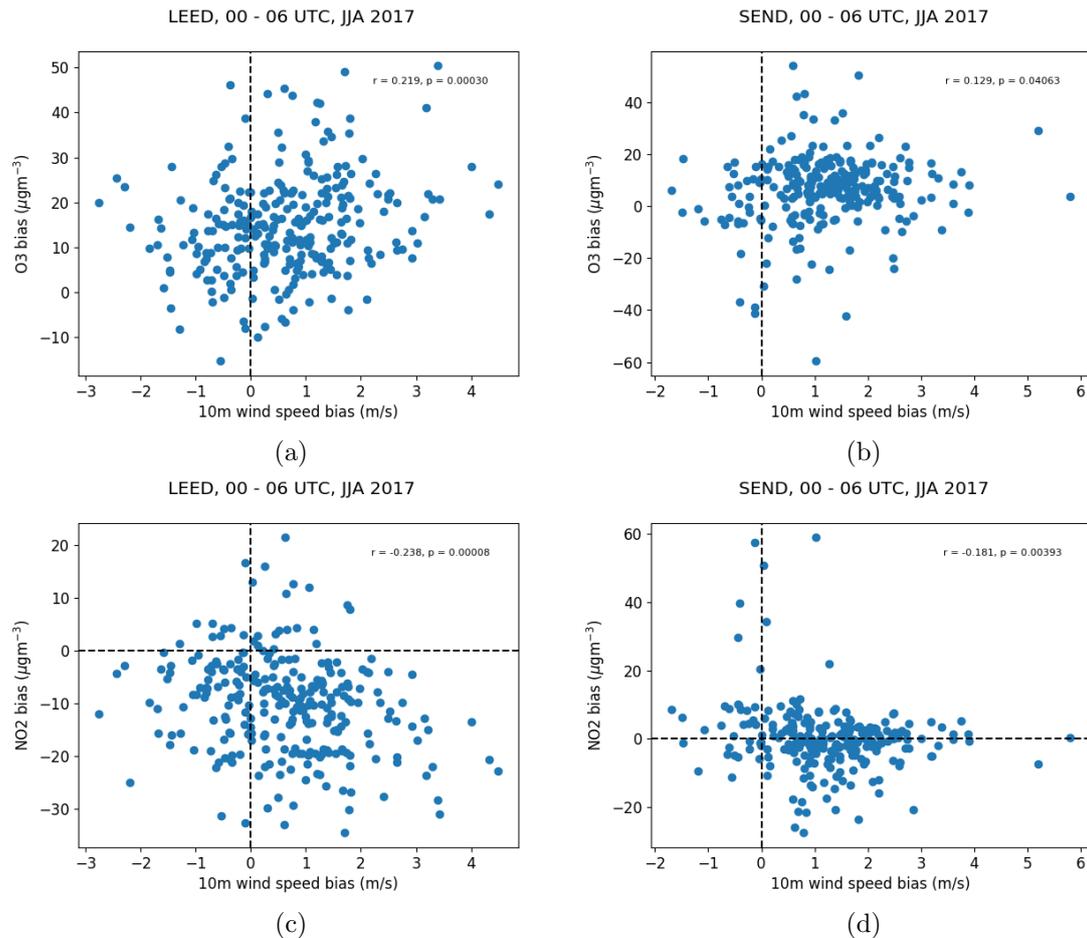
## 3.7 Relationships between air quality and meteorological forecast errors

### 3.7.1 10 m wind vs $O_3$ and $NO_2$

An evaluation of the diurnal profiles of seasonal 10 m wind speeds, as well as their forecast error distributions, revealed a night-time bias in the forecast (section 3.6.1). Using point-based correlations, forecast errors in 10 m wind speed are going to be tested for correlations with  $O_3$  and  $NO_2$  forecast error and lag discussed in section 3.5.1. Since it was shown that the 1.5 m temperature forecast is close to observations, thus errors are too small to influence the pollutant forecast, only the effect of over-forecasting 10 m wind speed on the pollutants will be considered. In light of results from section 3.6.1, this section aims to test the following hypothesis:

**Over-estimating night-time 10 m wind speed increases the negative forecast bias in  $NO_2$  and positive bias in  $O_3$ .**

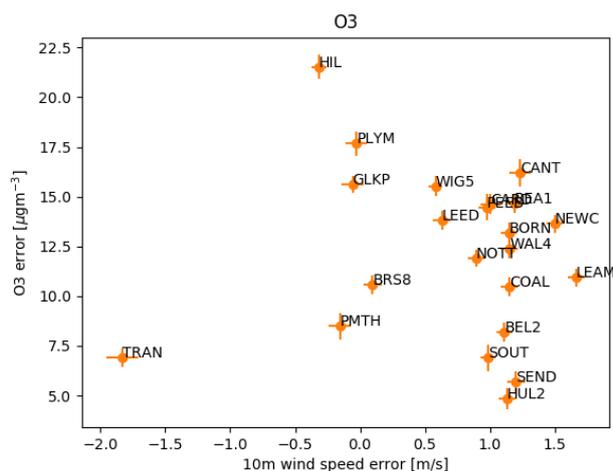
Firstly, the scatter of 10 m wind speed error against the  $O_3$  and  $NO_2$  error at individual sites is explored. Two sites have been chosen due to their distinct locations: Leeds Centre ('LEED') which is a large, landlocked city in the north of England (53.804° latitude, -1.546° longitude) and Southend-on-Sea ('SEND'), a smaller town on the south coast of England (51.544° latitude, 0.678° longitude). Because of the different nature of their locations, air quality and meteorology at these two measuring sites are expected to be different as a direct result of their surroundings, thus the two sites can be used as case studies for variability in relationships between air quality and meteorology within different surroundings. For example, the Leeds Centre site has a larger  $O_3$  and  $NO_2$  forecast error than the Southend-on-Sea site, irrespective of the magnitude or sign of the 10 m wind speed forecast error. This is shown in figure 3.16. However, it is difficult to assess the strength of any existing relation between 10 m wind speed and  $O_3$  or  $NO_2$  error at each individual site, as the variability of forecast error in both components is very large.



**Figure 3.16:** Scatter plots of night-time 10 m wind speed bias vs. O<sub>3</sub> or NO<sub>2</sub> forecast bias at Leeds Centre (a,c respectively) and Southend-On-Sea (b,d respectively) as example sites.

For example, considering PCC of 10 m wind speed error and O<sub>3</sub> or NO<sub>2</sub> error in figure 3.16, it is generally of the expected sign but small magnitude, although p-values are small ( $p < 0.05$  in all cases) so the result is statistically significant to 95% confidence level. For the sites of Leeds Centre and Southend-on-Sea,  $r = 0.219$  and  $0.129$  (for O<sub>3</sub> error vs wind speed error) respectively, and  $r = -0.238$  and  $-0.181$  (for NO<sub>2</sub> error vs wind speed error). It is clear from figure 3.16 that though the expected relationship between 10 m wind speed and O<sub>3</sub> and NO<sub>2</sub> forecast error exists, the large amount of scatter reveals that it is weak.

A number of different factors unique to each site can affect the O<sub>3</sub> and NO<sub>2</sub> errors, e.g. proximity to strong emissions sources, coastal areas or mountains. This means that there is a large variability in the magnitude of pollutant forecast errors across the sites in the dataset and as such, it is more useful to examine all the sites' aggregated data to identify the existence of a relationship, rather than evaluating them individually. Furthermore,

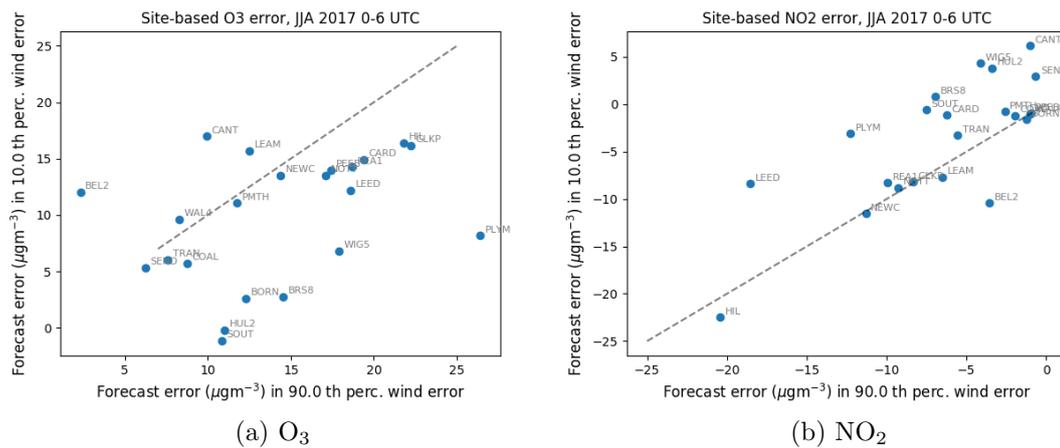


**Figure 3.17:** Site-based relations between average 10 m wind speed forecast error and O<sub>3</sub> forecast error, JJA 2017. Error bars are standard error on the mean.

a bias-correction method is needed to minimise the effect of any forecast bias (unique to each site) on the perceived relationship. Figure 3.17 shows the average wind speed forecast error and O<sub>3</sub> forecast error aggregated over all the sites, (i.e. calculating the mean over the scatter data as in figure 3.16 for each site). It is difficult to identify any kind of linear relationship between the 10m wind speed forecast error and the O<sub>3</sub> forecast error (similarly for NO<sub>2</sub>, not shown) in figure 3.17 because no bias-correction was applied to this data. One method of bias-correcting the data is by removing each site's mean bias from each data point in order to account for any systematic bias individual to each location. However, this does not make the relationship appear any stronger (not shown).

Another bias-correction method is to take the two tails of each site's forecast error distribution and compare them to each other. This method bypasses the problem of inter-site variability of the local aspects discussed above. By sub-sampling the wind forecast error distribution by the lowest 10th and highest 90th percentiles at each site, one can determine whether the corresponding O<sub>3</sub> and NO<sub>2</sub> mean error is of different sign or magnitude from the full dataset.

Figure 3.18 demonstrates the correlations between pollutant forecast errors during the lowest and highest wind error percentiles, focusing on night-time only (i.e. 0 - 6 UTC). Firstly, it is clear that pollutant error magnitude is strongly dependent on the site itself - perhaps its location, local emissions or topography - rather than whether 10 m wind speeds were strongly over- or under-predicted. For example, O<sub>3</sub> forecast error is generally larger at site 'GLKP' than 'SEND', and NO<sub>2</sub> mean error magnitude is larger at site 'HIL' than 'SEND', likely due to differences in local NO emissions. The meteorology of each



**Figure 3.18:** Site-based relations between mean forecast and observed  $\text{O}_3$  and  $\text{NO}_2$  during under-predicted (“lowest 10th percentile”) and over-predicted (“highest 90th percentile”) 10 m wind speed events during JJA 2017, sub-sampled by 0 - 6 UTC. Dashed line is one-to-one.

individual site is thus a second-order factor to determining pollutant error magnitude.

In general, the points lie close to the 1:1 line in figure 3.18b, suggesting a site-dependent relationship between the 10m wind speed error and  $\text{NO}_2$  forecast error. For  $\text{O}_3$ , there is more scatter. However, there are more sites underneath the one-to-one line in figure 3.18a than above it, meaning that within the sub-sample of large and positive 10 m wind speed bias,  $\text{O}_3$  also tends to have a larger positive bias than usual. Another way of thinking about this is that the mean  $\text{O}_3$  over-estimation is smallest when 10 m wind speeds are under-predicted. The same but opposite conclusion can be drawn from the  $\text{NO}_2$  profiles in figure 3.18b: the largest wind speed over-estimations are correlated with greater  $\text{NO}_2$  under-estimations than usual (note that  $\text{NO}_2$  already has a mean negative mean bias). And vice versa: small wind speed forecast error correlates with smaller  $\text{NO}_2$  forecast error. This is consistent with the notion that because  $\text{NO}_2$  is a primary species, it is likely to be affected by too much dispersion near sources (e.g. cities).

### 3.7.2 Precipitation vs $\text{PM}_{2.5}$ and $\text{PM}_{10}$ .

We have already discussed the efficiency of washout to remove PM from the air (section 2.3.3), evaluated the precipitation forecast (section 3.6.3) and PM forecast (section 3.5.3). The precipitation evaluation demonstrated that AQUM has a negative bias in the amount of rainfall predicted, especially during the winter months. We also know that both  $\text{PM}_{10}$  and  $\text{PM}_{2.5}$  are generally under-predicted in the winter, although the cause for this could be a combination of emissions, chemistry in the model and meteorological

factors, such as rainfall washing out the particulates, or dispersion through wind.

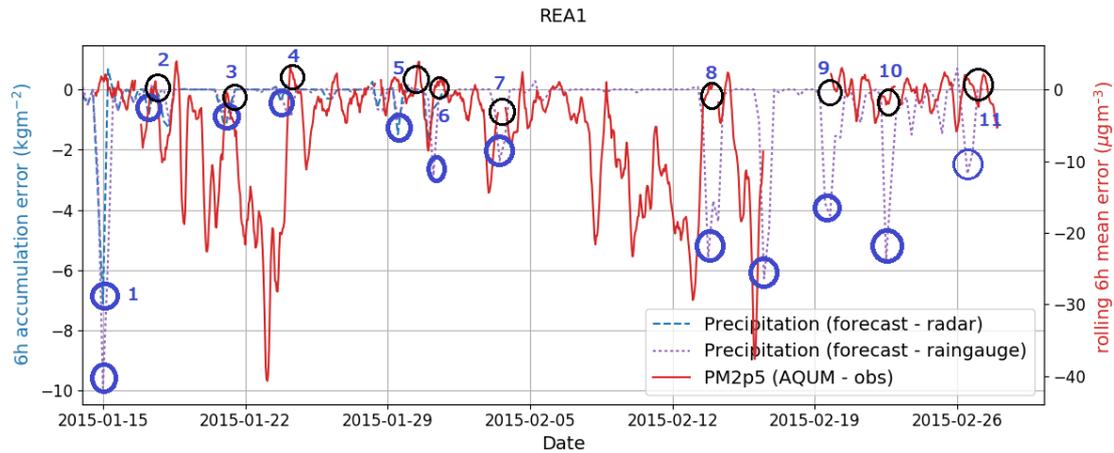
In this section, the influence of precipitation forecast error on the error in  $\text{PM}_{2.5}$  and  $\text{PM}_{10}$  is explored. When precipitation is under-predicted, one might expect insufficient amounts of PM to be removed from the atmosphere, leading to an over-estimation of PM. However, because the PM forecast already has a negative bias in AQUM (in both fine and coarse components), insufficient removal by precipitation could in fact lead to a ‘cancellation’ of the PM bias, thus resulting in a diminished bias overall. The aim of the present section is to test the following hypothesis:

**Under-estimating precipitation acts to decrease particulate matter forecast error.**

Although it is acknowledged that surface wind speed error could be a valuable parameter to evaluate against the PM error, as it was for  $\text{O}_3$  and  $\text{NO}_2$  in section 3.7.1, wind speed was not evaluated here. One reason for this is that it was felt that its distinct diurnal forecast error variability was inconsistent with the lack of a diurnal cycle in PM error.

This hypothesis is based on a non-bias-corrected version of the PM forecast. The forecast can also be assumed to not have a negative bias through a simple bias-correction method of subtraction of the mean error - in this case, the hypothesis would have to be amended to reflect this correction and read as ‘*under-estimating precipitation acts to **increase** particulate matter forecast error*’, consistent with our understanding of the well-known and observed relationship between rainfall and PM. Both versions will be discussed later in the present section.

Firstly, as an illustrative example of the error pattern that  $\text{PM}_{2.5}$  and precipitation exhibit alongside each other, figure 3.19 shows a time-series of the two variables’ errors for Reading New Town over the winter period. A centred 6-hour rolling mean was plotted because the hourly variability in precipitation is too sporadic to distinguish any relationship with  $\text{PM}_{2.5}$  errors simply by looking at it. Thus, any obvious patterns between the time-series of precipitation error and PM error can be identified visually. For example, in figure 3.19, it appears that the PM error (red line) reduces in magnitude whenever it follows a peak in either the blue or purple line (precipitation forecast error, verified by radar and rain gauge respectively). This data has not been bias-corrected. All the other sites in the evaluation show a similar pattern: both the  $\text{PM}_{2.5}$  and precipitation are generally under-estimated in AQUM (in line with results from sections 3.5.3 and 3.6.3) and peaks in precipitation are often followed by a reduction in PM error. The numbered circles identify



**Figure 3.19:** Example of a 6-hour centred rolling mean time-series for  $\text{PM}_{2.5}$  forecast error in Reading New Town (‘REA1’), overlaid with precipitation forecast error against rain-gauge and radar measurements. Note missing radar data at this site after 03/02/2015. This figure forms the basis of how figure 3.20 is created.

the precipitation error peaks, which form the centres of time-windows during which the  $\text{PM}_{2.5}$  error will be evaluated in the following section. Some precipitation peaks are clearly followed by a decrease in  $\text{PM}_{2.5}$  error (e.g. 2, 3, 8), while others lead to a sign change of the  $\text{PM}_{2.5}$  error from an under- to an over-prediction (e.g. 4, 5, 6). Figure 3.19 therefore illustrates the process of identifying the precipitation error peaks within the time-series of a single site. This process will be repeated across the time-series of all the sites, and the associated  $\text{PM}$  behaviour will be aggregated into a single composite figure, as described in the following section.

### 3.7.2.1 Identification of precipitation error events

To find out whether the precipitation errors have a statistically robust correlation with the  $\text{PM}_{2.5}$  or  $\text{PM}_{10}$  errors across all sites, the following steps are taken:

1. Identify paired WMO and particulate matter sites within a 2-model grid-box distance, i.e. 24 km <sup>8</sup>;
2. For each pair, identify local “peaks” in the precipitation hourly error time-series using (1) rain gauge measurements and (2) radar. For example, let us assume that a precipitation under-estimation peak occurs at 00:00 on 01-01-2015.
3. Define a 24-hour time window centred on each peak, ensuring that windows are independent. This means that windows do not overlap and only the first event

<sup>8</sup>Although this may seem like a long distance between sites, reducing it to e.g. one model grid-box (i.e. 12 km) would make the data sample insufficient.

counts. In the above example, this means 18:00 on 31-12-2014 to 00:00 on 01-01-2015 as the “before” window; and 00:00 to 06:00 on 01-01-2015 as the “after” window.

4. Match up the  $PM_{2.5}$  and  $PM_{10}$  hourly forecast errors within the 12-hour window (rejecting any window with missing data points);
5. Composite the PM error windows.

The magnitude of precipitation error “peak” is not taken into account, as the question under investigation is whether the presence of precipitation error alone influences the PM forecast, rather than by how much. Furthermore, since precipitation is an efficient PM sink, the mere presence of raindrops – regardless of the rainfall rate – already reduces particulate concentrations.

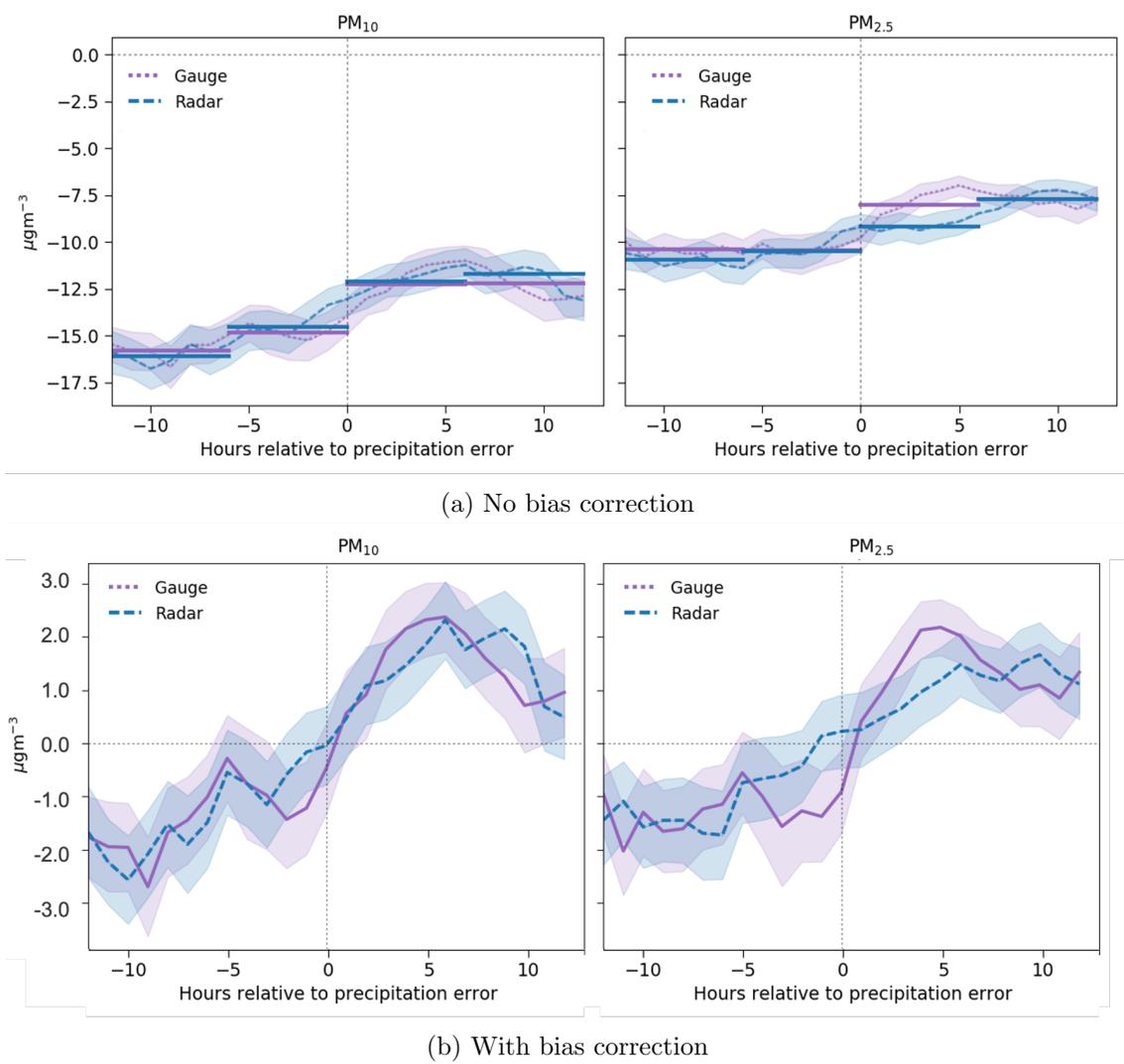
For both  $PM_{10}$  and  $PM_{2.5}$ , the total number of PM error peaks using rain gauge and radar data to verify the precipitation forecasts is 108 and 128 respectively. This is a sufficient number of data points to create a composite mean profile of the particulate matter errors (with standard error on the mean) before and after an instance of a precipitation error peak.

### 3.7.2.2 Analysis of PM errors following precipitation under-estimation

Figure 3.20 shows the resultant composite sub-sample of PM forecast error, for  $PM_{10}$  (left) and  $PM_{2.5}$  (right). The purple, dotted line is the average PM forecast error of the sub-sample based on precipitation forecast errors verified by rain gauge measurements; meanwhile the blue, dashed line is the PM error mean for the radar-verified precipitation. 6-hour averages in the forecast error were calculated across the 24 hour windows and shown as bold horizontal lines. In figure 3.20a, the mean error is shown with no bias correction, meanwhile in figure 3.20b mean bias has been removed from the mean error, such that the mean error lines pass through the axis origin, (0,0). This is done in order to provide a clarification in how to best interpret the results, as the implication from figure 3.20a may be counter-intuitive.

In figure 3.20a, the 6-hour PM error averages highlight the reduction in both  $PM_{10}$  and  $PM_{2.5}$  error after the precipitation error event. Table 3.3 gives the corresponding absolute PM error values and the error percentage decrease following the precipitation error event.

Mean  $PM_{10}$  forecast error in the 6 hours prior to the precipitation error is



**Figure 3.20:** Composite of forecast mean error in  $\text{PM}_{10}$  (left) and  $\text{PM}_{2.5}$  (right) before and after instances of precipitation under-estimation ( $x = 0$ ) verified with rain gauge (purple, dotted) and radar (blue, dashed) observations. Shaded area is standard error of the mean. (a) is the mean error, while (b) is the mean error bias-corrected by removing the mean bias.

$-14.65 \mu\text{gm}^{-3}$  and  $-14.14 \mu\text{gm}^{-3}$  for rain-gauge and radar respectively. In the 6 hour window following the precipitation error, those errors decrease to  $-11.76 \mu\text{gm}^{-3}$  and  $-11.80 \mu\text{gm}^{-3}$  respectively. This change in  $\text{PM}_{10}$  error amounts to an overall decrease when the precipitation forecast is verified by rain gauge (19.7% decrease) and radar (16.5% decrease) measurements.

In the 6 hours prior to the precipitation error, the  $\text{PM}_{2.5}$  error is  $-10.28 \mu\text{gm}^{-3}$  (rain-gauge) and  $-10.12 \mu\text{gm}^{-3}$  (radar). The mean error decrease after a precipitation event is similar to  $\text{PM}_{10}$  in absolute terms, but appears larger percentage-wise: the  $\text{PM}_{2.5}$  errors decrease by 25.8% for rain-gauge measurements (reaching  $-7.62 \mu\text{gm}^{-3}$ ). For radar-verified precipitation, the  $\text{PM}_{2.5}$  error has a smaller decrease than for the rain-

gauge precipitation, where the absolute error decrease is less than half ( $1.05 \mu\text{gm}^{-3}$  for radar,  $2.66 \mu\text{gm}^{-3}$  for rain-gauge). This value has an asterisk in table 3.3, as the absolute decrease in error is actually largest within the T+6 to T+12 hour window following the precipitation error. This can be seen in both sub-figures 3.20a and 3.20b, where the radar error local maximum (blue line) occurs between 6 to 12 hours after the precipitation error onset. Considering the delayed effect in  $\text{PM}_{2.5}$  error decrease with radar-verified precipitation in figure 3.20a, the peak decrease has an absolute value of  $2.53 \mu\text{gm}^{-3}$  (25% decrease). Therefore the peak error decrease with radar-verified precipitation matches that of the rain-gauge verified precipitation forecast, although it occurs within the T+6 to T+12 hour window rather than the T+0 to T+6 window.

Through removing the mean bias as in figure 3.20b, the result can be interpreted in the opposite way to reflect the assumption of no underlying bias in the  $\text{PM}_{10}$  and  $\text{PM}_{2.5}$  forecasts, and thus focusing only on the error fluctuations as a direct result of the precipitation error. The figure demonstrates that after the precipitation error onset, both the  $\text{PM}_{10}$  and  $\text{PM}_{2.5}$  forecast errors are positive, peaking at 5 - 6 hours after the precipitation error. Although the additional process of removing the mean bias was involved, this interpretation makes more intuitive sense than the original because it demonstrates that if there was no underlying PM forecast bias, the under-estimation of precipitation would indeed result in insufficient wet deposition of PM from the atmosphere, further leading to accumulation of the particulates in the model and hence a positive mean error.

Whichever interpretation is adopted, the original hypothesis is either proven true, or is rejected. The initial interpretation, i.e. that from figure 3.20a, supports the original hypothesis that an under-estimation of precipitation leads to a decrease in the PM forecast error in AQUM. The second interpretation, i.e. that from figure 3.20b, supports a flipped version of the original hypothesis, while making the assumption that the mean PM forecast

**Table 3.3:** Comparison of decrease in  $\text{PM}_{10}$  and  $\text{PM}_{2.5}$  mean error 6 hours prior to and after a precipitation error event, verified by rain gauge and radar measurements. Absolute values ( $\mu\text{gm}^{-3}$ ) and percentages.

	Precipitation verification	Mean error during T - 6 to T - 0 ( $\mu\text{gm}^{-3}$ )	Mean error during T + 0 to T + 6 ( $\mu\text{gm}^{-3}$ )	Absolute decrease ( $\mu\text{gm}^{-3}$ )	% error decrease
$\text{PM}_{10}$	gauge	-14.65	-11.76	2.89	19.7
	radar	-14.14	-11.80	2.34	16.5
$\text{PM}_{2.5}$	gauge	-10.28	-7.62	2.66	25.8
	radar	-10.12	-9.07*	1.05*	10.4*

bias is zero.

### 3.7.2.3 Discussion

Considering the underlying hourly error lines on figure 3.20, it is surprising to see that the maximum PM error peak occurs approximately 6 hours after the precipitation peak (except for PM<sub>2.5</sub> with radar verification, where the peak is closer to 9 hours after the precipitation error onset). One would expect a more dramatic change in PM error sooner, due to the efficiency of the washout process to remove PM from the atmosphere. For example, [Ouyang et al. \(2015\)](#) demonstrate that the first 5 - 10 minutes of rain significantly contribute to removal of PM<sub>2.5</sub> in a study over Beijing. The coarse component was not evaluated in the Beijing study, but perhaps it might be “washed out” even faster than PM<sub>2.5</sub> if a similar observational study were to be conducted. The delayed response of the PM forecast error to the precipitation error could be enhanced by secondary compounding factors, e.g. surface wind speed error, which could result in too much or too little horizontal transport and mixing of the particulates. The results presented in this section could therefore be made more robust by taking the potential effects of wind speed forecast error into account.

Furthermore, the differences between PM<sub>2.5</sub> and PM<sub>10</sub> error decrease are surprisingly small. One might have expected removal of particulates through washout to be more effective on the coarser particles (i.e. PM<sub>10</sub>) due to their larger aerodynamic diameter. Below-cloud scavenging is more efficient for coarse particles (because of physical impact and inertia) and for very fine particles on the sub-micron scale (because of Brownian motion) ([Dacre et al., 2020](#)). Particles in the intermediate range, such as PM<sub>2.5</sub>, tend to fall around streamlines of precipitation and are not captured as efficiently. Therefore, to see such similarities in these results is intriguing. Perhaps this is due to the make-up of PM<sub>10</sub> represented in AQUM, where PM<sub>2.5</sub> makes up a major sub-group of the coarse particles and therefore the effects of precipitation forecast error in the model are similar for both PM<sub>2.5</sub> and PM<sub>10</sub>.

## 3.8 Summary, discussion and conclusions

In this chapter, the forecasts of four air quality variables (O<sub>3</sub>, NO<sub>2</sub>, PM<sub>2.5</sub> and PM<sub>10</sub>) and three meteorological variables (10 m wind speed, 1.5 m temperature and precipitation accumulation) were evaluated using traditional point-based metrics such as mean error

and Pearson-r correlation coefficient (PCC). The study periods were summer and winter of 2015 and 2017.

### 3.8.1 O<sub>3</sub>, NO<sub>2</sub> and 10 m wind speed

Forecast concentrations of O<sub>3</sub> and NO<sub>2</sub> were compared to observations in urban background sites during JJA and DJF 2017. A systematic bias was found in the AQUM forecast of both variables, where O<sub>3</sub> observations are generally over-estimated, while the NO<sub>2</sub> observations are under-estimated (especially during the daytime in JJA). The diurnal profiles of observed and forecast concentrations, aggregated over 29 urban background sites, showed that AQUM replicates both the observed NO<sub>2</sub> and O<sub>3</sub> concentration profiles in DJF. It is therefore possible that there exists an additional source of forecast error in O<sub>3</sub> and NO<sub>2</sub> which is unrelated to the representation of atmospheric chemistry or surface emissions. The diurnal profiles of concentrations are less accurately forecast in JJA, suggesting that there is a mis-representation of some physical mechanism responsible for the temporal error in the forecast of concentrations dominant during the summer. My suggestion is that this mechanism could be related to the surface wind speed forecast, and therefore hypothesise that there is a relationship between O<sub>3</sub>, NO<sub>2</sub> and 10 m wind speed forecast error at the locations of interest.

It was found that 10m wind speed is overestimated between 0 - 6 UTC, which coincides with the O<sub>3</sub> night-time overestimation. A possible reason for the night-time wind speed bias could be due to the use of stability functions within the boundary layer scheme of [Lock et al. \(2000\)](#). These functions artificially introduce additional diffusion and mixing of higher momentum air down to the surface, in order to reduce a bias in surface temperatures. This could also explain why the temperature forecast has insignificant bias, although this was not investigated further within this thesis.

Furthermore, it was found that there is a 1 - 2 hour offset in the timing of the morning increase of the total oxidant (sum of O<sub>3</sub> and NO<sub>2</sub>) surface concentrations between AQUM forecast and AURN observations. Thus, a potential link between the temporal offset in total oxidant and inaccuracies in 10 m wind speed or 1.5 m temperature was explored.

The main findings of this study were:

- When only the 90th percentile of wind speed forecast error is considered, the corresponding forecast of NO<sub>2</sub> has a larger negative bias than usual. Conversely, O<sub>3</sub> forecast concentrations have a larger positive bias than usual. This suggests that over-estimating night-time 10 m wind speed coincides with too much O<sub>3</sub> and not

enough NO<sub>2</sub> in the forecast, which supports the hypothesis. This relationship could be explained by enhanced dispersion of NO<sub>2</sub> and mixing of O<sub>3</sub> from the residual layer, however exploring this further will require a process-based evaluation of the atmospheric constituents and wind speed, as well as a study of boundary layer entrainment. This is not possible with a point-based evaluation of the forecast-observation pairs alone, thus a process-based method to investigate this is proposed in chapter 5.

- Although other meteorological factors may have a larger influence on the air quality forecast, it can be concluded that **10m wind speed is one variable whose improvement within the model could lead to an improvement in O<sub>3</sub> and NO<sub>2</sub> summertime forecasts during the night and early morning hours.**
- 1.5 m temperature was found to have a negligible daytime bias which is unlikely to influence the pollutant forecast error.

Chapter 5 will explore the use of a process-based evaluation to further understand the physical relationship between wind speed error, the total oxidant and O<sub>3</sub> and NO<sub>2</sub> individually. Specifically, it will be further investigated whether the meteorological processes are causing the lag in the morning increase of modelled total oxidant.

### 3.8.2 PM and precipitation

The relationship between forecast errors in PM and precipitation were directly compared through the process of compositing errors. First, through the use of categorical (e.g. frequency bias; accuracy) and continuous (e.g. RMSE; mean error) statistics, it was found that the precipitation forecast was generally under-estimated in AQUM. It was also found that during the summer, non-zero rain events are forecast twice as much in AQUM as they are observed at point locations. These results are consistent with the model inter-comparison study of [Brunner et al. \(2015\)](#), who showed that many coupled chemistry-meteorology models slightly under-estimate precipitation in the European domain on a monthly basis. They conclude that differences in models' removal of PM is likely to be influenced more by the specifics of their wet scavenging schemes than the rainfall rates. It is also likely that biases are rooted in the cloud scheme and / or the aerosol feedback on radiation, which is an additional factor that could be influencing the precipitation bias seen in AQUM.

In the absence of other compounding factors, an under-estimation of rainfall is ex-

pected to result in an over-estimation of PM due to a reduced influence of the washout effect. However, by direct comparison of the AQUM forecast and AURN observations, PM was found to be generally under-estimated for the winter season. There is an overall negative bias in the DJF 2015 forecast for both  $\text{PM}_{2.5}$  ( $-4.6 \mu\text{gm}^{-3}$ ) and  $\text{PM}_{10}$  ( $-10.5 \mu\text{gm}^{-3}$ ). This suggests that other factors may be playing a role to result in a negative PM bias, such as wind speed or emissions. It has been suggested by [Savage et al. \(2013\)](#) and others via personal communication (P. Agnew, Met Office) that the PM forecast has a negative bias due to the emissions inventory used, alongside the fact that some particulates are not represented in the model. It was therefore the role of this chapter to evaluate the immediate influence of the precipitation forecast error on the PM error, outside of any statistical emissions inaccuracies.

The compositing of PM errors within a 24-hour window centred on the precipitation forecast error instance could be interpreted in two ways. Firstly, it was evaluated in a manner that is true to the forecast as it stands with no bias correction. The analysis served as a search for evidence to support the hypothesis that the precipitation under-estimation acts to reduce the PM forecast error (before bias-correction). However, for the purposes of finding out the near-immediate effect of precipitation error on PM, the systematic negative PM forecast bias can be removed by a simple subtraction of the mean error. Thus the original hypothesis can be flipped, to read more intuitively as ‘precipitation under-estimation acts to *increase* the PM forecast error’. Both scenarios were presented in the chapter, where the main findings are as follows:

- Without applying a bias-correction method to the PM forecast, its error decreases soon after the event of a precipitation under-estimation; however, the sharpest reduction in PM error does not occur until up to 6 - 7 hours following the precipitation error event;
- $\text{PM}_{10}$  bias is between 16% to 20% smaller after the precipitation error event than prior to it. For  $\text{PM}_{2.5}$ , the error decrease within 6 hours after the precipitation error is between 10% to 26%. However, the maximum  $\text{PM}_{2.5}$  error reduction occurs around 9 hours after the precipitation event when verified against radar measurements, and reaches a 25% reduction also.

The above results provide evidence for existence of a relationship between errors in the precipitation forecast and the PM forecast. Furthermore, supporting evidence for the original hypothesis that precipitation under-estimation reduces the PM forecast error has

been shown. When the mean PM forecast bias is corrected, the hypothesis is adapted to reflect the change: that precipitation under-estimation *increases* the PM forecast error. In both cases, this is likely to be because of the reduced effect of washout of PM in the model when precipitation is under-estimated. As it stands, the results in this chapter suggest that the PM forecast is more accurate when the precipitation is under-estimated than it would be if it were improved. The consequence for model developers is that if the precipitation forecast improves, it may actually increase the PM under-estimation if the main PM forecast bias is not addressed. One way of addressing this the systematic PM bias would be to improve the spatial and temporal representation of primary emissions in the model. Another factor which could be influential is the wind speed and direction, which has not been evaluated in this chapter in the context of the PM forecast due to lack of time. Given more time, it would have been beneficial to conduct sensitivity studies of how the PM forecast error is affected within different points on a joint precipitation - wind forecast error distribution.

Precipitation is widely recognised as a sink of soluble gases and aerosols via wet deposition ([Brunner et al., 2015](#)), and the findings from this chapter agree with previous studies such as [Gong et al., 2006](#), where an under-estimation of precipitation resulted in inadequate deposition of aerosols or tracers, hence over-estimating their concentrations. The results presented in this chapter are novel in the sense that, to the best of my knowledge, there is a gap in the literature with respect to systematically evaluating the near-immediate change in PM forecast error before and after a precipitation bias event.

It is worth asking whether replacing point-based evaluation metrics with a neighbourhood methods could reveal a stronger relationship between the precipitation and PM errors. Results presented in this chapter do not enable us to come to a strong conclusion, although a relationship clearly exists. One problem with the method of compositing individual precipitation error “events” is that although it considers a 12-hour time window at either side of the “event”, it does not account for any spatial errors which could falsely penalise the forecast. This relationship is explored further in [chapter 4](#) using probabilistic neighbourhood verification methods.



## 4. Neighbourhood forecast verification

### 4.1 Introduction

There exist verification methods which aim to provide a deeper understanding of the nature of the forecast error than is possible to extract with traditional grid-point based methods such as the root-mean-square-error (RMSE). Continuous advancements in grid resolution of AQ models pose a need for the development of verification techniques appropriate for the scale of the forecast, in order to avoid problems relating to verification of high resolution forecasts, such as the double-penalty problem. Where a precise matching between a high-resolution forecast and observation of a variable is required, the double-penalty effect can increase the false-alarm rate due to small displacement errors, which would have otherwise been a ‘hit’ in a coarser forecast (e.g. [Gilleland et al., 2009](#)). The need for development of non-traditional verification methods was identified within the last couple of decades, where some of the first discussions of novel verification methods for precipitation forecasts are presented in [Davis and Carr \(2000\)](#). Non-traditional techniques can generally be classified into two types: object-oriented and neighbourhood-based.

**Object-oriented techniques**, such as cluster analysis ([Marzban and Sandgathe, 2006](#)), are useful for evaluating well-defined ‘objects’ within the forecast, such as a thunderstorm or a squall line. [Hoffman et al. \(1995\)](#) presents a method which evaluates the structure and amplitude of the displaced ‘object’ within the forecast against observations. [Ebert and McBride \(2000\)](#) present another object-oriented technique which introduced ‘contiguous rain areas’ (CRA), where the accuracy of the forecast is determined from evaluating the overlapping area of the forecast and observed rainfall. A subsequent three-dimensional verification method called Structure - Amplitude - Location (SAL; [Wernli et al., 2008](#)) followed, with specific uses for e.g. quantitative precipitation forecasts (QPF) around a river catchment. It separately considers the three dimensions of the forecast in relation to the observations, thus providing useful information into which one needs the most improvement in order to create a more accurate forecast. Outside of precipitation

forecasts, SAL could also be used for evaluating the accuracy of e.g. an urban plume forecast (Dacre, 2011).

**Neighborhood-based verification methods** relax the requirement of a precise spatio-temporal matching between the forecast and the reference observation, thereby giving credit to “close” forecasts. Grid cells within a defined neighbourhood are assumed to have an equal probability of correctness. They also bypass the requirement of matching objects and are applicable to less well-defined features. Many of the commonly used neighbourhood verification methods were specifically developed for verifying precipitation forecasts against high-resolution radar or raingauge measurements (Ebert, 2008). Precipitation forecasts are generally difficult to predict accurately in both space and time due to the sporadic nature of convective rain. Neighbourhood-based verification of such forecasts takes this into account. They include Upscaling (Yates et al., 2006; Zepeda-Arce et al., 2000), “Fuzzy logic” (Damrath, 2004), Intensity-scale Casati et al. (2004), and Fractions Skills Score (FSS; Roberts and Lean, 2008), among others. Upscaling involves averaging across multiple grid-cells or time-steps to a coarser resolution in order to match the observations, but is prone to smoothing out sharper features, e.g. flash floods. Fuzzy logic determines whether a given forecast is correct more often than it is incorrect. Intensity-scale methods determine whether the forecast’s structure is more accurate than random observations. FSS answers the question, “What are the spatial scales at which the forecast resembles the observations?”

The neighbourhood methods mentioned thus far are all useful for understanding how modellers can improve the precipitation forecast; or even extend this analysis to e.g. air pollution from wildfire events. These methods are mostly of a ‘neighbourhood-observation neighbourhood-forecast’ (NO-NF) nature, which intuitively require a densely populated set of observations, e.g. precipitation radar or satellite imagery. Unfortunately for ground-based forecasts for AQ in particular, lack of regularity in the locations of measuring sites is one of the limitations to accurately verifying an operational pollutant forecast. Gridded observation datasets can be created through interpolation methods such as ordinary kriging and subsequently used in object-based evaluation techniques such as SAL. However, kriging can be an additional source of inaccuracy itself (Dacre, 2011). Single-observation neighbourhood-forecast (‘SO-NF’) methods are therefore more suited to evaluate forecasts whose observational datasets are irregularly spaced. Forecast neighbourhoods of varying sizes are centred on a single observation, and forecast grid-cells are compared against the observation with an equal probability of being the ‘correct’ forecast. Mittermaier (2014)

builds on the SO-NF techniques of [Ebert \(2009\)](#), using ground-based observing sites to verify meteorological variables from convection-permitting NWP models, not just precipitation. They demonstrate the usefulness of the SO-NF method, especially when directly evaluating the forecast skill of deterministic and ensemble forecasts in the same manner.

Furthermore, [Mittermaier and Csima \(2017\)](#) highlight that regardless of the origin of error relating to model resolution, the loss of predictability at the grid-scale, as introduced by [Lorenz \(1969\)](#), has important implications on forecast accuracy. Using varying sizes of a neighbourhood forecast centred around the observation, they demonstrate that the decrease in forecast accuracy with lead time of e.g. precipitation is faster for high than for low resolution forecasts, due to the larger error growth rate. An ensemble forecasting approach at the kilometre scale is therefore appropriate to minimise the influence of small-scale error on the forecast skill at longer lead times, with an appropriate verification framework to match ([Mittermaier and Csima, 2017](#)).

AQUM is a deterministic forecast; as such, the use of ensemble verification techniques is generally inappropriate. However, the SO-NF method enables a near-probabilistic method of evaluating the forecast, by creating a pseudo-ensemble from the neighbourhood around the observation. Since the AURN network of AQ measurement sites is neither densely populated nor regular, adapting an existing spatial verification method - such as SO-NF, which relaxes the requirement of a spatially regular dataset for use within AQ forecasting - could be beneficial. This idea is explored further in the present chapter.

#### 4.1.1 Chapter aims

The SO-NF technique is not commonly used within routine AQ forecast verification, thus we are presented with a unique opportunity to test its usefulness within the operational pollution forecast model, AQUM. For its routine forecast verification, AQUM relies on the previously discussed traditional point-based verification metrics such as RMSE, correlations and Odds Ratio. Instead, given the success of neighbourhood-based verification methods for rainfall ([Ebert, 2009](#)) or oceanic ([Crocker et al., 2020](#)) forecasts, a neighbourhood verification method such as SO-NF could provide additional information to aid improvement of the AQ model, where traditional methods cannot.

The overarching goal of this chapter is therefore to build on the point-based evaluation carried out in chapter 3 by using a SO-NF verification technique in both the AQ ( $O_3$ ,  $NO_2$  and PM) and meteorological (precipitation and 10 m wind speed) forecast variables from AQUM. Specifically, two skill scores will be used: Brier Score (BS; [Brier, 1950](#)) and

Continuous Ranked Probability Score (CRPS; Matheson and Winkler, 1976) to evaluate forecast accuracy within neighbourhoods of different sizes, centred on either the AQ or meteorological measurement. Research into using spatial verification methods for the forecast of surface pollutant species as well as meteorological variables within AQUM could create a pathway for its implementation within routine verification.

### 4.1.2 Research questions

Building on the results from chapter 3, the following specific questions will be addressed within the present chapter:

1. How do wind speed forecast errors within the vicinity of an air quality monitoring site influence the  $O_x$  forecast?
2. Does the precipitation forecast error have a spatial component of influence on the PM forecast error?
3. What is the optimal neighbourhood size for which the meteorological forecast errors influence the pollutant species forecast errors the most?

Firstly, the High Resolution Assessment (HiRA; Mittermaier, 2014) framework of verification is introduced in section 4.2. HiRA is a verification tool which has recently been jointly developed by a number of forecasting centres, in order to achieve a common verification tool across many types of forecast models. In this chapter, HiRA is used for its ease of skill score calculation across large datasets. Because HiRA uses a probabilistic approach to verifying an otherwise deterministic forecast, it is a novel method of analysis and serves both as 1) a unique exploration of the relationships between spatial forecast errors in meteorology and air quality variables, and 2) a ‘testbed’ to determine whether HiRA is a potential candidate framework for verifying future air quality forecasts of higher resolution.

Section 4.3 introduces the probabilistic verification scores used in HiRA and how they have been used in the analysis within section 4.4. Section 4.5 presents the results of the analysis for 10 m wind speed and  $O_x$  forecast errors, while 4.6 presents the results of finding relationships between precipitation and PM forecast errors.

## 4.2 High Resolution Assessment (HiRA)

HiRA is one of the many tools within the overarching verification framework, MET (Model Evaluation Tools). This is a package created and maintained by the DTC (Development Testbed Centre), initially with the purpose of verifying the WRF (Weather Research Forecasting) model. Since its initiation by DTC in 2006, MET has been further developed and freely used by the model research, verification and operational communities. The idea behind such a framework is to create a freely-available, unified verification package which can be commonly used in forecasting centres globally. This means that meteorological forecasts from different models can be compared with a common verification tool. The present work uses MET version 8.1 and the user documentation can be found in [Newman et al. \(2019\)](#).

The first versions of MET included statistical tools for evaluation by traditional verification methods for continuous and categorical forecasts, e.g. RMSE. It also included two categories for evaluating spatial forecasts: “object-based” and “neighbourhood-based”.

A strategy for inclusion within MET, proposed by [Mittermaier \(2014\)](#), HiRA is based on evaluating the square neighbourhood around a point observation. Throughout this chapter, this technique will be referred to as ‘SO-NF’, as described earlier. SO-NF verification is based on the concept of creating a pseudo-ensemble of the forecast within a defined size of a neighbourhood around the observing site. Each model grid-box (akin to a single ensemble member) is equi-probable, in that the value of a variable in the grid-box near to the observation point is assumed to be just as likely to be the outcome as a grid-box farther away within the neighborhood. Therefore if the neighbourhood becomes too large, this assumption is violated. What neighbourhood size is “too large” will depend on the variable being tested. Thus, the skill of an ensemble forecast is affected by the size of the ensemble ([Ferro et al., 2008](#)). Similarly, it is expected that the size of the neighbourhood used in a SO-NF technique for an otherwise deterministic forecast is also of importance. The verification score should reflect the changes in the accuracy of the pseudo-ensemble, thus probabilistic measures like the BS and CRPS can be used to verify a deterministic forecast in a probabilistic way. It is important to note that the conventional use of BS and CRPS as measures of ‘skill’ (rather than just ‘accuracy’) involves comparison to a reference forecast (e.g. persistence or climatology), in order to evaluate the ‘skill score’ (SS):

$$SS = 1 - \frac{BS}{BS_{\text{ref}}} \quad (4.1)$$

This is usually done e.g. in order to compare the perceived improvement of a model upgrade from one version to another; or to compare the skill of two or more different forecasts against a common baseline. The skill of a forecast is therefore a second order measure, while accuracy is a first-order measure. It is acknowledged that the importance of calculating the full skill score in reference to a baseline; however in this chapter, only the raw values of the scores (i.e. not the ‘skill score’) are presented, as the aim here is not to compare the skills of two forecasts - rather, the aim is to use a non-traditional probabilistic metric to glean potential relationships between variables within the meteorological and AQ forecasts.

## 4.3 Probabilistic Verification Scores

### 4.3.1 Brier Score

The Brier score (BS; [Brier, 1950](#)) is used for categorical variables such as thresholded wind speed and 6 h precipitation accumulations. Using a threshold value for a continuous quantity such as wind-speed translates the forecast into a binary value which answers the question: when the observed 10 m wind-speed is  $< X \text{ms}^{-1}$ , what is the fractional proportion of the model neighbourhood which satisfies the threshold limit? The larger the neighbourhood fraction which matches the observed value (either 0 or 1), the smaller the difference between the forecast and observation. This difference is represented by the BS, thus ranging from 0 to 1, where 0 is the perfect score.

BS is a simple and familiar metric within the forecast verification community. [Murphy \(1973\)](#) demonstrated that the score can reveal more information beyond a single value by partitioning it into three components of reliability, resolution and uncertainty. These partitions are insightful for checking the forecast’s statistical properties relative to the observed event. The value of BS shows the magnitude of the probability forecast error:

$$\text{BS} = \frac{1}{N} \sum_{i=1}^N (p_i - o_i)^2 \quad (4.2)$$

where  $o_i$  is the observed quantity at time-step  $i$ , with value  $o_i = 1$  if the event happened, or 0 if it did not. In theory,  $p_i$  is the probabilistic forecast composed from the model grid-boxes within the neighbourhood of size  $K$ , such that

$$p_i = \frac{1}{K} \sum_{k=1}^K p_k \quad (4.3)$$

$$p_k = \begin{cases} 1, & \text{if event is forecast} \\ 0, & \text{otherwise} \end{cases}$$

but within MET, while the above holds true,  $p_i$  is taken to be the mid-point of the probability bin in which it resides. Probability bins are used within BS because it is a discrete score, unlike e.g. CRPS, which does not require probability bins. Thus, while in theory the range of BS is  $[0, 1]$ , where 0 is perfect, in practice HiRA gives a range of  $[0.0025, 0.9025]$  due to a squared factor of  $p_i$  and a dependence on whether the ‘event’ is observed or not. Let  $n_i$  represent the total number of ‘events’ within each binary observation category. For instance,  $n_{.1} = 1$  and  $n_{.0} = 0$  if the event is observed; but  $n_{.1} = 0$  and  $n_{.0} = 1$  if it is not. Then

$$\text{BS} = \frac{1}{T} n_{.1} (p_1 - 1)^2 \quad (4.4)$$

$$= \frac{1}{T} (p_1 - 1)^2 \quad (4.5)$$

if the event is observed, and

$$\text{BS} = \frac{1}{T} n_{.0} (p_o)^2 \quad (4.6)$$

$$= \frac{1}{T} (p_o)^2 \quad (4.7)$$

if it is not observed.  $T$  represents the number of time-steps over which the BS is taken - in the present evaluation,  $T = 1$  as only the spatial - not temporal - constraint is relaxed. More details about equations 4.5 and 4.7 can be found in [Newman et al. \(2019\)](#).

As an example, consider the following case of a grid-scale evaluation (i.e 1x1 neighbourhood), where the only two possible values of BS are 0.025 or 0.9025 due to the binary nature of this metric:

**Case 1: Correct forecast.**

$\mathbf{o} = \mathbf{1}$  and  $p_i = 0.95$ . Then  $n_{.1} = 1$  and  $n_{.0} = 0$ . So by equation 4.5,

$$\begin{aligned} \text{BS} &= \frac{1}{T} [n_{.1}(p_1 - 1)^2] \\ &= \frac{1}{T} [1 (0.95 - 1)^2] \\ &= 0.0025 \end{aligned}$$

which is the lowest possible score in this framework.

Using only the nearest grid-point to the point observation is not recommended for deterministic forecasts (Mittermaier and Csima, 2017). As soon as a neighbourhood is introduced, more probability thresholds are involved:

**Case 2: False alarm.**

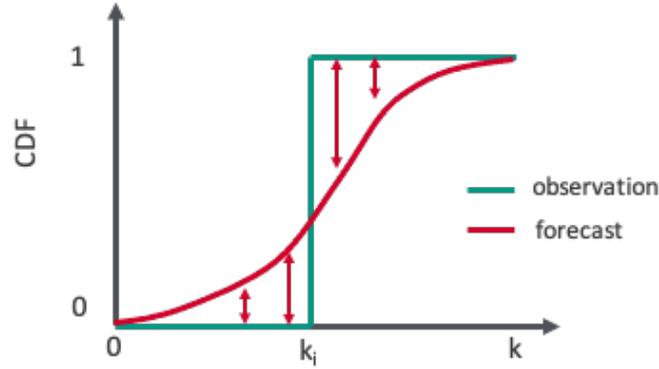
$\mathbf{o} = \mathbf{0}$ , so  $n_{.1} = 0$  and  $n_{.0} = 1$ . Consider a 3x3 neighbourhood where 7 out of 9 of the model grid-boxes forecast an event. In this case,  $p_i = 0.75$  (because  $\frac{7}{9}$  sits in the probability threshold bin of [0.7, 0.8]) and thus  $p_i$  takes on the mid-point value. By equation 4.7,

$$\begin{aligned} \text{BS} &= \frac{1}{T} [n_{.0}(p_0)^2] \\ &= [1 (0.75)^2] \\ &= 0.5625 \end{aligned}$$

BS is a base for other error metrics when a number of thresholds are evaluated at the same time, e.g. the Ranked Probability Score (RPS) is a multi-threshold version of the BS useful for categorical variables like cloud cover; or CRPS, which is a continuous version of the RPS and useful for continuous variables such as temperature. Therefore BS is a useful metric when a single threshold is considered in order to distinguish the skill of specific forecasts, e.g. how well does a model predict low wind speeds, or high precipitation? Identifying forecast skill for specific parts of the distribution of the meteorological variable will be useful for understanding its influence on the pollutant forecast.

**4.3.2 Continuous Ranked Probability Score**

CRPS is a probability score with an infinite number of classes of infinitesimal width, which can be expressed as an integral of the Brier Scores over all possible thresholds (Hersbach, 2000). CRPS is therefore a multi-threshold extension of the BS and is useful



**Figure 4.1:** Schematic of the Continuous Ranked Probability Score (CRPS) in terms of the cumulative distribution function (CDF) of an observed event and an ensemble forecast.

for continuous variables<sup>1</sup> which follow a normal distribution, such as surface temperature.

CRPS is calculated by evaluating the integral of squared differences between the cumulative distribution functions (CDF) of the forecast and observed event,

$$\text{CRPS} = \int_{-\infty}^{\infty} [P_f(x) - P_o(x)]^2 dx \quad (4.8)$$

where  $P_f(x)$  and  $P_o(x)$  are the CDF values for the forecast and observation evaluated at  $x$ , as in figure 4.1. In figure 4.1, the continuous variable  $x$  is shown in terms of discrete probability bins  $k_i$  (which are of infinitesimal width). The red arrows represent the differences between the observation and ensemble forecast CDFs. Note that the observation CDF is a step function due to it being a binary variable - either the variable in question (e.g. temperature) has value  $x$  or not. The ensemble forecast has a larger spread of probability bins, therefore its CDF forms a curve based on a normal distribution. If reduced to a deterministic forecast, the CDF would also become a step function and thus CRPS is equivalent to the mean absolute error (MAE) because of the square in equation 4.8. This is also true for when a grid-scale forecast is considered rather than a neighbourhood, i.e. the pseudo-ensemble in the present evaluation.

## 4.4 Using spatial forecast verification with non-gridded observations

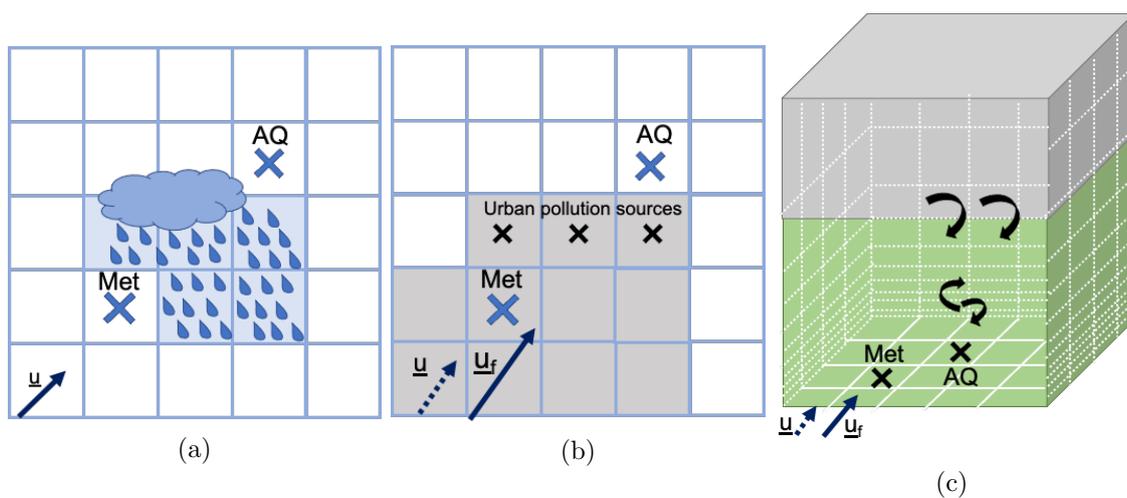
Verification methods where the spatial constraint is relaxed are better suited to forecasts where a precise spatio-temporal matching is not required. To demonstrate the ben-

<sup>1</sup>Although continuous in nature, wind speed is a variable which does not follow a normal distribution, therefore is better suited to the BS.

efit of this, figure 4.2 depicts a hypothetical example of a 5x5 neighbourhood ( $60^2 \text{ km}^2$ ), containing a WMO meteorological site and an AQ measurement site. In all three sub-figures, there is an observed mean 10 m wind field  $\underline{u}$  from the bottom-left corner of the grid, with a forecast wind field  $\underline{u}_f$  of a larger magnitude but same direction. The AQ site is located downwind of the WMO meteorological site, for hypothetical illustration purposes.

In example 4.2a, precipitation is forecast downwind of the meteorological site, but not within the grid-box of its location. Let us assume that the grid-box forecast is correct, i.e. no observed and no forecast precipitation. This means that if we were to only consider a traditional, point-based verification metric such as the mean error, we would not be able to see that the forecast incorrectly predicted some precipitation downwind of the meteorological site. Although no rain is forecast at the AQ site, forecast concentrations of horizontally advected pollutants like PM will be affected by the incorrect rain prediction upstream of the AQ site. This is a case of an inaccurate but “close” precipitation forecast, which still deserved merit.

In example 4.2b, there is an urban source of pollution situated downwind of the meteorological site, i.e. between it and the AQ site. As indicated by the different sizes of the arrows in the bottom left corner, forecast wind speed is larger than observations within the region marked by the grey grid-boxes. This means that pollutants such as PM or  $\text{NO}_2$  (which tend to have larger concentrations in urban areas) may be advected downwind,



**Figure 4.2:** Schematic depicting a model neighbourhood around a meteorological site paired to an AQ site.  $\underline{u}$  and  $\underline{u}_f$  represent the observed and forecast wind vectors respectively. (a) precipitation upwind of AQ site (plan view); (b) urban pollution upwind of AQ site (plan view); and (c) enhanced vertical mixing / entrainment from the residual layer above AQ site (3-d view). Diagrams are not to scale.

resulting in an over- or under-estimation of concentrations at the AQ measurement site, due to dispersion which is stronger in the model than is observed. Using a neighbourhood approach enables us to determine the size of the wind speed forecast error ‘spatial footprint’ for each individual pollutant species. For example, species which have more spatial variability are likely to be affected by wind speed errors which are more local to the pollutant sources than species which are more homogeneous in nature, such as  $O_3$ .

A traditional metric only gives information about the wind speed at the location of observation - it might be known that there is a positive bias, but it is not known how “close” the forecast was to being accurate. However, if a neighbourhood-based, threshold error metric - such as BS - is used, there is more information to be gleaned about the shape of the “footprint” in the forecast, and subsequently how it affects the pollutant measurement at the AQ site. Consider a wind speed threshold of particular interest,  $U_{\text{thresh}}$  (where  $U = |\underline{u}|$ ). If the observed wind speed  $U < U_{\text{thresh}} < U_f$  in a significant proportion of the neighbourhood, BS can give a simple interpretation of the error (i.e. that the wind forecast is too high in this case) without additional information about the magnitude by which the threshold was exceeded, if this information is not required.

However, neighbourhood verification of wind speed could also be useful for evaluating the influence of its forecast error on species such as  $O_3$ . Therefore, the relationship between horizontal winds and turbulent mixing is considered. In example 4.2c, imagine that the wind speed is still over-estimated (as in example 4.2b) but now let us focus on the process of entrainment. One could hypothesise that within regions of over-estimated wind speed, there is enhanced vertical mixing within the boundary layer and possibly from the residual layer above it (as outlined in section 2.2.1.5). Considering a forecast evaluation within a neighbourhood instead of only a grid-box, one could determine the spatial extent of the wind speed error, and therefore how it might affect entrainment further downwind, directly above the AQ site.

Although the AURN point observations in this dataset are irregularly spaced, they form a set of long-running and reliable ground-based air quality measurements. In the absence of a dense and regularly distributed network of surface pollution measurements, the AURN data is the best choice for this analysis. However, a secondary issue is that the air quality and the meteorological sites (WMO) are not co-located. Therefore in order to conduct an evaluation of how meteorological variables affect pollution concentrations, it was necessary to first create a set of matched pairs of air quality sites to their nearest WMO sites, subsequently filtering out pairs where the distance is  $> 24\text{km}$  (approximately

2 AQUM grid-lengths). This is done to ensure that the meteorological conditions at the WMO site resemble the conditions at the air quality site. A quantitative analysis has not been conducted of the correlations in meteorological conditions between the location of the air quality and its corresponding WMO site as a function of distance, but one would expect that in non-mountainous terrain (i.e. a large part of the UK), synoptic weather conditions will be similar at two locations within a 24 km distance of each other.

## 4.5 Relationship between total oxidant and 10 m wind speed forecast errors

With an understanding of the potential benefits of using spatial verification to assess the skill of a forecast, a hypothesis can be formed on how neighbourhood verification could help us understand some of the issues with the AQ forecast introduced in chapter 3. To recap, in chapter 3 a point-based comparison of total oxidant ( $O_x$ ) and 10m wind speed forecast is presented.

As a reminder,  $O_3$  concentrations are generally over-predicted throughout the day, with a mean bias of  $17 \pm 5 \mu\text{gm}^{-3}$ . Meanwhile, the  $\text{NO}_2$  concentrations have a smaller diurnal variability than  $O_3$  and are generally under-predicted. The  $\text{NO}_2$  mean bias has a diurnal cycle, where the largest increase in bias ( $8 \mu\text{gm}^{-3}$ ) occurs between 4 - 6 UTC and remains near-constant until 22 UTC. Error tendencies in  $O_3$  and  $\text{NO}_2$  are anti-correlated at night due to chemical reactions between NO and  $O_3$ , producing  $\text{NO}_2$  (see section 2.1.2). Thus combining  $\text{NO}_2$  with  $O_3$  and examining the error in the total oxidant eliminates errors due to NO emissions, particularly during the early morning hours.

Results of chapter 3 showed that accuracy of the  $O_3$  and  $\text{NO}_2$  forecasts had a site-based relationship with errors in the 10 m wind speed forecast. It was also shown that night-time wind speeds are overestimated, which could influence the amount of shear-generated turbulence within the NBL and, consequently, morning entrainment. However, a point-based evaluation does not reveal information about any spatio-temporal forecast errors. It is therefore worth considering the strength of the relation of wind speed and total oxidant  $O_x$  within a gridded neighbourhood, in order to gain a more comprehensive understanding of the spatial forecast error features.

In this section, the first hypothesis to be tested is that during night-time, forecast error in  $O_x$  is positively correlated with night-time over-estimation of 10m wind-speed, and the correlation is stronger when a forecast neighborhood is considered rather than a

single location. This is tested through an evaluation of night-time concentrations of  $O_x$  for the summer of 2017 (June, July, August) at 18 urban background and 6 rural background sites. Included in this evaluation are only the sites which measure both  $O_3$  and  $NO_2$ , rather than exclusively only one of the species. Omitted are urban sites predominantly located in London, i.e. Hillingdon, Kensington, Haringey Priory Park South, as these sites systematically skew the distributions due to anomalous  $NO_2$  underestimation in the forecast. It is possible that the  $NO_2$  forecast is systematically underestimated at these sites, due to findings of a recent campaign which used eddy-covariance methods to measure  $NO_x$  in central and outskirts of London, showing that the NAEI emissions inventory tends to underestimate ground  $NO_x$  by up to 150% (Vaughan et al., 2016). The NAEI inventory is used in AQUM as input emissions, and thus may contribute to the systematic underprediction at the London-based sites. The site at Rochester Stoke is also omitted from the analysis, as its anomalous concentrations may be a result of re-circulation of polluted air by the sea-breeze effect due to its proximity to the coast (e.g. Harrison and McCartney, 1979; Blumenthal et al., 1977; Lyons and Olson, 1973; Mavrakou et al., 2012). The above sites have therefore been excluded from the evaluation.

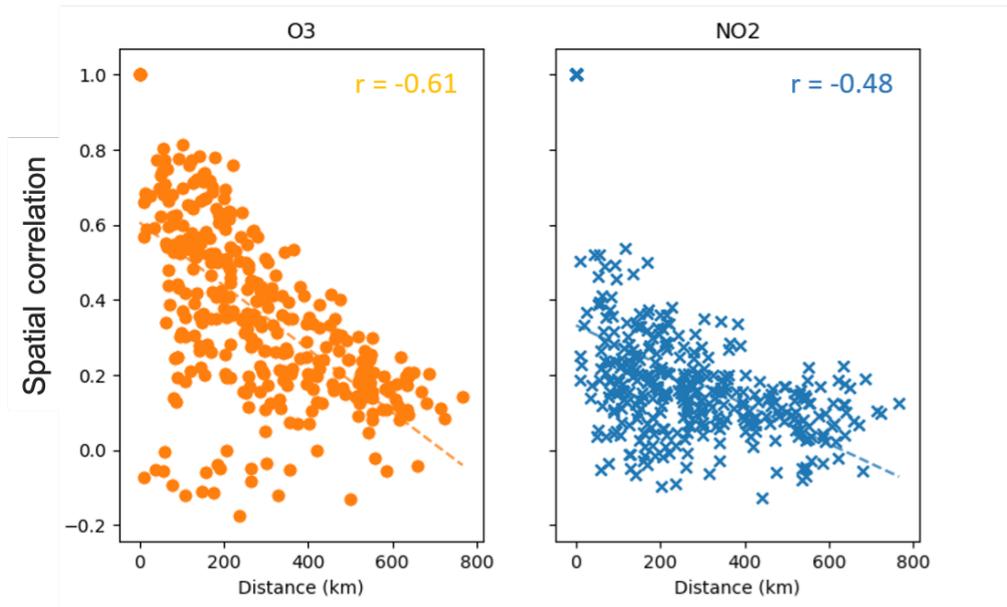
#### 4.5.1 Spatial correlations of $O_3$ and $NO_2$ observations

The heterogeneity of a forecast within a neighbourhood around a measurement site will vary on the nature of the atmospheric species, e.g. its average life-time and sink processes. Figure 4.3 shows a simple correlation method for determining the similarity of hourly observations taken at sites located at an increasing distance away from each measurement location, in JJA 2017. The y-axis represents the overall correlation of one site's collection of hourly measurements to another site's, located at a particular distance away marked on the x-axis.

The slower drop-off rate in correlation with distance in surface  $O_3$  than  $NO_2$  concentrations implies that  $O_3$  concentrations form a more horizontally homogeneous field than  $NO_2$ .

For  $O_3$  concentrations, the line of best fit suggests that correlation is high (0.5 on the y-axis) for sites located  $\leq 100$  km away, whereas for  $NO_2$  at the same distance, spatial correlation is 0.3. In fact,  $NO_2$  correlations between the measurement sites are  $\leq 0.5$  at most distances. This implies that  $NO_2$  concentrations are characteristic to each individual site, whereas there exists some similarity in  $O_3$  concentrations between locations.

Differences in homogeneity of fields between the two pollutants is in part due to



**Figure 4.3:** Spatial correlation of hourly  $O_3$  and  $NO_2$  observations between site locations.

the local nature of  $NO_2$  emissions, some of which are subsequently converted into  $O_3$  and other chemical species. The lifetime of tropospheric  $O_3$  (days within the boundary layer; weeks within the free troposphere) is longer than the lifetime of  $NO_2$  (hours).  $NO$  concentrations are also lower at higher altitude, which decreases titration and loss of  $O_3$  through formation of other species (e.g.  $N_2O_5$ ). The longer lifetime of  $O_3$  means that it can be transported over larger distances by the wind, and is therefore more well-mixed than  $NO_2$ .

Because the forecast bias of  $O_3$  and  $NO_2$  is of different sign and magnitude, the implication is that AQUM does not capture some of the observed sources or sinks of the two species (which have already been discussed in section 2.1.2). One of the physical sources of  $O_3$  into the boundary layer could be entrainment from the residual layer, which may act as a night-time reservoir of  $O_3$  following the de-coupling of the daytime mixed boundary-layer from the surface. In the morning, entrainment can therefore mix down  $O_3$ -rich air into the developing boundary layer from above, especially during windy conditions when generation of atmospheric turbulence is dominated by shear. This section aims to analyse the contribution of  $O_3$  vertical transport from the free troposphere to the forecast surface  $O_x$  concentrations by examining the 10 m wind speed forecast within the neighbourhood of an AQ measurement site. The next section introduces the methodology for the 10 m wind speed verification.

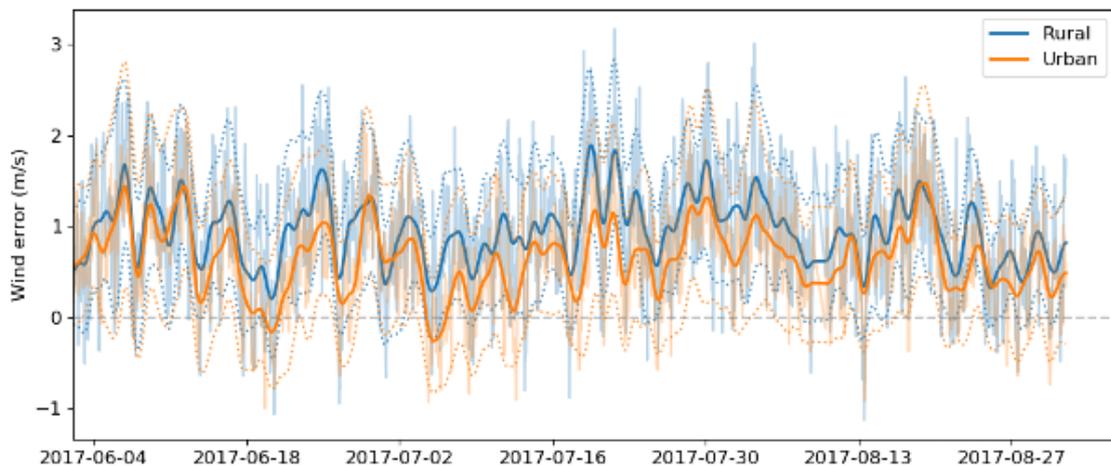
### 4.5.2 Neighbourhood verification of surface wind speed

Mean error in 10 m wind speed from AQUM at 24 sites is shown for the period of JJA 2017 in figure 4.4. 24 hr smoothing has been applied to filter out the diurnal cycle in forecast error across all sites. There is a clear positive bias in both urban and rural sites, with the rural locations generally exhibiting a greater positive bias than urban sites by around  $0.3 \text{ ms}^{-1}$  (which is small relative to the observed mean diurnal range of  $3 - 5 \text{ ms}^{-1}$ ).

In this section, 10 m wind forecast from AQUM is assessed on a neighbourhood basis using a threshold – based BS. To calculate BS, wind speed ( $v$ ) threshold criteria were set as:  $v < 4.0 \text{ ms}^{-1}$ ,  $v < 6.0 \text{ ms}^{-1}$ , and  $v < 9.0 \text{ ms}^{-1}$ . One would expect the relationship between  $v$  and turbulent kinetic energy (TKE) at low wind-speeds during stable conditions to be weak (Sun et al., 2012). Furthermore, because turbulence is suppressed by buoyancy when mechanical production is small, a strong relationship between wind speed BS and  $O_x$  forecast error is not expected for wind speed  $< 4.0 \text{ ms}^{-1}$ .

Using a threshold higher than e.g.  $9 \text{ ms}^{-1}$  is unsuitable in this analysis because the wind distribution is non-Gaussian, i.e. the probability distribution function has positive skew, where wind speeds above  $9 \text{ ms}^{-1}$  are a rare event. This implies that for a large threshold value, the binary threshold is satisfied and BS is 0 most of the time. On the other hand, assessing cases where the wind speed is too low, e.g.  $< 4 \text{ ms}^{-1}$ , may be mis-informative due to the weak relationship between TKE and low wind speeds.

Figure 4.5 shows the percentage of observed and forecast wind speed data which



**Figure 4.4:** Mean error in 10 m wind speed evaluated at 20 urban (orange) and 4 rural (blue) sites over JJA 2017, with 24 hr smoothing.

satisfies  $v < 4.0$ ,  $< 6.0$  or  $< 9.0 \text{ ms}^{-1}$  in the form of an average diurnal cycle within the study period. It is clear in all three sub-plots that a larger sample of the data counts satisfies the threshold criteria during the night. It was also shown in section 3.6.1 that the diurnal cycle of the observed and forecast 10 m wind speed (for the combination of urban and rural locations), has a night-time (0 - 5 UTC) forecast error ranging between  $-5$  to  $5 \text{ ms}^{-1}$ , with mean error of  $0.9 \pm 0.1 \text{ ms}^{-1}$  which decreases throughout the day.

When the thresholds are used to create a binary forecast for evaluation using the BS, these wind speed data for 0 - 5 UTC can be understood as follows:

- (a) The sample size of observations where  $v < 4.0 \text{ ms}^{-1}$  is between 70 to 75% of the entire study period. The sample size of wind speed forecasts satisfying this condition is smaller (between 57% to 62%) than the observations. This means that between 10% to 20% of the forecasts are ‘missed events’, due to the over-estimation of the wind speed. Therefore, one would expect the equivalent percentage of the wind speed BS for  $v < 4.0 \text{ ms}^{-1}$  to be non-zero.
- (b) For  $v < 6.0 \text{ ms}^{-1}$ , between 85% to 90% of forecasts and 90% - 93% of observations satisfy the threshold criterion. This means that only around 5% to 8% of the Brier scores are going to be non-zero.
- (c) Most of the data ( $> 95\%$ ) from the entire study period is included when considering wind speeds  $v < 9.0 \text{ ms}^{-1}$ , consequently a very small proportion of the data will have a non-zero BS.

The smaller the proportion of non-zero BS values, the more difficult it is to extract any existing correlation in the forecast error between wind speed and  $O_x$ . Therefore, a high

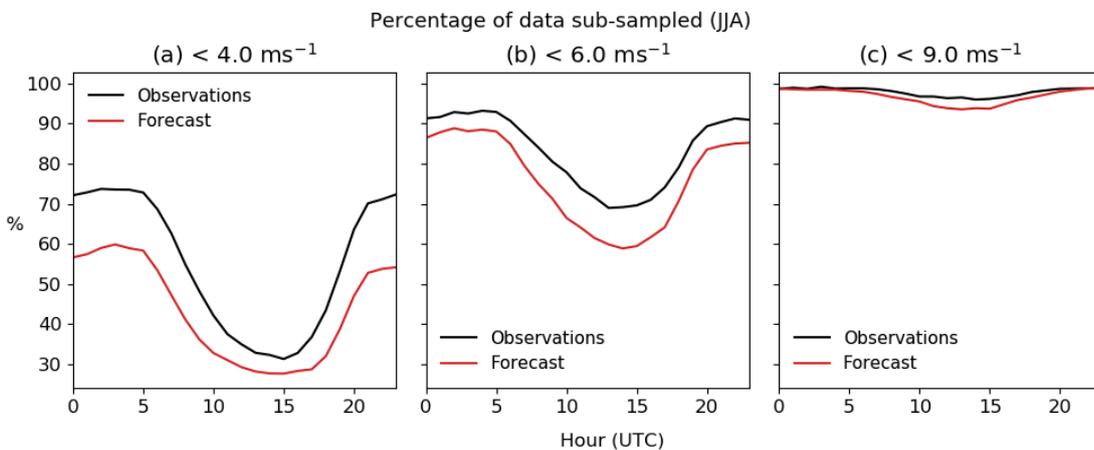


Figure 4.5: Percentage of data used for different wind speed thresholds.

threshold such as  $v = 9.0\text{ms}^{-1}$  is unsuitable for this evaluation, as results will not be statistically significant. Much of the following analysis rests on evaluating relationships in forecast errors for a wind speed threshold of  $v = 6.0\text{ms}^{-1}$ .

### 4.5.3 Forecast error correlations

In this section, correlations between forecast errors in 10 m wind speed (using BS) and in  $O_x$  (using mean absolute error) are presented. A reason for using mean absolute error for  $O_x$  is that for now, the magnitude rather than sign of the forecast error and its relationship to the errors in wind speed is of interest. In addition, mean absolute error is the grid-scale equivalent of CRPS, which will also be used in section 4.5.3.4 and thus consistency between the analysis metrics used is maintained.

The Pearson correlation coefficient  $r$  was calculated over the study period for each hour of the day, in order to recognise which points of the day there may exist a relationship between wind speed forecast error and  $O_x$  forecast error. The hypothesis is that if such a relationship exists, it would be most prominent at night because this is when 10 m wind speed forecast has the largest mean error. Furthermore, there is no net photochemical production of  $O_3$  during the night, so in order to focus on errors rooted in the wind speed forecast (thus minimising potential sources of error arising from the modelled chemistry), the present evaluation is based on early morning hours. Since AQUM is initialised at 18 UTC, the forecast lead time evaluated in this work is T+6 to T+30.

#### 4.5.3.1 Grid-scale comparison

If, for each hour of the day, the wind speed forecast BS (for  $v < 6.0\text{ms}^{-1}$ ) is directly compared against the mean absolute error for either  $O_3$ ,  $NO_2$  or  $O_x$  at the grid-scale, there is no clear correlation. This makes sense, because the wind speed BS has only two possible values when calculated over a single grid-box. It is more useful to evaluate diurnal correlations between the forecast errors when a neighbourhood on the wind speed forecast is considered, as this could inform us of the importance (or not) of accurately predicting the wind speeds around an AQ measurement site.

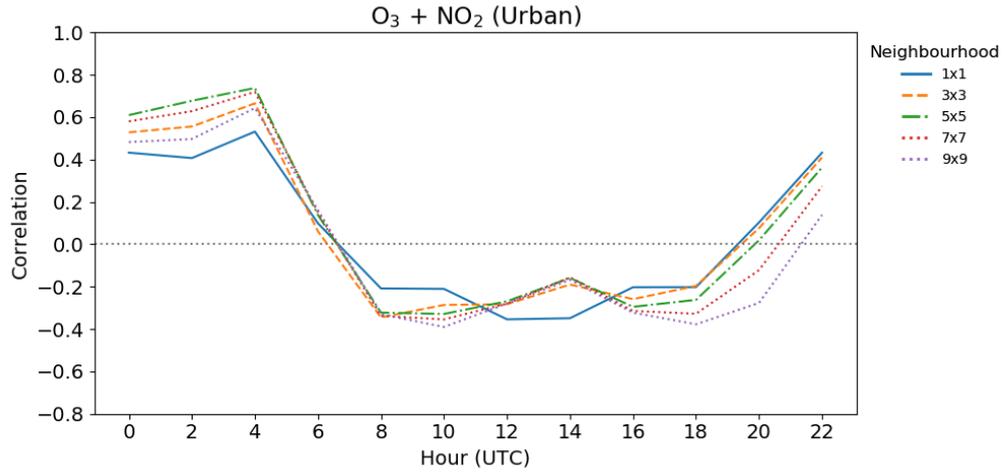
### 4.5.3.2 Diurnal variability of forecast error correlations at urban and rural sites

First, the  $O_x$  forecast error is compared with the 10 m wind speed BS for  $v \leq 6.0\text{ms}^{-1}$  separately for urban and rural sites, and presented as a mean diurnal profile in figures 4.6a and 4.6b. Wind speed BS will then be compared with CRPS for  $O_3$ ,  $\text{NO}_2$  and  $O_x$  in section 4.5.3.4.

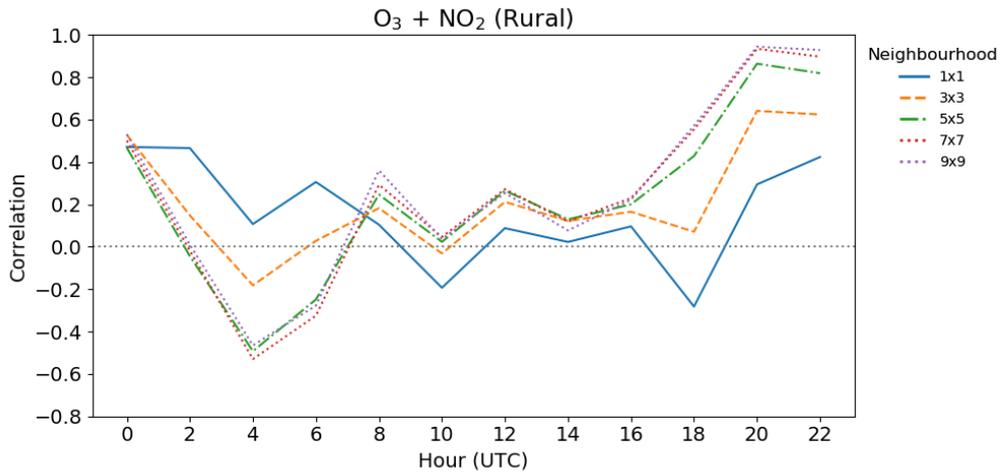
The analysis for urban background sites (18) in figure 4.6a shows a statistically significant (90th percentile confidence level) positive correlation between MAE of  $O_x$  and wind speed BS ( $r > 0.4$  for all neighbourhood sizes between 00 – 05 UTC). Although there is some suggestion of a weak negative correlation for during the day (6 UTC onward), the correlations are not statistically significant. For rural sites (6) in figure 4.6b, there is some evidence to suggest that night-time correlations between  $O_x$  and wind speed forecast error are weak but positive. However, the only statistically significant correlations occur during the evening and night (20 – 00 UTC), and are only significant for neighbourhoods 5x5 and larger. This means that the null hypothesis (i.e. that this could have happened by chance) cannot be rejected outside of the 20 – 00 UTC window and for small neighbourhoods. A reason for the low level of significance in the analysis of rural sites is likely to be the small sample size of 6 sites.

This above results suggest that for urban sites, the wind speed forecast error is correlated with  $O_x$  error at night only. This is reasonable, because (1) wind speed forecast error in AQUM is larger (and usually positive) during the night than daytime; and (2)  $O_x$  is a chemically conserved quantity during the night, in the absence of photolysis through sunlight. The existence of statistically significant positive correlations at urban sites for 00 – 05 UTC supports the hypothesis that night-time  $O_x$  concentrations in urban regions are strongly influenced by errors in the wind forecast, more so than during the day.

Separate  $O_3$  and  $\text{NO}_2$  correlation diurnal profiles are not shown here. This is because both pollutants exhibit a similar pattern: significant correlations at 00 – 05 UTC at urban sites ( $r$  is between 0.2 to 0.5 for  $O_3$ ) while rural  $O_3$  errors have no significant correlations with wind speed throughout the day.  $\text{NO}_2$  has negative error correlations during the night at urban sites, but they are significant to a lower (80th percentile) confidence level. However, positive correlations during the daytime ( $r$  is between 0.25 to 0.6 during 10 – 19 UTC) suggest that daytime  $\text{NO}_2$  MAE in urban areas is larger if wind speed over-estimation is larger. This is reasonable, because one might expect an over-estimation of wind speed to result in too much  $\text{NO}_2$  dispersion from anthropogenic sources in urban



(a) 18 Urban background sites



(b) 6 Rural background sites

**Figure 4.6:** Diurnal correlations of the BS for wind speed  $v \leq 6.0\text{ms}^{-1}$  and  $O_x$  mean absolute error over the study period 01/06 - 01/09/2017, for (a) urban background sites and (b) rural background sites. Different coloured lines are neighbourhood sizes for the wind speed forecast around the measurement sites.

regions. For rural sites, there is very little significance in results, except for negative correlations during the night ( $r$  is between  $-0.6$  to  $-0.8$  for  $22 - 05$  UTC) at grid-scale and  $3 \times 3$  neighbourhoods (80th percentile confidence level). The lack of significance could also be because the  $\text{NO}_2$  forecast error at rural sites is near zero on average, whereas the urban sites exhibit a systematic negative bias ( $-8 \pm 6 \mu\text{gm}^{-3}$ ).

Furthermore, there is some spread in correlation strength between neighbourhood sizes in figures 4.6a and 4.6b, but it is small. For both urban and rural sites, correlations are generally weakest for the grid-scale evaluations throughout the day (except rural  $\text{NO}_2$  during the night). Since only night-time correlations have significance across all three pollutant species, the following section will explore the relationships between wind speed

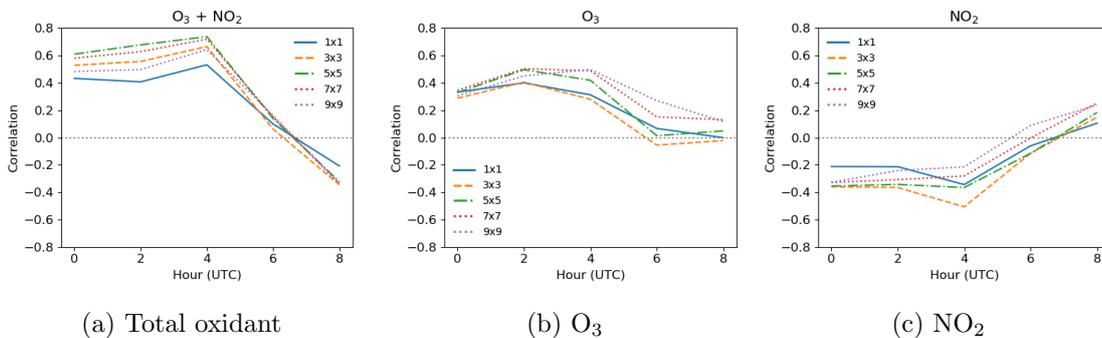
forecast error and  $O_x$ ,  $O_3$  and  $NO_2$  forecast error during 00 – 05 UTC.

### 4.5.3.3 Night-time correlations for $O_3$ and $NO_2$ at urban sites

In order to focus on the time of day when the correlations are statistically significant, figure 4.7 only considers the morning hours and urban background sites. Here, sub-plots 4.7a, 4.7b and 4.7c simply show the correlations between 00 – 08 UTC for  $O_x$ ,  $O_3$  and  $NO_2$  forecast error respectively (in the same format as figure 4.6a). The BS for wind speed has been calculated by using the threshold of  $v \leq 6.0 \text{ ms}^{-1}$ . Figure 4.7b shows that the  $O_3$  error correlation with wind speed forecast error is positive throughout the night. However,  $p \leq 0.1$  only for the largest neighbourhoods (7x7 and 9x9) between 02 – 05 UTC, when  $r$  ranges between 0.4 – 0.5. Sensitivity of the correlation strength to neighbourhood size is smaller for  $O_3$  than it is for  $O_x$ , as shown by lines which are closer together.

For  $NO_2$  across all the neighbourhoods,  $r = -0.35 \pm 0.1$  but results are only significant (to 80th percentile confidence) for neighbourhood sizes 3x3 and 5x5. The existence of statistical significance in the results for smaller neighbourhoods for  $NO_2$ , but larger neighbourhoods for  $O_3$ , suggests that the spatial footprint of influence of the wind forecast is smaller on the  $NO_2$  than the  $O_3$  forecast errors. In other words, only local wind speed forecast errors (on the scale of e.g.  $\leq 36^2 \text{ km}$ ) influence the surface  $NO_2$  forecast error.

The night-time differences in correlations between  $O_3$  and  $NO_2$  could be due to many factors.  $O_3$  and  $NO_2$  have different sources and sinks throughout the night: there is no chemical production of  $O_3$ , whereas  $NO_2$  emissions (e.g. from vehicles and manufacturing) do still occur near the surface. Both species are involved in chemical reactions producing  $N_2O_5$ , which results in a net loss of both  $NO_2$  and  $O_3$ . A physical explanation for the



**Figure 4.7:** Forecast error correlations in wind speed BS and the mean absolute error for (a) total oxidant, (b)  $O_3$  and (c)  $NO_2$ , averaged over 18 urban sites throughout the study period (JJA 2017).

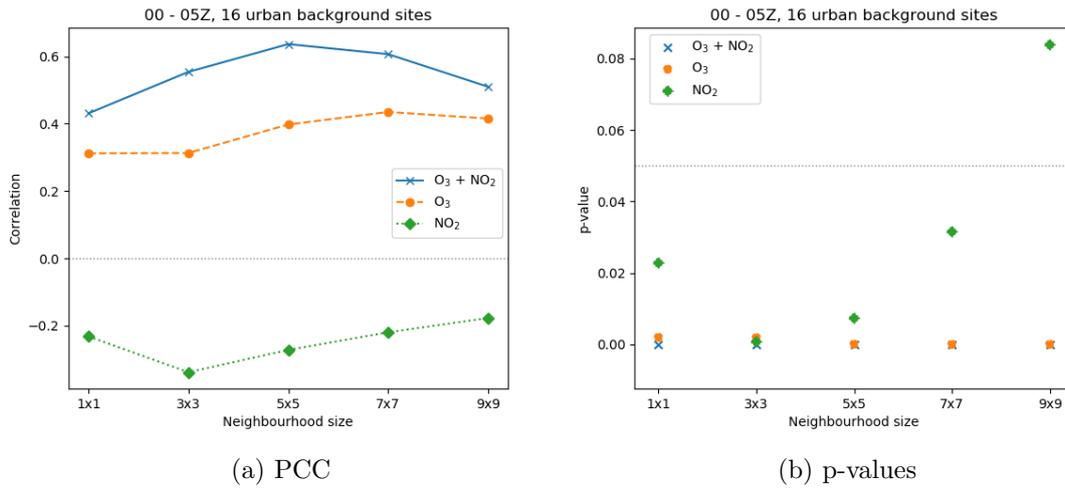
differences in the effect of wind speed over-estimation on both species is that enhanced winds could be causing too much vertical mixing in the model, thus creating a spurious source of  $O_3$  from the residual layer (as discussed in section 3.8.1). This effect is likely to be spatially uniform, therefore it would be expected that the horizontal gradients in  $O_3$  concentration fields are small. This reasoning is consistent with the spatial homogeneity in observed  $O_3$  concentrations which were discussed in section 4.5.1. Conversely for  $NO_2$ , enhanced winds could lead to too much dispersion, which may have local effects such as dilution of forecast  $NO_2$  concentrations near the sources (i.e. urban regions).

#### 4.5.3.4 CRPS for $O_3$ and $NO_2$

The previous section demonstrated that a neighbourhood verification metric can be used to evaluate the forecast of a meteorological variable (i.e. 10 m wind speed) in order to glean more information about its spatial error. It was compared against the mean absolute error of  $O_x$ ,  $O_3$  and  $NO_2$  at a point location. In the following section, we are able to find out more information about the spatial forecast error of the AQ variables as well, through the use of CRPS (introduced in section 4.3.2). The 10 m wind speed BS for increasing neighbourhood sizes around the AQ site is directly compared against CRPS for the pollutant at equivalent neighbourhood sizes. This method could be useful to understand how the spatial error in the wind speed forecast affects the spatial error of the AQ variables.

In figure 4.8, the wind speed BS for increasing neighbourhoods around 16 AQ urban background sites has been directly compared against CRPS for  $O_x$ ,  $O_3$  and  $NO_2$ . Panel 4.8a is the Pearson-r correlation coefficient (PCC), while panel 4.8b shows their respective p-values as functions of neighbourhood size. The correlations at each neighbourhood size were calculated from the pool of CRPS and BS scores aggregated across 00 - 05 UTC, i.e. night-time. Only the evaluation for urban background sites is shown here, as it was demonstrated in section 4.5.3.2 that night-time correlations in forecast errors are not significant at rural sites due to a small sample size.

In figure 4.8a,  $O_x$  and  $O_3$  forecast error correlations are positive for all neighbourhood sizes, with the strongest correlations for  $O_x$  at a 5x5 neighbourhood ( $r = 0.62$ ), while  $O_3$  error correlations are strongest for a 7x7 neighborhood ( $r = 0.42$ ).  $NO_2$  forecast errors are anti-correlated with wind speed forecast errors at all neighbourhood sizes, with the strongest relationship at a 3x3 neighbourhood ( $r = -0.35$ ). Figure 4.8b reveals that all but one of the correlations are statistically significant to 95% confidence (the insignificant



**Figure 4.8:** Forecast error correlations for wind speed BS based on the threshold of  $6.0 \text{ ms}^{-1}$  and  $O_x$  (blue, cross),  $O_3$  (orange, circles) and  $\text{NO}_2$  (green, diamonds) CRPS as a function of neighbourhood size. This is calculated for 00 - 05 UTC.

correlation is for  $\text{NO}_2$  at a 9x9 neighbourhood size).

Evaluating the wind forecast around a neighbourhood larger than a 3x3 area and trying to link it to influences on local  $\text{NO}_2$  is inappropriate due to the local nature of  $\text{NO}_2$  and its short lifetime in the boundary layer. It will travel shorter distances than  $O_3$ , and thus will be influenced by wind fields from around a smaller area nearby. This could be a reason for the strongest correlations for  $\text{NO}_2$  and  $O_3$  forecast errors to fall at different neighbourhood sizes: 3x3 and  $\geq 7x7$  respectively.

The same analysis was also done for rural locations (not shown). The only statistically significant results (p-value  $\leq 0.05$ ) is for  $\text{NO}_2$  forecast error correlations at the grid-scale and 3x3 neighbourhood, where  $r = -0.65$  and  $r = -0.4$  respectively. The negative correlation of  $\text{NO}_2$  forecast error with 10 m wind speed error is stronger in rural than urban regions on local scales. This is interesting, because the mean  $\text{NO}_2$  bias in rural areas throughout 00 - 05 UTC is close to zero, meanwhile in urban areas it is around  $-10 \mu\text{gm}^{-3}$ . The wind speed bias is larger in rural than urban areas by around  $0.3 \pm 0.3 \text{ ms}^{-1}$  (see figure 4.4), which is small relative to the absolute values, and the urban / rural difference in mean error during 00 - 05 UTC is within  $0.1 \text{ ms}^{-1}$ . This suggests that there is no significant difference in wind speed error between urban and rural regions during the night / morning, and that the difference in correlation strengths is instead related to the difference in magnitude of  $\text{NO}_2$  error.

#### 4.5.4 Summary and conclusions for this section

Assessing the performance of a forecast model with a neighbourhood method across varying neighbourhood sizes can reveal information about the spatial scales over which meteorological errors influence AQ errors. In that vein, the neighbourhood size at which the forecast error correlations are strongest have been assessed. In physical terms, this questions the spatial footprint of influence of the wind field on the pollutant concentrations at the AQ measurement site in question.

The main results of the evaluation are:

- There is a statistically significant positive correlation between 10 m wind speed forecast error and both  $O_x$  and  $O_3$  forecast errors between 00 – 05 UTC at urban locations only.
- The  $NO_2$  forecast error correlation with wind speed error is negative between 00 – 05 UTC at both rural and urban sites.
- At rural sites, correlations across all neighbourhood sizes between  $O_x$  and  $O_3$  forecast errors and wind speed errors are weak and not statistically significant. Only the  $NO_2$  forecast error correlations show significance, for 3x3 neighbourhoods and smaller. The correlations are negative, which means that over-estimating wind speeds under-estimates  $NO_2$  at local scales (3x3 neighbourhoods and less). This effect is stronger at rural sites ( $r = 0.5 \pm 0.1$ ) than urban sites ( $r = 0.3 \pm 0.1$ ).

It can be concluded that the spatial footprint of 10 m wind speed forecast error influence is smaller for  $NO_2$  forecasts than for  $O_3$  forecasts. Differences in the correlation strengths at varying neighbourhood sizes may be because of the opposing signs of the forecast error and locality of the sources: enhanced winds may be causing mixing and even entrainment of  $O_3$  into the boundary layer (as discussed in section 2.2.1.5), while  $NO_2$  experiences too much dispersion of the local pollutant. This means that  $NO_2$  is more sensitive to neighbourhood size used for the wind speed forecast evaluation than  $O_3$ , as the spatial footprint of influence of the wind speed error is small.

## 4.6 Relationship between PM and precipitation forecast errors

It is well established that wet scavenging by precipitation serves an important role for removal of particulates from the atmosphere, which was introduced in section 2.3.3. Due to the efficiency of wet deposition acting to reduce PM concentrations in the boundary layer, one might expect to find a relationship between forecast errors in precipitation and both PM<sub>2.5</sub> and PM<sub>10</sub>. It was confirmed in section 3.7.2 that under-estimating precipitation at a point location was soon followed by a reduction in the (negative) PM bias.

However, sometimes precipitation may be observed at a point location, but the forecast could be misaligned in space and / or time. Point-based metrics (RMSE and PCC at individual sites) do not acknowledge the potential skill of the forecast within the vicinity of the matched point location. It is therefore appropriate to extend the point-based analysis of PM and precipitation errors to the neighbourhood methods described in section 4.3. A neighbourhood technique may reveal more information about how forecast errors in rainfall within a neighbourhood of the PM receptor site may influence it. Therefore the second hypothesis explored in this chapter is:

- Precipitation forecast error is positively correlated with PM forecast error when a neighbourhood verification technique is used.

Furthermore, one might expect that increasing forecast neighbourhood size will be able to capture the observed relationship between PM and precipitation better than at the grid-scale because precipitation elsewhere in the neighbourhood may influence the PM at the monitoring station. Therefore, comparing forecast errors in precipitation and PM at the grid-scale only is likely to miss this relationship. In this section, it will be tested whether expanding the neighbourhood of evaluation will increase the relationship between precipitation forecast error and PM forecast error.

### 4.6.1 Data and methodology

AQUM provides an hourly forecast for PM concentrations, which is operationally verified against the AURN network. However, the precipitation forecast data available for this study has a coarser temporal resolution of 6 h accumulations. Therefore 6-hour mean PM observations and forecasts were calculated, resulting in mean PM forecast error

values every 6 hours to match the precipitation accumulations.

Rainfall observations are from the ground-based WMO gauge network. The instruments are well maintained and provide long-term, quality controlled precipitation data. Gauge data is more accurate in measuring precipitation which actually reaches the ground than measurements from radar, which may include precipitation aloft. Radar data is also susceptible to noise from e.g. wind farms or clutter, although quality control does well to remove data defects. There are different benefits for using either type of precipitation measurement as ‘truth’ as outlined in section 3.6.3.3, but one should not be used to verify the other due to e.g. representativeness issues (Ciach and Krajewski, 1999). The merit of using a neighbourhood verification technique within HiRA is that it compensates for the sparseness of the gauge network. A gridded radar dataset is more suited to a verification method such as the Fractions Skill Score. It also provides area-averages which are not strictly consistent with localised rainfall accumulations specific to a site in question.

Therefore, rainfall gauge observations were chosen for this evaluation in order to keep the consistency with point-based forecast verification work discussed in section ???. WMO rainfall gauge locations were matched with nearest AQ sites which measure PM: 25 urban and rural background sites were sub-sampled for the evaluation. The study period for this section is 22/11/2016 to 22/02/2017. This is a different time period than the one used for the analysis in chapter 3, partly due to convenience of the data format which was used within HiRA, and partly due to the fact that AQUM would have experienced model upgrades since the 2015 data used in chapter 3, which one might expect would mean that the model is at its most accurate. Using a different winter period should make little difference to the statistical analysis of relationships between errors in precipitation and PM forecasts. This is because the PM sites of interest are background (urban and rural), which are representative of the ambient concentrations, and should therefore be unaffected by local emissions e.g. from vehicle transport. A winter season was chosen rather than the summer because of the generally higher frequency of elevated PM episodes which occur during winter.

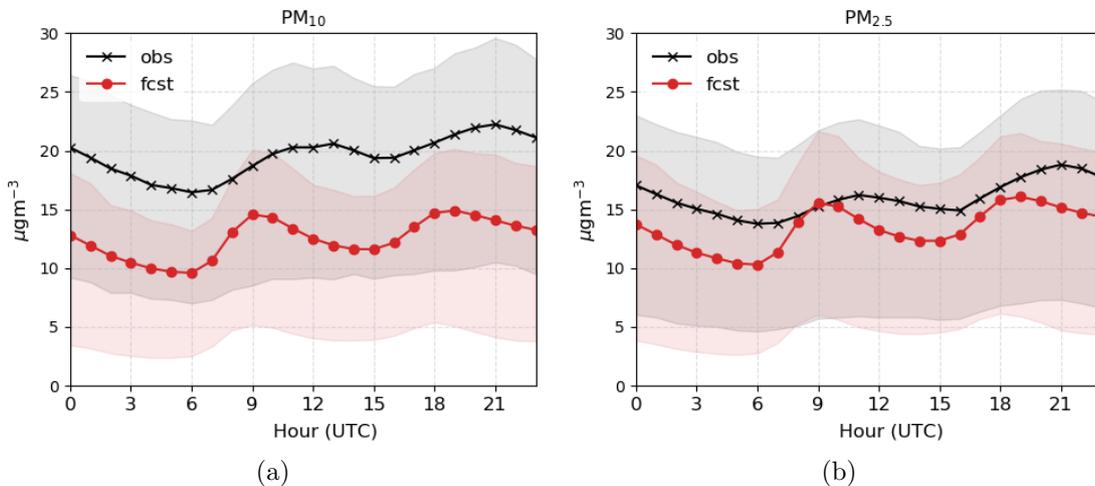
First, PM data for the study period is described with diurnal cycles as per section 3.5.3. For comparison with precipitation forecast error, first the mean error in both variables at point locations is used in order to seek out any obvious relationship for this study period. Zero-precipitation events (either observed or forecast) are filtered out due to their dominance in the dataset. Then, the BS is used for analysis of forecast error correlations between precipitation and PM. When using the BS, one has to be cautious in

choosing an appropriate threshold for the case of ‘no rain’, as rainfall gauges may register some amount of rainfall  $> 0$  over a given time period - e.g. an hour - where the rainfall may just be highly localised and brief. Because rain gauge measurements are recorded in discrete values of 0.2 mm/6h, the term ‘no rain’ is re-defined as accumulation amount  $R < 0.2$  mm/6h.

#### 4.6.2 Relationships between forecasts and observations

Before analysing relationships between forecast errors of PM and precipitation, it is useful to know how the diurnal cycles of the forecasts and observations compare within the study period. Here, it will also be confirmed whether an anti-correlation between precipitation and PM forecasts and observations exists for the study period.

Figure 4.9 depicts the DJF diurnal cycle in observed and forecast concentrations of PM<sub>10</sub> and PM<sub>2.5</sub>, averaged over 25 urban and rural background sites in the UK. It is clear that the forecast concentrations for the coarse component (figure 4.9a) are similar to the fine component (figure 4.9b), whereas the observed concentrations are more different between the two sub-groups. In other words, despite the difference in observations between PM<sub>10</sub> and PM<sub>2.5</sub>, the mean diurnal profile of the forecasts looks similar because PM<sub>2.5</sub> dominates the PM<sub>10</sub> in the model. This means that the PM<sub>10</sub> mean forecast error is  $-6.4 \pm 2.3 \mu\text{gm}^{-3}$ , while for PM<sub>2.5</sub> it is  $-1.7 \pm 2 \mu\text{gm}^{-3}$ . There is a greater negative bias in the PM<sub>10</sub> forecast because in AQUM, a significant proportion of PM<sub>10</sub> is made up of the fine particulates, and any generation of secondary aerosol is always of the PM<sub>2.5</sub> sub-



**Figure 4.9:** Average diurnal cycle of observed (black, crosses) and forecast (red, circles) concentrations of (a) PM<sub>10</sub> and (b) PM<sub>2.5</sub> at 25 locations, for DJF 2016 / 2017. Shading represents the 25th and 75th percentiles.

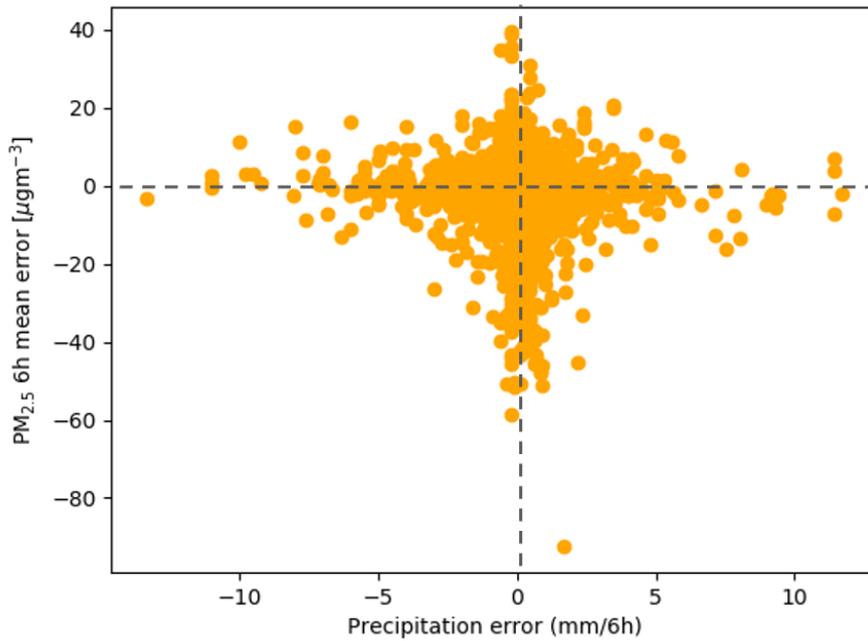
group. Furthermore, some coarse component (e.g sodium nitrate) is not represented in the CLASSIC scheme used in AQUM, while other species such as sea salt are not emitted over land. Any inaccuracies in the representation of other coarse component sub-groups, such as dust, pollen or smoke, may further increase the difference between forecast  $PM_{2.5}$  and  $PM_{10}$ . This means that the  $PM_{10}$  forecast more closely resembles the  $PM_{2.5}$  forecast than it does observations.

There is also a potential effect of the boundary layer representation on the differences between observed and forecast diurnal cycles of PM, previously suggested in section 3.5.3. It is expected - and indeed observed - that the PM concentrations rise during the morning ‘rush hour’ (7 - 9 UTC). However, the forecast peak in PM concentrations is higher relative to that observed. A possible reason for this could be because of a delay in the increase of the boundary layer depth in AQUM (which will be discussed in further detail in section 5.4.2), which may result in the trapping of pollutants within a layer which is too shallow, hence a sharper rush-hour peak. Furthermore, dilution of pollutants associated with entrainment of cleaner air aloft doesn’t occur until 1 - 2 hours after the modelled boundary layer starts growing.

6h-accumulations of precipitation have a log-normal distribution due to the frequency of zero and small-valued accumulations, and as such it is useful to log-transform the data and / or remove zero-precipitation events from the data sample. In DJF 2015 / 2016, PCC for the observed 6-hourly mean  $PM_{2.5}$  and log-transformed precipitation is  $r = -0.254$ , whereas for the forecast values the relationship is weaker, at  $r = -0.134$ . These results suggest that the anti-correlation of forecast 6-hour precipitation accumulations and  $PM_{2.5}$  in AQUM is not as strong as the observed relationship. There could be many reasons for this; one of them being the systematic under-estimation of precipitation, which could result in insufficient washout of particulates in AQUM.

### 4.6.3 Comparison of errors in $PM_{2.5}$ and precipitation

The next step was to compare the 6-hour precipitation accumulation forecast error with the 6-hour averaged  $PM_{2.5}$  mean error. In section ??, a first-order relationship between PM and precipitation was discussed. As a reminder, because both  $PM_{10}$  and  $PM_{2.5}$  are generally under-estimated, under-prediction of rainfall would wash out less PM from the atmosphere than is observed, thus any negative PM error would be made less negative (i.e. a negative error correlation). Similarly, if there is too much rainfall forecast, it would be expected that there is too much wet deposition of PM, thus worsening the

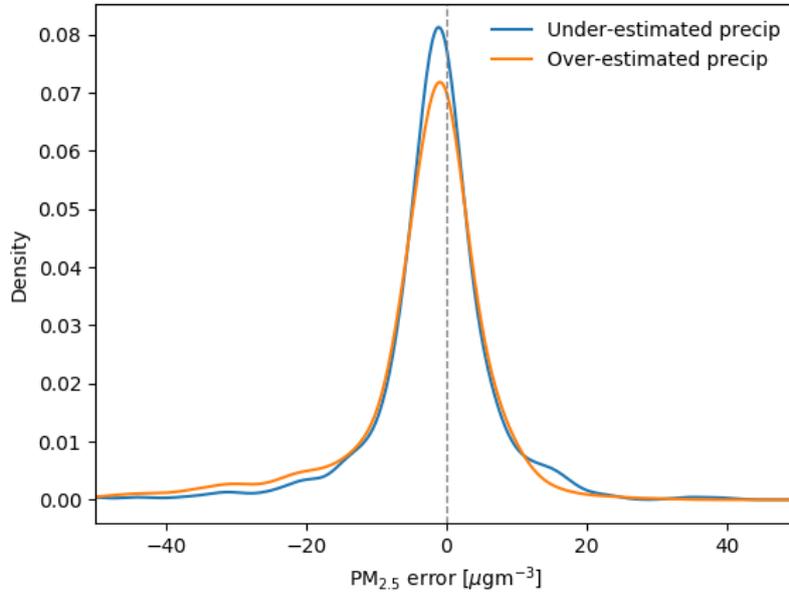


**Figure 4.10:** Relationship between the 6-hour averaged  $\text{PM}_{2.5}$  mean error and precipitation mean error, calculated at 25 sites for every 6-hour window in DJF 2016 / 2017.

negative bias. However, unlike the total oxidant and wind speed evaluation, one would not expect to see any diurnal relationship in the precipitation and PM forecast errors because although PM may have a diurnal cycle due to emissions, the precipitation does not.

Figure 4.10 shows a direct comparison of the two error types, composited over 25 sites in the evaluation. All zero-precipitation events (either observed or forecast) have been filtered out because the large number of non-precipitation events was dominating the data. This originally resulted in a dominant cluster of scatter points around  $x = 0$  on the plot, because non-precipitation events tend to be forecast correctly in DJF (as shown in section 3.6.3.1, where 78% of forecast data and 74% of observations in DJF 2015 were zero-valued). However, the remaining data which forms the scatter plot shown in figure 4.10 does not show any obvious relationship between the precipitation and  $\text{PM}_{2.5}$  forecast errors. In fact, there is significant clustering of data around both the  $x$  and  $y$  axes, i.e. near-zero precipitation error and near-zero  $\text{PM}_{2.5}$  error. This is counter-intuitive because, as discussed earlier, one would have expected a precipitation forecast with a large error to have a strong influence on the PM forecast error. Perhaps a case study of a particularly poor precipitation forecast episode is required to gain more insight into how it affects the PM forecast, instead of analysing the whole study period as a whole.

However, the mean error in precipitation has very little bias (mean bias =



**Figure 4.11:** The distribution of 6-hourly  $PM_{2.5}$  forecast error, partitioned by cases of over- and under-estimation of 6-hour precipitation at 25 locations (DJF 2016 / 2017).

0.02 mm/6h) and is centred on zero. Therefore it makes sense to partition the  $PM_{2.5}$  error distribution based on whether the corresponding precipitation forecast is under- or over-estimated. Figure 4.11 demonstrates that there is very little difference in the  $PM_{2.5}$  error distributions when the precipitation is over- or under-estimated. In both cases, the distribution peak is  $< 0$ , but it is clear that there is no obvious relationship between the precipitation and PM forecast errors. Given the above analysis, the use of BS as a neighbourhood technique will be explored next to learn about whether the spatial error in precipitation is influencing the  $PM_{2.5}$  forecast error - a relationship that would not be obvious at the grid-scale.

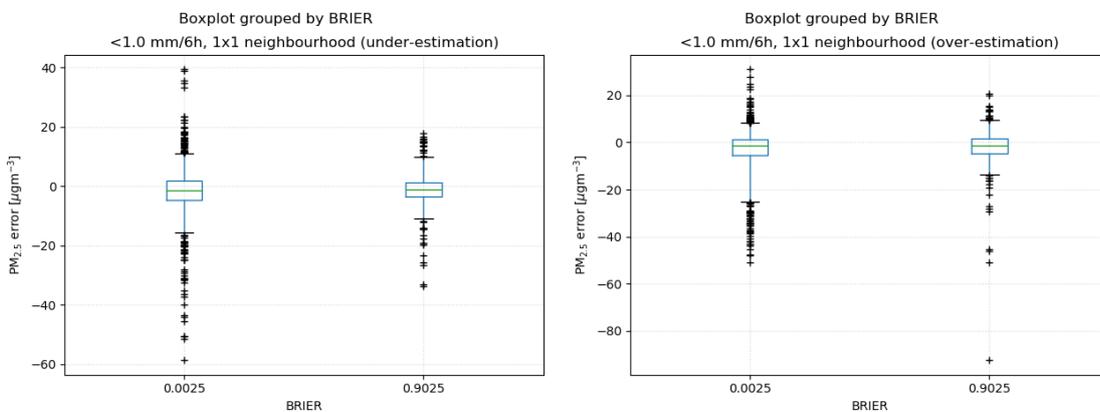
#### 4.6.4 Precipitation BS and its relationship with PM

Finally, this section presents the methodology and results of comparing the precipitation forecast error in terms of the BS against the PM mean error. Using the thresholds of 6 h rainfall accumulation  $< 0.1$ ,  $< 0.5$ ,  $< 1.0$ ,  $< 2.0$  and  $< 4.0$  mm/6h, BS were calculated for each 6 h window over the neighbourhoods of 1x1, 3x3, 5x5, 7x7 and 9x9. This was repeated for all 25 locations.  $PM_{2.5}$  forecast error in this case is treated instantaneously, i.e. the value of the forecast error was taken at the end of every 6 h window instead of calculating the 6 h mean error. Due to the hourly variability of  $PM_{2.5}$ , it is appropriate to consider the instantaneous errors rather than averaging over a temporal window.

#### 4.6.4.1 Precipitation BS at grid-scale

Figure 4.12 shows the 6 hr precipitation BS at grid-scale, for the threshold of  $R < 1.0$  mm/6h. This relatively low threshold is evaluated because using a higher threshold such as 4 mm/6h means that few of the BS values are non-zero due to the rarity of extreme rainfall. As a reminder,  $BS = 0$  is the perfect score (i.e. both forecast and observation falls either within (‘correct forecast’) or outside (‘correct negative’) the specified threshold value); meanwhile  $BS = 1$  indicates maximum possible error (i.e. only one of either the forecast or observation is larger than threshold value). BS takes on a binary value at the grid-scale, because the forecast grid-box can only either satisfy the threshold criterion or not. It is impossible to distinguish whether a non-zero BS value is due to an under- or over-estimation of precipitation, because it only indicates whether the threshold criterion was met. As such, the data has been split into two cases within figure 4.12: under-prediction (sub-figure 4.12a) and over-prediction (sub-figure 4.12b) of precipitation.

Figure 4.12a depicts the distribution of  $PM_{2.5}$  error when precipitation is under-predicted. In this case, it does not matter by how much; it only matters that the precipitation forecast error is negative. If the outlier values are considered, there is large variability in  $PM_{2.5}$  forecast error (y-axis) when the precipitation BS is small; i.e. when both the forecast and observation are within the 1.0 mm/6h threshold. Some of the largest  $PM_{2.5}$  under- and over- predictions occur in this scenario, as the error ranges from -60 to  $40 \mu\text{gm}^{-3}$ . Where the cluster of points sits at  $BS = 1$ , this indicates that either the forecast or observed precipitation is larger than 1.0 mm/6h (but is still under-predicted). In this scenario,  $PM_{2.5}$  error has smaller variability (range from -35 to  $20 \mu\text{gm}^{-3}$ ).



(a) Under-estimated precipitation

(b) Over-estimated precipitation

**Figure 4.12:** 6h precipitation BS for grid-scale forecasts and  $PM_{2.5}$ , aggregated over 25 locations and separated into (a) under-predicted precipitation and (b) over-predicted precipitation. Whiskers are 5th and 95th percentile.

In section 3.7.2, it was shown that an under-estimation of precipitation is followed by a reduction of PM<sub>2.5</sub> forecast error. The grid-scale evaluation in figure 4.12a supports the above statement, as the PM<sub>2.5</sub> error inter-quartile range (IQR) is smaller when BS = 0.9025 (“bad” forecast) than when BS = 0.0025 (“good” forecast). However in both cases, median PM<sub>2.5</sub> error is negative ( $-1 \mu\text{gm}^{-3}$ ).

In figure 4.12b, the effect of over-predicting the precipitation is not very clear. The PM<sub>2.5</sub> forecast error IQR is almost the same regardless of whether BS is low or high, and the median error in both cases is negative.

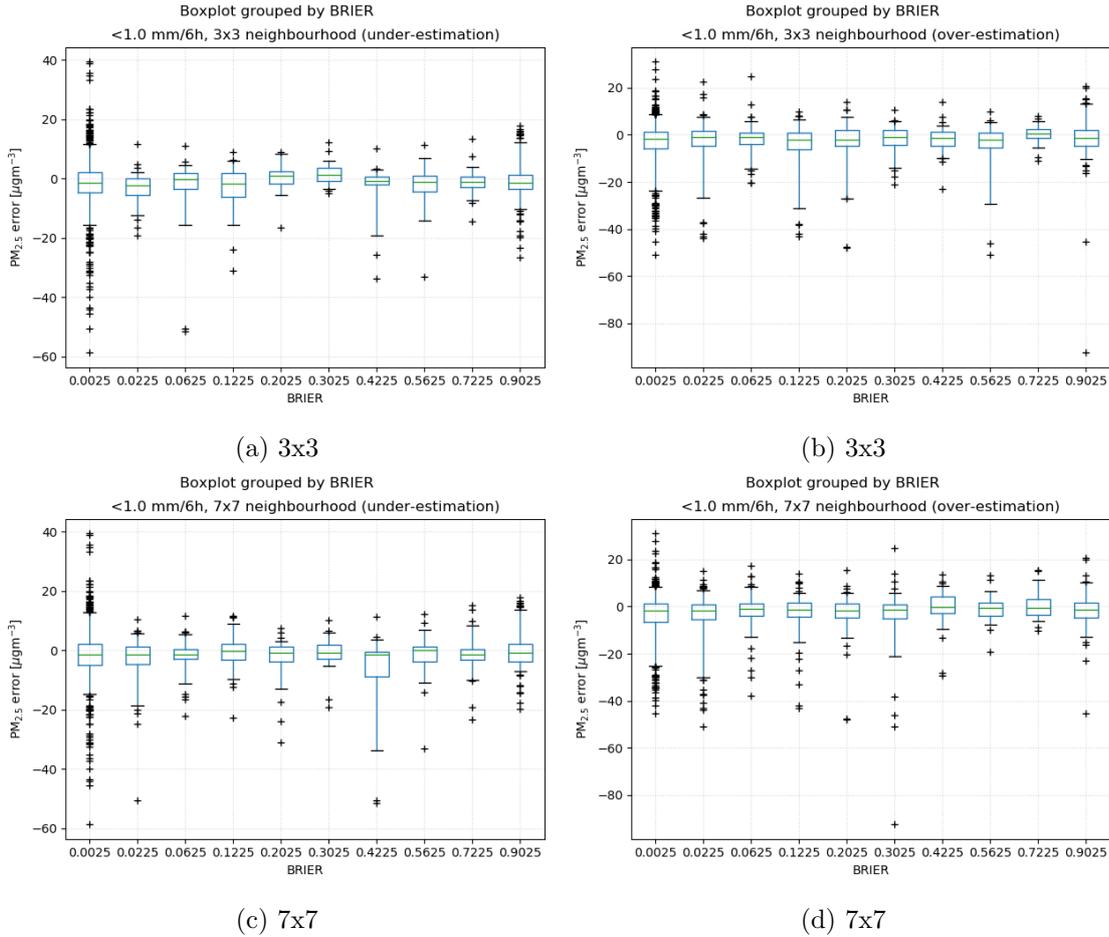
The result is similar for threshold values of 0.5 mm/6h and 2.0 mm/6h, therefore the equivalent plots are not shown.

#### 4.6.4.2 Precipitation BS over larger neighbourhoods

When a neighbourhood around the precipitation observation is introduced, BS can take on more values due to its probabilistic nature. Physically, a ‘mid-range’ BS value means that only some of the neighbourhood satisfies the threshold criterion. When a 3x3 or larger forecast neighbourhood is taken into account, a pseudo-ensemble forecast has been created under the assumption that each grid-box is equally-likely to be a correct forecast.

Figure 4.13 shows the 3x3 and 7x7 neighbourhood equivalent of figure 4.12. The panels have been split into under- and over-estimation of precipitation (as it happens at the grid-scale). It appears that using a 3x3 and 7x7 neighbourhood slightly reduces the spread of the PM<sub>2.5</sub> error at either end of the BS range in both cases, e.g. the range of PM<sub>2.5</sub> error values for BS = 0.9025 is smaller in panel 4.13c than in panel 4.13a. This is mainly due to the migration of the large negative outlier values at BS = 0.9025 (present in the 3x3 neighbourhood) which are not there in the 7x7 neighbourhood. This makes sense, because although the bin-size has been kept the same in both cases (each probability bin is 0.1), evaluating the forecast over more grid-boxes enables the BS to take on more of the ‘mid-range’ values, thus some of the large PM<sub>2.5</sub> forecast errors appear to migrate over to the mid-range of the precipitation error values. However, increasing neighbourhood size for the precipitation evaluation has little effect on the IQR of PM<sub>2.5</sub> error, and the median values remain near-zero or negative for most possible BS values.

Of course, a different threshold could be used other than 1.0mm/6h to further analyse the impact of using a neighbourhood verification for the precipitation forecast. It would be expected that using a smaller threshold would decrease the number of precipitation

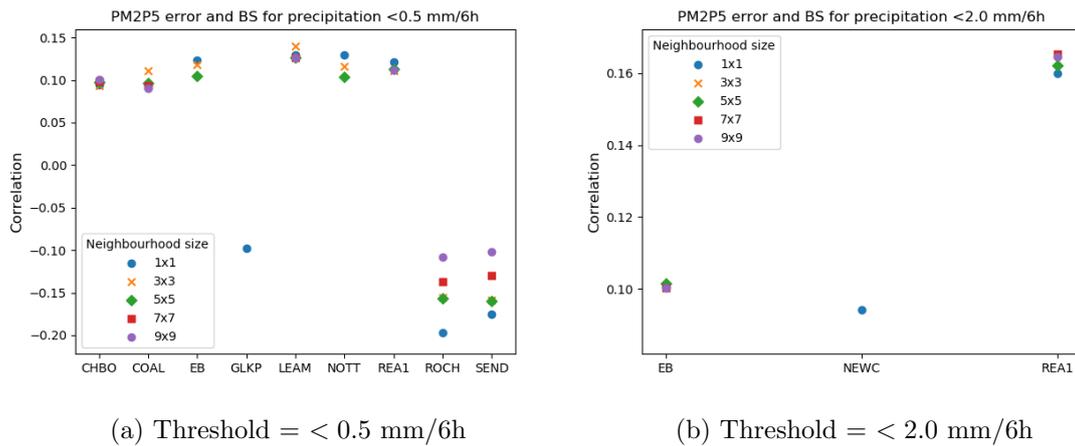


**Figure 4.13:** The same as figure 4.12, but for a 3x3 and 7x7 neighbourhood. Panels (a), (c) are based on under-estimation of precipitation at grid-scale; panels (b), (d) are based on over-estimation of precipitation at grid-scale.

forecasts and observations which match the threshold criterion. This may increase the PM<sub>2.5</sub> forecast error variability towards the BS = 1 end of the scale for grid-scale precipitation forecasts, as well as increasing the abundance in the mid-range BS for larger neighbourhoods. The opposite would be true for higher thresholds, as more forecasts would satisfy the criteria and therefore BS would be closer to 0 across all neighbourhood sizes.

#### 4.6.4.3 Site-specific BS evaluation

The aim of the neighbourhood evaluation is to seek systematic behaviours and error correlations within the entire dataset. However, the evaluation presented in section 4.6.4.2 showed no systematic change in PM<sub>2.5</sub> error when the BS was used to evaluate precipitation over varying-size neighbourhoods. The aim of this section is to determine the extent of specificity within the individual sites, and whether this is affecting the



**Figure 4.14:** Site-based error correlations between PM<sub>2.5</sub> forecast error and precipitation BS, based on a threshold of (a) < 0.5 mm/6h and (b) < 2.0 mm/6h. Only showing sites and neighbourhood sizes where correlations are statistically significant ( $p \leq 0.1$ )

overall picture. For each site, PCC of the BS and PM<sub>2.5</sub> forecast error time-series was calculated at varying neighbourhood sizes and BS thresholds. Only sites whose results are statistically significant are shown in figure 4.14.

The most obvious comparison between figures 4.14a and 4.14b is that the number of sites with significant correlations decreases as the BS threshold increases. This is reasonable: when too much data satisfies the high ( 2.0 mm/6h) threshold criterion, the pool of BS scores is saturated with low (‘good’) scores. This means that there is little variability in the precipitation error, which is not comparable against the large variability in the PM<sub>2.5</sub> forecast error. Consequently, only three sites display statistically significant correlations in figure 4.14b, though their values are low (range of  $r$  is 0.1 to 0.18 across all neighbourhood sizes).

For the lower threshold of 0.5 mm/6h, 6 sites exhibit statistically significant positive error correlations (which range between 0.08 to 0.14), and 3 sites have negative correlations (ranging between -0.2 to -0.1). Within these results, the strongest correlations seem to occur at the grid-scale and 3x3 neighbourhoods (sites ‘EB’, ‘COAL’, ‘GLKP’, ‘LEAM’, ‘NOTT’, ‘REA1’, ‘ROCH’ and ‘SEND’). However, even the highest PCC values are low, indicating that the relationship in PM<sub>2.5</sub> and precipitation forecast errors is weak, and the sign of the correlation is site-dependent.

#### 4.6.5 Conclusions: neighbourhood evaluation of relationships in precipitation and PM errors

The present section presented an evaluation of (1) the skill of the precipitation forecast and the PM<sub>10</sub> and PM<sub>2.5</sub> forecasts for the study period of DJF 2016 / 2017, and (2) attempted to evaluate the relationships between the forecast errors in terms of a probabilistic metric and varying neighbourhood sizes. Similar to section 4.5, the aim of introducing a neighbourhood verification metric for the precipitation forecast was to further our understanding of the PM forecast errors associated with spatial errors in precipitation. At the beginning of this chapter, the question was asked whether the forecast verification technique should consider the skill of the precipitation forecast in the neighbourhood of PM<sub>10</sub> and PM<sub>2.5</sub> measurement site, rather than only at the point location? The answer to this question is not simple, because it appears that this depends on how heavy the rainfall is. In conclusion:

- Considering a neighbourhood around the measurement site and using the BS shows some reduction in precipitation forecast error, but only for rainfall accumulation thresholds of < 1.0 mm/6h and less. Otherwise, the reduction in forecast error due to using a larger pseudo-ensemble is negligible because almost all of the grid-points considered meet the threshold, resulting in a reduced variability in the BS.
- There exist correlations between precipitation BS and mean forecast error of PM<sub>10</sub> or PM<sub>2.5</sub>, but they are small.
- PM<sub>2.5</sub> forecast error has a similar distribution when the precipitation forecast errors are partitioned by sign. This translates to very little difference in PM<sub>2.5</sub> errors when the precipitation BS is small or large.
- Where relationships between precipitation BS and PM<sub>2.5</sub> exist, they are weak and their sign is site-dependent. This suggests non-systematic behaviour.

The purpose of this section was to seek out systematic relationships between the PM forecast error and the precipitation error, when the spatial constraint is relaxed. However, the results are not systematic, indicating that the relationship between PM and precipitation forecast errors is not as simple as initially suggested. One reason could be that errors in the PM forecast are more strongly influenced by factors such as emissions (in the vicinity of specific measurement sites) and transport than precipitation errors. Wind speed and direction could be another source of meteorological forecast error influence on

the PM error, which has not been evaluated in this work. We have already seen from the point-based analysis in chapter 3 that the magnitude of mean precipitation error in AQUM is not large; perhaps this simply means that precipitation is thus not a considerable factor in the magnitude of PM forecast error. Perhaps an evaluation involving a case study would be more insightful. A series of poor precipitation forecast events could be evaluated, alongside their individual impact on the PM forecast error at a single site, calculating a skill score in reference to a baseline forecast. This approach could be more insightful than conducting a study which involves averaging over many sites or a long time-period.

## 4.7 Chapter summary and conclusions

In this chapter, two neighbourhood verification metrics were used to evaluate the role of meteorological forecast error on air quality forecast error within AQUM. The BS, used for probabilistic and ensemble forecasts, is useful when considering a range of thresholds to evaluate forecasts in a binary manner, pertaining to a contingency table.

First, the BS is used to evaluate wind speed forecasts in AQUM for a range of thresholds in order to find out whether overestimating wind speeds affects the pollutant forecast. A night-time overestimation of wind speed was identified, and it was hypothesised that it could be related to the forecast errors in night-time total oxidant, as well as  $O_3$  and  $NO_2$  separately. The hypothesis that night-time overestimation in wind is influencing the forecast of  $O_3$  and  $NO_2$  was supported by the results presented in this chapter, which evaluated forecast errors during the JJA 2017 period. The results showed a positive significant correlation between night-time (00 - 05 UTC) surface wind speed forecast error and both  $O_3$  and total oxidant forecast errors at urban background sites only. The wind speed forecast error was anti-correlated with  $NO_2$  forecast error at both urban and rural sites.

Extending the neighbourhood verification to the AQ forecasts, the Continuous Ranked Probability Score (CRPS) was used to evaluate  $O_x$ ,  $O_3$  and  $NO_2$  forecasts. Using this approach was an exciting opportunity, as it is novel to AQ verification. Results presented in this chapter demonstrate that there are potential benefits to employing this verification technique in the future, as resolution of AQ forecasts increases. All three AQ variables showed significant correlations in their CRPS values against wind speed error across all neighbourhood sizes at urban sites.  $O_x$  showed the strongest correlations in a 5x5 neighbourhood, for  $O_3$  is was a 7x7 neighbourhood, meanwhile  $NO_2$  CRPS had strongest correlations with wind speed forecast at a 3x3 neighbourhood. This suggests

that the  $\text{NO}_2$  forecast is affected more by local wind speed errors than errors farther away within urban areas; meanwhile  $\text{O}_3$  forecast errors have more dependence on wind speed errors farther afield.

A reason for this could be that if wind speed is stronger than e.g.  $5 \text{ ms}^{-1}$ , it can contribute to vertical exchange of air masses from the residual layer into the boundary layer. Given that the  $\text{O}_3$  forecast error shows a positive correlation with night-time wind speed values below  $6 \text{ ms}^{-1}$ , the results suggest that boundary layer entrainment could be acting as an additional reservoir of  $\text{O}_3$  from the night-time residual layer, affecting its forecast at the surface. This effect could further dilute  $\text{NO}_2$  concentrations, which are also anti-correlated with wind speed. This could be due to enhanced dispersion from local sources within and around urban areas.

The second part of this chapter focused on neighbourhood verification using BS for the forecast error relationships between precipitation and  $\text{PM}_{2.5}$  and  $\text{PM}_{10}$ . The results show that this relationship is weak. As for the precipitation forecast alone, there is some improvement in forecast skill when a pseudo-ensemble is created by including a neighbourhood around the point location instead of just the grid-box. This improvement is only visible for rainfall accumulation thresholds of  $1.0 \text{ mm}/6\text{h}$ . This suggests that for smaller rainfall thresholds, there may be merit in evaluating the precipitation forecast by using a neighbourhood, however it does not have a significant effect on the PM forecast error. The analysis involved a certain amount of averaging in space and time, therefore conducting some case studies into poor precipitation forecasts and / or poor PM forecasts at specific locations could help us understand more about the influence of precipitation error within a neighbourhood of the PM measurement site than we have learnt from the present evaluation. Another point to consider is the varying spatial scales of precipitation: convective rainfall may be more responsive to a neighbourhood verification method (in terms of assessing the accuracy of localised rainfall and its impact on nearby PM receptor sites), than large-scale precipitation as a front comes through at a point location. Furthermore, wind speed and direction could be a secondary variable affecting the PM forecast error in relation to the precipitation.

# 5. Process-based evaluation of boundary layer development and forecast errors in O<sub>3</sub> and NO<sub>2</sub>

## 5.1 Introduction

Near-surface O<sub>3</sub> contributes to the formation of photochemical smog and can trigger breathing difficulties. In order to make accurate ground-level O<sub>3</sub> forecasts, it is important to understand the processes that control the magnitude and evolution of surface O<sub>3</sub> concentrations. As highlighted in chapter 1, sources of O<sub>3</sub> and its precursors may originate from emissions near the surface, from long-range inter-continental transport, or from the free troposphere acting as a reservoir for O<sub>3</sub> produced the previous day (e.g. [Purvis et al., 2005](#)). This chapter examines the role of morning exchange of air masses between an O<sub>3</sub>-rich residual layer and the boundary layer in determining diurnal profiles of surface O<sub>3</sub> concentrations, alongside one of the other atmospheric species key to its life-cycle: NO<sub>2</sub>. The Lagrangian atmospheric dispersion model NAME (Numerical Atmospheric-dispersion Modelling Environment) is used to perform an idealised study of the impact of boundary layer characteristics on downward entrainment of pollutant from the decoupled night-time residual layer. The sum of O<sub>3</sub> and NO<sub>2</sub> will henceforth be referred to as the total oxidant, O<sub>x</sub>.

Vertical mixing is a physical process which enables the movement of atmospheric constituents - including O<sub>3</sub> molecules - within as well as above the planetary boundary layer (PBL). A temperature inversion at the top of the PBL usually prevents air masses from mixing across the boundary layer top, unless a strong up- or down-draught enables the mixing. This physical process is called ‘entrainment’ (i.e. the movement of air masses from the free troposphere or residual layer down into the PBL) or ‘detrainment’ (i.e. ventilation of air masses) - see also section 2.2.1.5. Maximum average entrainment veloc-

ities are reached during the morning period of the PBL development, as the night-time boundary layer (NBL) transitions into the daytime mixed layer. Buoyant convective thermals originating near the ground rise into the troposphere and, as they rise, they may overshoot into the residual layer, mixing down air upon returning to the boundary layer (Stull and Ahrens, 2000; Trousdell et al., 2016).

At certain times of the day, atmospheric species such as  $O_3$  may have lower concentrations within the boundary layer than above it. This is because as the mixed daytime boundary layer transitions into the shallower NBL,  $O_3$  is deposited to surfaces or undergoes titration by NO (Logan et al., 1981). When the daytime boundary layer erodes, a ‘residual layer’ decouples from the diminishing surface layer and some  $O_3$ -rich air masses end up residing in the separated residual layer overnight. Because the residual layer is decoupled from the PBL,  $O_3$  in this region undergoes no physical deposition processes and fewer chemical reactions due to a lower abundance of  $NO_x$  in the free troposphere. Thus the  $O_3$  lifetime is longer in this region than near the surface (Brown and Stutz, 2012). A positive  $O_3$  concentration gradient can therefore exist between the boundary layer and the residual layer in the early hours of the morning.

When air masses containing a scalar quantity (e.g. water vapour, or an atmospheric species like  $CO_2$  or  $O_3$ ) are entrained from the residual layer, they will either enhance or dilute concentrations within the boundary layer and even near the surface, depending on the concentration gradient. For example, observational studies show that entrained air masses from the residual layer with low  $CO_2$  concentrations relative to the boundary layer dilute the surface concentrations, particularly during the ‘morning transition’ period (de Arellano et al., 2004). For  $O_3$ , there exists observed evidence of the opposite effect due to a positive gradient across the boundary layer top: entrainment of  $O_3$ -rich air can enhance surface concentrations both on land (Jaffe, 2011) and over the marine boundary layer (Parrish et al., 2010).

While many studies concentrate on accurate representations of local photochemical production and loss processes of  $O_3$ , the contribution of existing residual layer  $O_3$  to surface background concentrations is often overlooked. For example, in an extensive AQ-model inter-comparison framework in Europe and North America, (Solazzo et al., 2013) attributed boundary layer  $O_3$  error mostly to surface emissions, transport and photochemistry. In a subsequent study, Solazzo and Galmarini (2016) conclude that  $O_3$  bias due to input fields (e.g. emissions and boundary conditions) has the largest influence on the mean squared error and is likely to be the primary focus for improving AQ

modelling systems in the near future. However, their  $O_3$  error decomposition technique reveals that modelling processes including boundary layer depth are also important for the diurnal variance component of the overall mean squared error metric. Another case study by [Travis et al. \(2016\)](#) evaluating the GEOS-Chem model with ozonesonde data in the Southeast US concludes that although  $NO_x$  emissions inventories are an important source of error, excessive boundary layer mixing is a major source of positive model bias for daytime surface  $O_3$ .

Some model studies have been conducted in recent years to estimate the contribution of entrained  $O_3$  and  $O_x$  to surface concentrations. For example, [Kaser et al. \(2017\)](#) compare observed surface  $O_3$  concentrations against concentrations modelled by WRF-Chem. By calculating an Eulerian  $O_3$  budget, they show that growth of the boundary layer through entrainment before 11 LT (local time) leads to an increase in boundary layer  $O_3$  (and  $O_x$ ) concentrations. Further growth after 11 LT has the opposite effect on  $O_3$  concentrations due to the reversal of the  $O_3$  vertical gradient across the boundary layer top. The study concludes that morning entrainment contributes  $\sim 4.8 \text{ ppbv h}^{-1}$  ( $\sim 10 \mu\text{gm}^{-3}\text{h}^{-1}$ ) to surface  $O_3$  concentrations during the morning transition period. For  $O_x$ , this value is  $\sim 3.3 \text{ ppbv h}^{-1}$  ( $\sim 7 \mu\text{gm}^{-3}\text{h}^{-1}$ ). The findings of the [Kaser et al. \(2017\)](#) study form a useful base for verifying the results of the Lagrangian simulations presented in this chapter, and will be further discussed in section [5.6.2](#).

It is well-established that certain synoptic meteorological conditions, such as anti-cyclonic conditions associated with clear skies and high surface temperature are favourable to  $O_3$  production, while minimising ventilation because of stagnant conditions due to subsidence ([Mukammal et al., 1982](#)). Boundary layer structure can be strongly affected by synoptic-scale circulations. [Sinclair et al. \(2010\)](#) conduct an idealised experiment to demonstrate that the sign and magnitude of heat fluxes in the presence of a passing cold front between a high- and low-pressure region dictates the boundary layer structure. For example, heat fluxes behind a passing cold front are positive, leading to deep and convection-driven boundary layers due to an unstable potential temperature profile. Due to the positive heat fluxes, entrainment of air from the free troposphere to the boundary layer occurs, regardless of whether winds throughout the boundary layer are strong or weak. Meanwhile in the warm sector ahead of the cold front, heat fluxes are negative and the boundary layer is shallow (or moderately deep if wind speeds are high) and shear-driven, with warm air initiating large-scale ascent (i.e. a low pressure system). This means that in the warm sector of the cyclone, boundary layer conditions are generally

neutral or stable.

Therefore, a factor which could affect the performance of an AQ forecast model is its responsiveness to synoptic-scale meteorological conditions. Modelling studies such as [Eder et al. \(2006\)](#) demonstrate that the Community Multiscale Air Quality (CMAQ) model over-predicted O<sub>3</sub> concentrations in August 2004 during periods of cyclonic and cloudy conditions, but performed well during clear-sky, anti-cyclonic conditions. In the case of CMAQ, too much O<sub>3</sub> forecast near the tropopause was transported down to the surface by downdraughts associated with convective clouds. It is therefore worth conducting a process-based investigation of the impact of synoptic conditions on AQUM's O<sub>x</sub> forecast.

## 5.2 Chapter aims

In chapter 2, a 1 - 2 hour offset between forecast and observed O<sub>3</sub> concentrations was identified, but point-based verification could not tell us much about the source of the timing discrepancy. This chapter investigates the influence of potential errors in the modelling of boundary layer processes on forecast errors in surface O<sub>x</sub> concentrations. Specifically, the chapter concerns the delay in the onset of modelled boundary layer growth and entrainment, and the relation of those processes to the timing offset in morning O<sub>x</sub> increase. Having already evaluated O<sub>3</sub> and NO<sub>2</sub> forecasts against observations in chapter 2, the evaluation is extended to include O<sub>x</sub> in urban background regions, for June 2017.

In the presence of a positive concentration gradient across the boundary layer top, the process of morning entrainment should act to enhance the boundary layer O<sub>3</sub> concentrations, and subsequently the amount of O<sub>3</sub> and NO<sub>2</sub> forecast at ground-level. In this chapter, a semi-idealised tracer-release experiment has been designed and carried out to quantify the proportion of tracers which are entrained from the free troposphere into the boundary layer and which eventually reach the surface, against 'locally' emitted tracers, i.e. those originating and always residing near the ground. The experiment implements a novel tracer tagging methodology using the Lagrangian off-line dispersion model NAME III ([Jones et al., 2007](#)) to investigate vertical exchange. The tracer has been constructed to simulate aspects of the behaviour of tropospheric O<sub>x</sub> without the complication of chemical transformations. The novelty of this work is that - to the best of my knowledge - a Lagrangian process-based evaluation of top-down mixing processes in terms of air quality modelling has not been done before. Although boundary layer top entrainment and mixing has been widely studied for the purposes of improving the representation of e.g. convection and the 'grey zone' in NWP, there is little information in

the way of implications for air pollution modelling. A benefit of using a Lagrangian dispersion model to run this experiment is that the dynamics are grid-independent, therefore allowing a more general investigation into the influence of top-down mixing on modelled atmospheric species, which is not bound to a particular air quality model.

Another possible route of evaluation would be to use a simple box model, as in the thesis of Peake (2012), where the influence of a coastal outflow diurnal cycle on  $O_3$  is assessed. Box model studies are useful for simplifying the research question and breaking it down into known parameters (i.e. sources and sinks of tracers, advection into or out of the box, ventilation, etc). Box model studies are often used for studying chemical emissions, as the equilibrium budget analysis of the variable in question (e.g. a specific chemical species) is computationally inexpensive. While using a box model to represent and evaluate a particular process can be useful for understanding the generalities, it is usually an idealised configuration and therefore not suitable for representing specific scenarios. Because a systematic evaluation is required of the process of entrainment at a number of locations (rather than just with one general, simplified representation), a Lagrangian dispersion model was chosen as the best approach.

This chapter aims to answer the following specific scientific questions:

1. How does the diurnal evolution of  $O_3$  depend on the rate of growth in modelled boundary layer depth?
2. What is the contribution of non-locally produced  $O_x$  to local surface concentrations during stagnant or unsettled meteorological conditions?

Background information and experimental details are set out in this chapter as follows:

- Section 5.3 is an introduction to the three models used in this study: the off-line NAME model, the UKV and the UM Global configurations of the Met Office UM;
- Section 5.3 also introduces the boundary layer scheme used in UM Global and the UKV, as well as the land-surface representation;
- Comparison of physical parameters in the NWP, such as surface sensible heat flux, boundary layer depth and vertical velocity, follow in section 5.4;
- Methodology for initiating semi-idealised tracer release experiments with NAME in order to study entrainment of free-tropospheric air masses are outlined in section 5.5, and results follow in section 5.6;

- A summary, key findings and further work conclude the chapter in section 5.7.

Primary data generated through experiments is available at [Milczewska and Dacre \(2021\)](https://doi.org/10.17864/1947.325), <http://dx.doi.org/10.17864/1947.325>.

### 5.3 Model configurations and key parametrisation schemes

The following section introduces three NWP models whose meteorological data provides input for the experiments using the NAME atmospheric dispersion model. The three models are the UK variable resolution model (UKV), the global configuration of the Unified Model (UM Global) and AQUM. The study period for this chapter is June 2017, where differences between UKV, UM Global and AQUM are evaluated in terms of boundary layer depth and the surface sensible heat flux,  $Q_H$ . This leads to a discussion of the boundary layer parametrisation schemes used in the models, and the role of land-surface representation in determining boundary layer stability. Understanding the differences in these parameters is relevant to understanding the physical processes which contribute to surface concentration forecasts of  $O_3$  and  $NO_2$ .

AQUM, UKV and UM Global are all configurations of the Met Office Unified Model (MetUM), while NAME is a separate atmospheric dispersion model. The MetUM is a "seamless" modelling strategy such that the fundamental dynamics and physical processes are closely related in the different configurations. Although the three configurations stem from the MetUM and therefore share numerous characteristics, some differences exist and are important to note - the following section will highlight those relevant to this study and introduce the configurations' shared parametrisation schemes. Please refer to section 3.4 for a full description of AQUM.

#### 5.3.1 NAME

The Numerical Atmospheric-dispersion Modelling Environment (NAME III vn. 7.1; [Jones et al., 2007](#)) is a Lagrangian model which simulates dispersion of particles, tracking their movement either in forward or inverse mode. It can have a large number of variable concentration particle sources, released within any domain on a user-specified co-ordinate system. It is an off-line chemistry-transport model (CTM) which requires three-dimensional gridded meteorological data input. This can be from deterministic or ensemble configurations of the MetUM (e.g. UM Global and UKV), or global forecasts and / or reanalysis products from ECMWF. This data is required by the flow module in

NAME, which supplies parameters (e.g. mean flow) to the parametrisation scheme. Representation of processes such as advection, turbulence and both dry and wet deposition will influence the transport and accumulation of particles in a NAME simulation. Particles can either be passive tracers, or can be modelled as chemical species by specifying the use of one of NAME's chemistry scheme configurations (e.g. a comprehensive gas and aqueous tropospheric chemistry scheme). Tracers are specified with properties such as source emissions, lifetime and deposition characteristics.

Because the horizontal and temporal grid is not prescribed, the user can tailor the output of particle concentrations and meteorological fields over a uniform grid or a non-uniform grid e.g. for time-series at point locations or defined regions. In line with the point-based verification of chapter 3, the focus is on the time-series of boundary-layer averaged tracer concentrations at AURN site locations, rather than entire regions.

### 5.3.1.1 Homogeneity of the NAME boundary layer scheme

NAME can read boundary layer depth from the NWP input and apply it during the simulations. Resolved mean wind velocities  $\bar{u}$  from NWP input advect particles in three dimensions, while deviations from the mean wind are simulated with a random walk scheme. The un-resolved scale transport is further sub-divided into mesoscale motions and turbulence, which are treated independently. The equation of motion for a particle from position  $x_t$  in time  $\Delta t$  is:

$$x_{t+\Delta t} = x_t (\bar{u}(x_t) + u'(x_t) + u'_m(x_t)) \Delta t \quad (5.1)$$

where  $\bar{u}$  is the resolved advection,  $u'$  is motion due to treatment of turbulence scales and  $u'_m$  is that of the unresolved mesoscale motions.

Turbulent velocity components required for modelling boundary layer turbulence include velocity variance profiles ( $\sigma_{u,v,w}^2$ ) and Lagrangian timescales ( $\tau_{u,v,w}$ ). Velocity variances are used in the calculation of Lagrangian timescales, which have different formulations depending on the boundary layer stability - for details, see [Jones \(2017\)](#).

Two different versions of the NAME boundary layer scheme can be used:

**Inhomogeneous:** vertical profiles of the velocity variances vary with height; or

**Homogeneous:** vertical profiles are treated as a boundary-layer average.

The homogeneous scheme is typically used with a long-range Wiener ('diffusive') random-walk model and is less computationally-intensive than the inhomogeneous scheme,

thus it can be used for long-range dispersion. Short-range dispersion is better suited to a scheme with particle velocity memory, which models particle dispersion from a point source linearly along the travel direction axis (for short travel times) and parabolically (for longer travel times), which is closer to observations (Taylor, 1922). However, because the inhomogeneous scheme has a memory of particle velocity from the previous time-step, it is naturally much more computationally intensive to run simulations. Therefore, where the main concern is to analyse long-range particle transport, the homogeneous scheme provides adequate accuracy while minimising computational power.

The downside of using the homogeneous scheme is that it results in a discontinuity in turbulent fluxes across the boundary layer top, due to the use of a fixed value within the free troposphere. An entrainment parametrisation scheme (Webster and Thomson, 2011) permits particle movement across the discontinuity, using vertical velocities at either side of the interface to determine whether to transmit or to reflect each particle. For the scheme to work, particles are assumed to be non-sedimenting. The vertical velocity of a transmitted particle ( $w_t$ ) depends on its incident velocity ( $w_i$ ) and relationship of the effective velocity variance of the particle at the incident side ( $\sigma_w(i)$ ) and the other side of the interface ( $\sigma_w(t)$ ). The velocity of the transmitted particle follows the relation:

$$w_t^2 = \sigma_w^2(i) \left[ \frac{w_i^2}{\sigma_w^2(i)} + \log \left( \frac{\sigma_w^2(t)}{\sigma_w^2(i)} \right) \right] \quad (5.2)$$

$$= \sigma_w^2(i) \beta \quad (5.3)$$

In other words, the exit velocity of the particle incident on an interface is determined by the incident velocity. In the case of  $\beta > 0$ , the particle is transmitted with velocity  $w_t$ , otherwise it is reflected with velocity  $w_r = -w_i$ .

Because resolved velocity is used in the NAME entrainment scheme, this parameter is worth investigating further within the model configurations used in this chapter. The next section will therefore focus on the evaluation of vertical velocities  $w$  at boundary layer top, boundary layer depth and  $Q_H$  from the UKV and UM Global meteorology fields used for the NAME simulations outlined in section 5.5.

### 5.3.2 UKV

The necessity of producing an accurate weather forecast at a local scale has led to development of the variable resolution model over the UK domain. The result is the UKV, which has 70 vertical model levels up to 40km. At its highest resolution,

UKV has a uniform 1.5 km horizontal grid ( $0.0135^\circ \times 0.0135^\circ$ ) and smoothly transitions to 4 km towards the outer edges of the domain. A horizontal grid resolution which is fine enough permits convection processes to be modelled explicitly in place of using a parametrisation scheme. UKV belongs to the family of such near-convection-permitting models, which result in e.g. improved forecasts of precipitation over land terrain (Clark et al., 2016). Because UKV uses LBCs provided by the UM Global, the coarser resolution at the edges allows for better compliance between the two grid-lengths. As with AQUM, the boundary layer scheme in UKV is also that of Lock et al. (2000), but cloud and convective precipitation are modelled explicitly. This may lead to differences in the two models' forecast of solar radiation reaching the surface. Another difference between UKV and AQUM is their representation of land use. Differences between urban and rural regions are better represented within the land surface parametrisation in UKV than in the simpler version within AQUM. This is because parameters such as buildings and street canyons are considered in UKV's land scheme, but not in AQUM (Porson et al., 2010). The surface energy balance is thus also better represented in UKV than AQUM and will be discussed in more detail in section 5.3.5.

### 5.3.3 Global configuration of the Unified Model

UM Global provides a deterministic, medium-range weather forecast as well as boundary conditions for higher-resolution configurations (such as UKV and AQUM). Although the operational grid resolution of the UM Global now is 10 km at mid-latitudes, model resolution was at 17 km during the study period of this chapter (i.e. June 2017). It has 70 vertical model levels up to 80 km. UM Global also uses the Lock et al. (2000) boundary layer parametrisation scheme, whose details are outlined in the following section.

### 5.3.4 The Lock boundary layer parametrisation scheme

AQUM, UKV and UMG all share the Lock et al. (2000) parametrisation scheme to represent most sub-grid scale processes and the dynamical boundary layer processes outlined in section 2.2. The scheme diagnoses the mixing layer height (i.e. depth of the boundary layer) by two separate methods based on the atmospheric boundary layer stability profile:

- **Stable layers:** local Richardson number ( $R_i$ ) based scheme, first-order mixing length closure scheme;

- **Unstable and neutral layers:** dry parcel adiabatic ascent based scheme, non-locally specified.

Boundary layer depth is taken as the maximum height from the  $R_i$  number or parcel ascent methods. This generally results in using a method involving the Richardson number  $R_i$  for stable conditions, and the parcel ascent method for unstable or neutral conditions (see section 2.2 for a reminder on  $R_i$ ).

The critical Richardson number ( $R_{ic}$ ) is used to determine the height at which buoyant suppression and shear generation of turbulence result in a laminar flow, which is taken as the boundary layer depth. In atmospheric models, the value of  $R_{ic}$  often ranges from 0.25 to 1, but can be set to as high as 1.3 (e.g. in NAME). This method is used in stable atmospheres (initially identified at the surface when  $Q_H < 0$ ). This is a first-order K-closure, meaning that the turbulent fluxes are defined in terms of local gradients (of momentum, potential temperature etc) and eddy diffusivity profiles for momentum ( $K_m$ ) and scalars ( $K_h$ ). These are defined in terms of  $R_i$  – see also section 2.2.1.2. The lowest vertical half-level of the model at which  $R_i > R_{ic}$  is used to define the boundary layer top. In the parcel ascent method used for unstable boundary layers ( $Q_H > 0$ ), the height at which the adiabatic lapse rate intersects the environmental profile is determined. Simultaneously, a descent from the cloud top (diagnosed by the Smith (1990) cloud scheme) takes place because sources of turbulence may originate at both ground level and cloud top. In the unstable boundary layer, turbulent fluxes and the K profiles are calculated using a non-local scheme, as air parcels are influenced by characteristics of large-scale convection. The non-local scheme also incorporates an explicit entrainment parametrisation for momentum fluxes, based on the depth of the inversion. If the unstable layer is well-mixed, then the height at which the ascending moist parcel reaches its level of natural buoyancy,  $z_{par}$ , is taken to be the boundary layer height.

In the case of an unstable but cumulus-capped atmospheric boundary layer, its height is set to the cumulus cloud base at the lifting condensation level (LCL), and all processes above this height are parametrised either by the mass-flux convection scheme of Gregory and Rowntree (1990) in AQUM, or modelled explicitly in the case of UKV due to it being a convective-scale model. Further details can be found in Lock et al. (2017).

For non-cumulus-capped boundary layers, the depth is calculated within the boundary layer scheme with the parcel ascent method. Correct diagnosis of boundary layer stability is important for cloud-top turbulence and entrainment of air from the free-troposphere. For example, entrainment is closely related to the rate of change of the

boundary layer depth, as it is one of two processes which causes its growth (the other process being convergence of air below the boundary layer top, causing upward motion), i.e.

$$\frac{\partial z_i}{\partial t} = w_e + w_s \quad (5.4)$$

where  $z_i$  represents boundary layer top,  $w_e$  is the entrainment velocity and  $w_s$  is the large-scale motion (i.e. convergence / subsidence) (Stull, 1988).

The model inter-comparison study of Brunner et al. (2015) finds that AQUM calculates a significantly higher summertime planetary boundary layer depth between 00 - 06 UTC over the European continent than most other regional models. The UK4 (AQUM) model was found to exhibit a night-time mean bias of approx. 350 m when evaluated against 17 radiosonde locations in continental Europe. A difficulty in diagnosing the stable boundary layer depth may originate in the representation of surface heat fluxes and therefore the stability of the surface layer - especially in urban regions - as will be described in section 5.3.5. Alternatively, any errors in the local  $R_i$ -based approach could propagate to the boundary layer depth diagnosis, as Mauritsen and Svensson (2007) point out that even in strongly stable regimes where  $R_i > 1$ , observed turbulent fluxes are non-zero which should make the modelled boundary layer even deeper. NWP models therefore use different fine-tuning methods to correct for too much or too little stability, such as the stability functions based on  $R_i$  (Lock et al., 2017).

The diagnosed boundary layer depth, alongside other variables such as surface temperature, surface heat flux and stomatal conductance are derived with the boundary layer scheme and subsequently used in the dry deposition processes within UKCA sub-routine (see section 2.3.2). It follows then that the boundary layer scheme is important not only for the correct representation of turbulence, energy dissipation, advection of moisture, heat and pollutants among other quantities; but it also plays a role in the physical and chemical transformations of atmospheric species within the coupled UKCA configuration.

Many inter-related processes take place within the boundary layer parametrisation, some of which are related to the representation of surface stability via buoyancy fluxes and  $Q_H$ . Their sign and magnitude is variable depending on the kind of land-use (i.e. whether the model grid-box is representing a rural or urban area) and terrain. Given the complexity of different land structures - streets, buildings, rivers etc - a kilometer-scale model (or coarser) needs to use a land-use parametrisation of some kind. The following section introduces the current land-use schemes used in the UKV, UM Global and AQUM and gives an overview of how well they represent urban regions in particular.

### 5.3.5 Urban land tile schemes: Best and MORUSES

Complex terrain types, such as those in urban areas, have a simplified representation in models through the use of surface land schemes. The approach used for this simplification in AQUM and the UM Global is the [Best \(2005\)](#) scheme, based on the concept of multiple tiles of varying terrain type within a grid-box outlined by the Met Office Surface Exchange Scheme (MOSES2.2; [Essery et al., 2003](#)). In [Best \(2005\)](#) (henceforth referred to as the ‘Best scheme’), each model grid-box is a linear combination of up to 9 land types, of which the urban tile is one. Based on the type of land cover, each tile calculates its own surface energy budget:

$$C \frac{dT_*}{dt} = R_N - Q_H - Q_E - G \quad (5.5)$$

where  $C$  [ $\text{JK}^{-1}\text{m}^{-2}$ ] is the heat capacity,  $T_*$  is the canopy temperature,  $R_N$  [ $\text{Wm}^{-2}$ ] is the net radiation,  $Q_H$  [ $\text{Wm}^{-2}$ ] is the sensible heat flux,  $Q_E$  [ $\text{Wm}^{-2}$ ] is the latent heat flux and  $G$  [ $\text{Wm}^{-2}$ ] is the ground heat flux (see also section 2.2.1.4). There is no anthropogenic heat source in equation 5.5, which is likely to have a large contribution during the winter ([Best, 2005](#)). Within each tile, also calculated are the canopy moisture content; snow masses and snow melt rates. The Best scheme uses bulk parameter values for e.g. emissivity and albedo, and does not depend on building geometry.

Due to the complex nature and sub-grid variability of urban areas, a single tile repre-

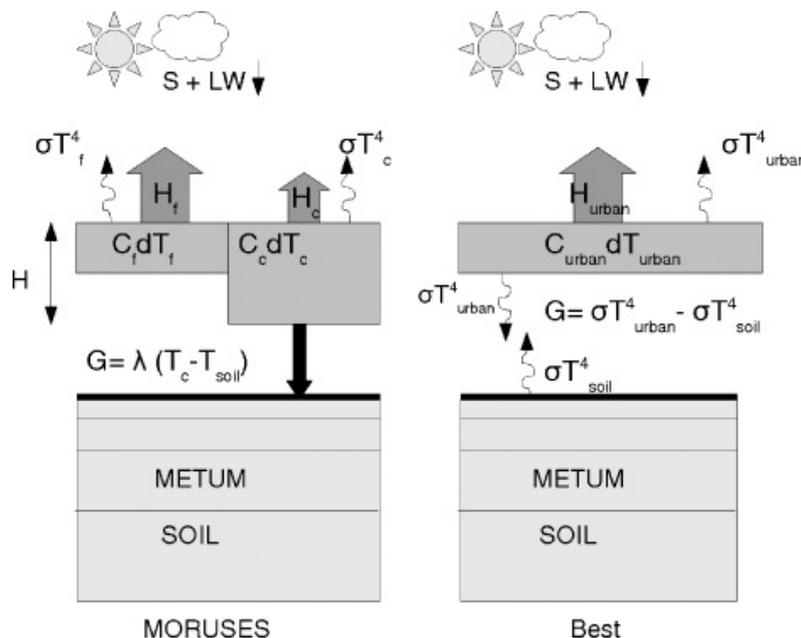


Figure 5.1: From [Porson et al. \(2010\)](#). Schematics of MORUSES and the Best scheme.

sentation is inadequate for accurately capturing the surface energy budget, or representing the urban heat island effect (Bohnenstengel and Hendry (2016), internal report). In particular, building roofs and street canyons are made of different materials and therefore store different amounts of heat due to differences in the surface energy fluxes (Harman and Belcher, 2006). The Met Office Reading Urban Surface Exchange Scheme (MORUSES, Porson et al., 2010) therefore replaces the single-urban tile (urban- 1t) with a 2-tile approach, which distinguishes the energy fluxes between roofs and street canyons. Figure 5.1 is a schematic from Porson et al. (2010) which demonstrates the differences between MORUSES and Best scheme in terms of the coupling between the ground heat flux  $G$  and the surface, and the effect of using two tiles instead of one to represent the urban canopy. MORUSES takes into account urban geometry, which means that parameters such as emissivity, albedo and heat capacity have variable bulk values rather than one, prescribed value. Therefore MORUSES better represents the variety of terrain and building types in urban areas on sub-grid scales, which is beneficial to the modelling of surface heat fluxes and thermal eddies. This also has implications for modelling dispersion of pollutants and air quality in cities (Porson et al., 2010), as will become evident in the experiments performed in this chapter. Some of the main differences between the MORUSES and Best scheme are outlined in table 5.1. MORUSES was implemented in UKV during the Parallel Suite 37 (PS37) upgrade of the MetUM in March 2016.

PS37 sensitivity experiments described in Bohnenstengel and Hendry (2016) suggest that differences in modelling surface wind speeds and fluxes arising from the implementation of a more complex urban tile is most pronounced during calm, sunny days and cloud-free nights, as any insulating effects by cloud-reflection of long-wave radiation back into the boundary layer are minimised. This means that errors in the modelling of low-level urban temperatures and thermal instabilities can be, in a large proportion, explained by the modelling of surface fluxes and heat storage of the variable surface materials. On

**Table 5.1:** Major differences in the land representation scheme used in UM Global and AQUM (Best, 2005), and in UKV (MORUSES, 2011)

Best (2005)	MORUSES (2011)
Single tile to represent urban areas	Two tiles: street canyon and roof
Bulk values for emissivity, albedo, heat capacity, conductivity, roughness length for momentum and heat	Variable parameters, as it takes urban geometry into account. Bulk values for albedo and emissivity for radiative exchange.
Error in timing and amplitude of $Q_H$	Realistic phasing and amplitude of $Q_H$
Total ground heat flux calculated by aggregation of all tiles	Parametrised storage of heat and heat transfer to soil

clear and calm days, differences in wind-direction are considerable, but the sensitivity tests do not reach a conclusion on any differences in boundary layer depth or structure, recognising that more research is needed in this area. This finding, combined with the role of surface heat fluxes in diagnosing boundary layer stability as explained in section 2.2.1.4, may relate to the difficulty experienced by AQUM (among various other regional meteorological models) to represent the stability of the nocturnal boundary layer depth, particularly in urban regions (e.g. [Savage et al., 2013](#); [Brunner et al., 2015](#)).

In summary: representation of the surface energy budget and heat fluxes, through accurately modelling heat storage and capacity in urban areas, is important for capturing an accurate boundary layer development. For example, if the sign of  $Q_H$  is diagnosed incorrectly, it may impact the amount of turbulent mixing throughout the boundary layer, and also near the inversion. Whether there is too little or too much buoyancy-driven turbulence is important for entrainment at the boundary layer top, which, as discussed earlier, may act as a top-down reservoir of  $O_3$ . If this effect is not accurately represented, it will have implications for the surface concentrations of  $O_3$  and  $NO_2$ . Sub-grid heterogeneity related to heat storage within the urban canopy is better represented when the bulk parameters are allowed to vary with building geometry, as is the case in the MORUSES scheme implemented within UKV after 2016. It is therefore useful to compare output of  $Q_H$  from UKV against that from UM Global, since UM Global was still using the Best scheme for the study period of June 2017.

In the following section,  $Q_H$ , boundary layer depth and  $w_s$  from UKV and UM Global are directly compared.

## 5.4 Evaluation of physical parameters

AQUM and UKV not only vary in horizontal grid resolution, but also in their representation of urban land-use. By studying modelled values of  $Q_H$ , one can gain some insight into differences between the two models' treatment of boundary layer stability in urban areas as a consequence of the differing urban land-tile approach. Differences in boundary layer stability may subsequently impact tracer mixing and concentrations, which will be explored later.

This section focuses on modelled night-time and morning boundary layer depth and  $Q_H$  from the employment of two different surface land schemes. Also discussed is the vertical velocity  $w_s$  at model level nearest to the modelled boundary layer height. The

UKV and UM Global meteorology input to NAME have hourly and 3-hourly temporal resolution respectively.  $Q_H$  and boundary layer depth were output from NAME at 53 locations corresponding to urban-type AURN sites, over a  $0.2 \times 0.2^\circ$  latitude-longitude (approximately  $22 \text{ km}^2$ ) area around the point locations. This area size is in line with the AURN definition of an ‘urban’ location as representing an area of “a few  $\text{km}^2$ ”. For AQUM, archived gridded model output for  $Q_H$  and boundary layer depth was interpolated using bi-linear interpolation of gridded model fields to the same 53 latitude-longitude point coordinates as above.

#### 5.4.1 Meteorological conditions

The study period for this evaluation is 01-06 to 30-06-2017. The period was marked by varied conditions across the UK, as summarised in [Met Office \(2020a\)](#)<sup>1</sup>. The first part of the month (1 - 10th) was unsettled and characterised by strong winds alongside heavy and persistent precipitation, caused by a series of cyclone passages (figure 5.2a). Between 10th - 16th June, the weather conditions became more settled across England and Wales, though with some intense showers at the beginning of the period. Between 14th - 16th June, the south of the UK was more settled and sunny than the north, which experienced cloudier conditions and also some rain (figure 5.2b).

In general across the UK, higher pressure conditions prevailed throughout the period of 16 - 22nd June, with lighter winds and warm weather, with thunderstorms towards the end of this period (figure 5.2c). For the rest of the month, cloudier and cooler conditions set in across the UK, with rain and wind in the north (figure 5.2d). These variable conditions render June 2017 a suitable study period for relating meteorological conditions with the ground-level air quality forecast. Within the following sections, references to ‘low’ or ‘high’ pressure days relate to a classification by visual inspection of surface pressure charts, such as those shown in figure 5.2.

---

<sup>1</sup><https://www.metoffice.gov.uk/research/climate/maps-and-data/summaries/index>, last accessed 2021-02-16

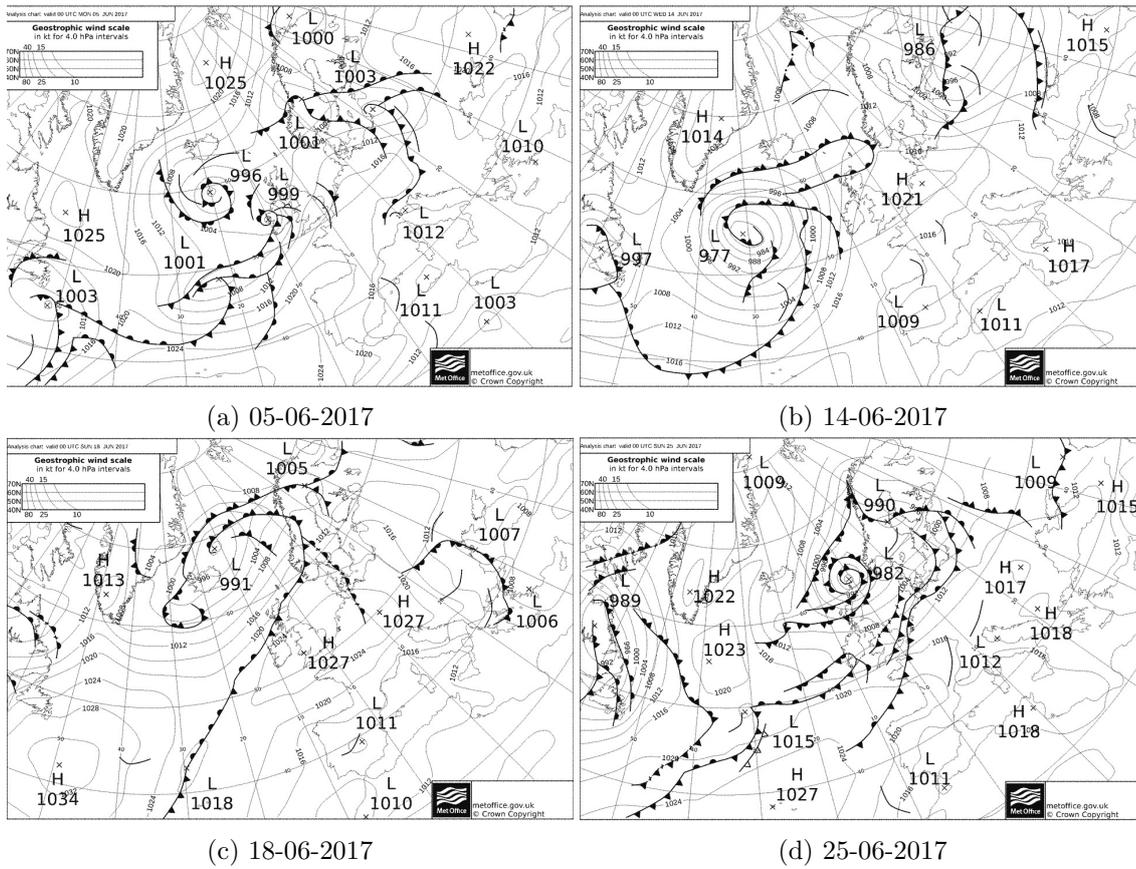


Figure 5.2: Surface reanalysis charts at 00:00 UTC from the Met Office, sourced from <https://www.wetterzentrale.de/>

## 5.4.2 Boundary layer depth

Since the boundary layer depth defines the depth through which pollutants experience vertical and horizontal mixing, modelling its extent correctly is important for achieving an accurate regional forecast of surface pollutant concentrations. This section analyses differences in modelled boundary layer depth within urban and rural regions.

### 5.4.2.1 Urban regions

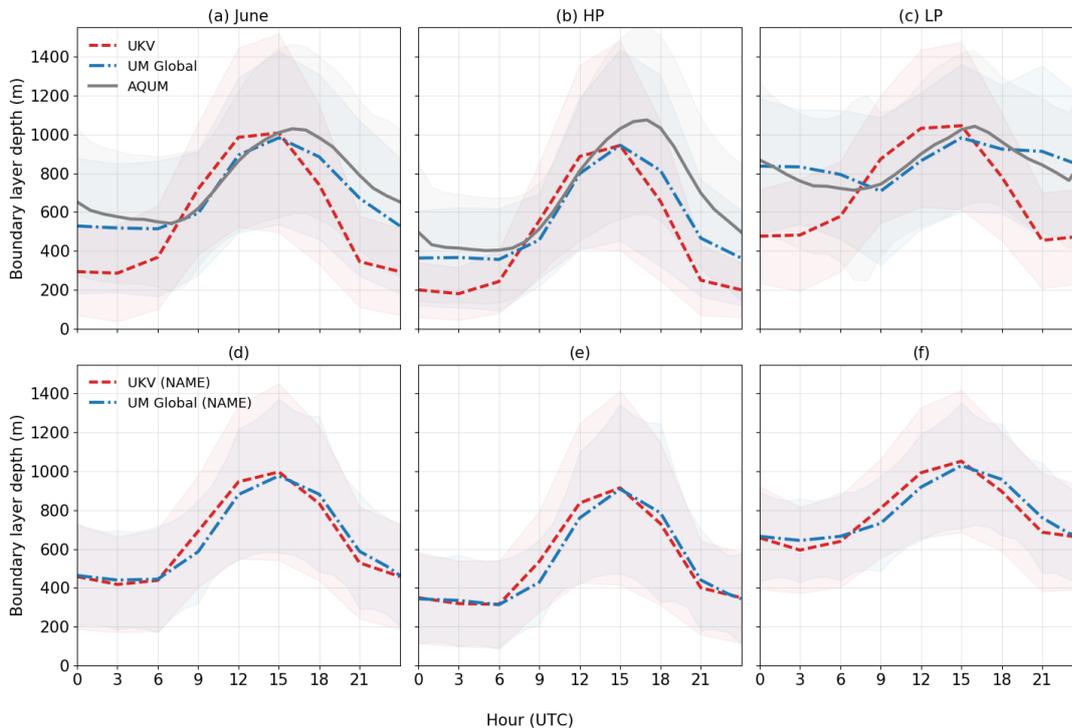
Results of the evaluation of boundary layer depth in urban regions from UKV, UM Global (and, for completion from AQUM) are shown in subplots of figure 5.3. The top panel (plots a, b and c) represents average boundary layer depth in UKV, UM Global and AQUM over (a) the entire study period, (b) during high-pressure days and (c) low pressure days. The bottom panel (plots d, e and f) shows average boundary layer depths but simulated with NAME's own boundary layer parametrisation scheme. Since the purpose of the plots in the bottom panel is to depict how the change in parametrisation scheme influences the modelled boundary layer depth (and hence the change in boundary layer-average tracer concentrations in the experiments), in addition to not using AQUM meteorology within the experiments due to incompatibility with NAME, AQUM is omitted from the bottom panel of figure 5.3. Boundary layer depths from AQUM are only shown in the top panel for comparison against the UKV and UM Global configurations of the Met UM.

UKV data has been coarse-grained from hourly to 3-hourly resolution in order to eliminate the question of whether differences in the boundary layer evolution between the UKV and UM Global exist only because UKV has a finer temporal resolution than UM Global. Note that AQUM remains in its original hourly resolution, as it is only shown to confirm that UM Global data can be used as a proxy for AQUM meteorology. It can be seen in figure 5.3 that this is not the case, and there are clear differences between the two profiles even when the resolutions match. The sub-plots can be analysed in terms of 4 distinct stages of the boundary layer evolution:

1. nocturnal boundary layer (NBL);
2. morning transition;
3. daytime mixed layer;
4. evening collapse.

From the shape of the diurnal cycle, the NBL can be defined as the period from 21 - 06 UTC in the UKV, but 00 - 06 UTC in the UM Global and AQUM because the modelled boundary layer depth is still decreasing, while the UKV rate of change is much smaller between 21 - 00 UTC. NBL depth in UM Global and AQUM is largely comparable (550 - 600 m agl), whereas the UKV boundary layer depth is significantly shallower (300 - 400 m agl). The morning transition period in UKV happens between 06 - 12 UTC, when mean boundary layer depth increases steadily from 600 m to 1000 m. In AQUM and UM Global, the mean boundary layer deepens from 550 m to 900 m until 12 UTC, but continues on to reach a peak of 1000 - 1050 m between 15 - 16 UTC which can be called the daytime mixed layer. Note that from 15 UTC, the UKV boundary layer already begins to deepen steadily, while the UM Global has a shallower decrease until 18 UTC, when the collapse intensifies.

It is evident from figure 5.3(a) that the overall behaviour of boundary layer development in UM Global and AQUM is comparable. This is expected because both models use the same  $Ri_c$  criteria for diagnosis of boundary layer depth. When the UKV data is not



**Figure 5.3:** Diurnal cycle of mean modelled boundary layer depths averaged over 53 urban locations in (a,d) June (01-06-2017 to 30-06-2017), (b,e) high-pressure days and (c,f) low-pressure days. Top panel shows the modelled boundary layer depths calculated within each NWP separately; bottom panel shows the UKV and UM Global when switched to the NAME boundary layer scheme, i.e. using a common  $Ri_c$ . Shading represents the standard deviation of boundary layer depth across all locations.

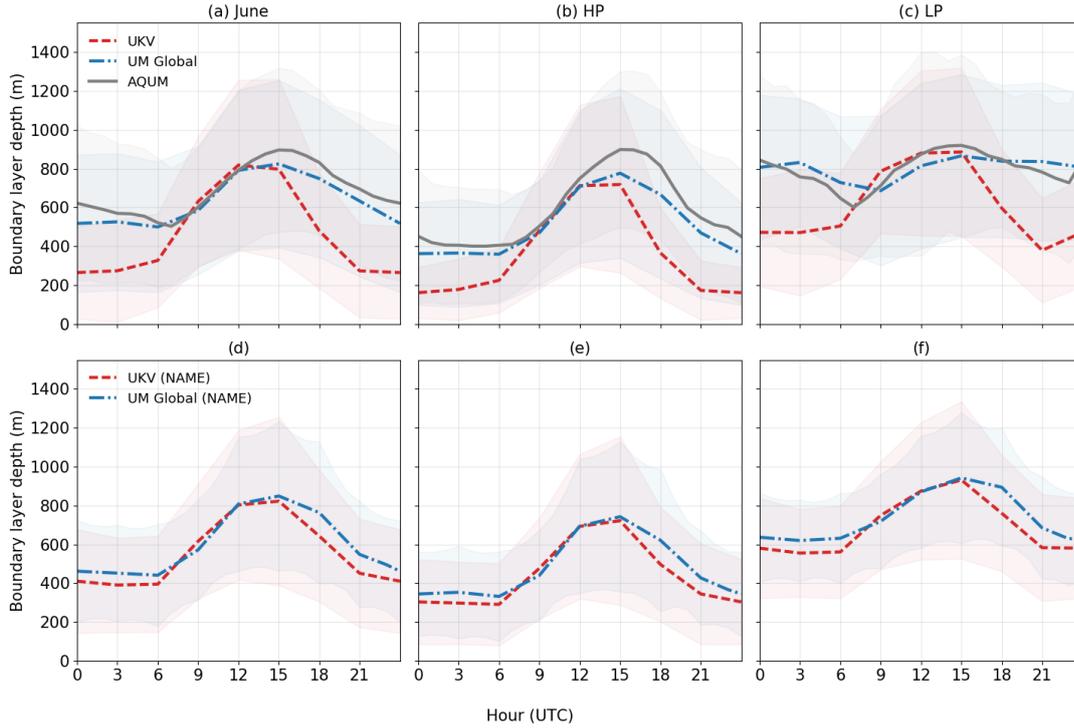
coarse-grained, it shows that UM Global and AQUM lag the UKV maximum boundary layer growth in the morning by at least 1 hour. This lag coincides with the lag in the rise of  $O_3 + NO_2$  concentrations discussed in section 3.5.1.1.

When only high-pressure days are considered (figure 5.3b), the UKV and UM Global NBL profiles are closer in magnitude (200 m and 400 m respectively). Although the UKV profile again shows a steeper increase than the UM Global during the morning transition, the mean boundary layer depth is shallower than when all days are considered, as both configurations reach a maximum average depth of 900 m at around the same time. Meanwhile, the AQUM boundary layer continues to increase until 17 UTC, before it starts collapsing in the evening. For the low-pressure days only (figure 5.3c), mean boundary layer depths across all three configurations are deeper at all times of the day than during high-pressure days. Interestingly, while the UKV profile deepens from 3 UTC, the UM Global profile decreases between 0 - 9 UTC, which is closely matched by the AQUM profile.

Significant differences between the UKV and UM Global representation of the boundary layer depth could be related to their differences in the diagnostic used during stable conditions. One way to determine whether the differences described above are caused by different criteria for critical Richardson number ( $Ri_c$ ) during stable conditions is by enabling NAME to use diagnostics from UKV and UM Global in order to diagnose boundary layer depths with a common value of  $Ri_c$ . Figure 5.3d shows that the NBL, daytime mixed layer and evening collapse are all consistent between the UKV and UM Global when using the same  $Ri_c$ , but the morning growth is not. This behaviour is consistent for low and high pressure synoptic conditions (figure 5.3(e,f)). The only obvious differences between the UKV and UM Global profiles is the timing of the start of boundary layer growth and collapse.

#### 5.4.2.2 Rural regions

When 22 rural locations are considered (figure 5.4), the daytime maximum peak in June reached by UKV and UM Global is 800 m (900 m for AQUM), which is shallower by around 200 m than in urban areas. The most significant observation about rural areas is that the timing offset in the morning transition is not evident. Furthermore, the offset disappears completely upon using NAME's own boundary layer scheme over rural areas, whereas the urban regions still experience an offset during this time. This suggests that something about differences in the representation of urban regions within UM Global and

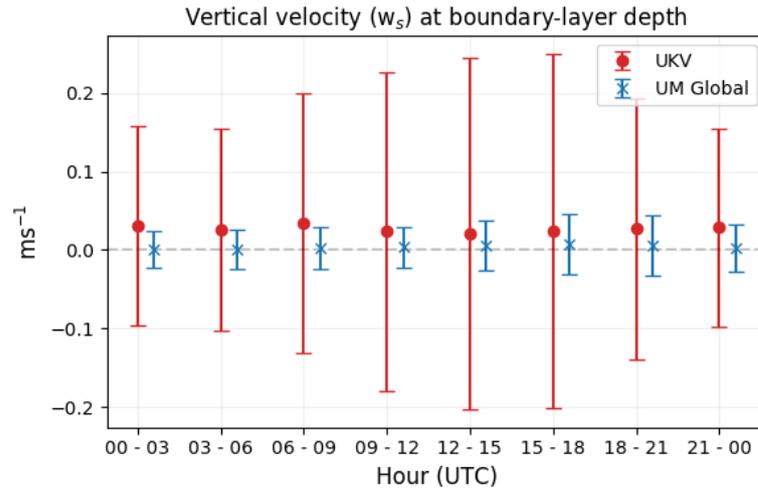


**Figure 5.4:** Diurnal cycle of mean modelled boundary layer depths averaged over 22 rural locations in (a,d) June (01-06-2017 to 30-06-2017), (b,e) high-pressure days and (c,f) low-pressure days. Top panel shows the modelled boundary layer depths with own NWP scheme; bottom panel shows the UKV and UM Global when switched to the NAME scheme, i.e. using a common  $Ri_c$ . Shading represents the standard deviation.

UKV influence the treatment of the boundary layer evolution. Physically, a significant difference between a rural and an urban region is the heterogeneity of land use, e.g. the spatial and vertical distribution of buildings, green areas, farms, etc. In particular, the surface energy budget has varying contributions -  $Q_H$  represents how much heat the ground emits, which will be different depending on whether a grid box is covered by open grassland or buildings. In light of the results of the boundary layer depth analysis discussed in this section, section 5.4.4 explores the representation of  $Q_H$  in urban regions within UKV and UM Global configurations of the MetUM. First, let us consider the vertical velocity  $w_s$  contribution to boundary layer evolution as per equation 5.4.

### 5.4.3 Vertical velocity

Vertical motion across the boundary layer top is associated with two processes: turbulent entrainment of free-tropospheric air into the boundary layer, and mesoscale vertical motion across the inversion due to horizontal divergence or convergence of air, which causes subsidence or upward motion (Stull, 1988). Here, entrainment velocity is denoted as  $w_e$ , while the mesoscale vertical advection velocity is  $w_s$ . In order to parametrise



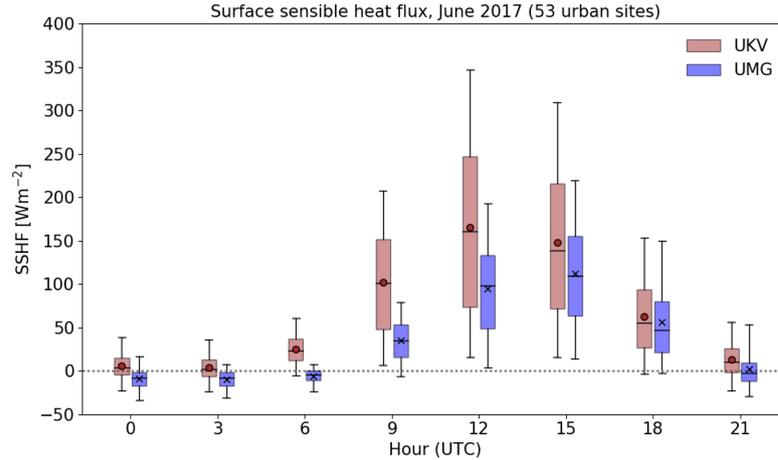
**Figure 5.5:** Vertical velocity  $w_s$  distributions at boundary layer height from 53 locations in UKV and UM Global for June 2017. Filled circles are 3-hour averages and vertical bars are standard deviation,  $\sigma$ .

entrainment of free-tropospheric air into the boundary layer, NAME requires values for vertical velocities  $w_s$  and the variances,  $\sigma_w$  at boundary layer top. Figure 5.5 shows the diurnal distribution of instantaneous  $w_s$  from both the UKV and UM Global, discretised to the nearest 50m of the boundary layer top. Values of  $w_s$  were pooled into 3-hourly windows across the entire month (essentially, each category 00 - 03, 03 - 06 UTC etc altogether had 30 days  $\times$  3 hours  $\times$  53 locations), and the mean and standard deviation  $\sigma$  were calculated.

Figure 5.5 shows that throughout the day, UKV consistently resolves higher average values of  $w_s$  than UM Global, which are one order of magnitude greater ( $10^{-2}$  ms<sup>-1</sup>, while mean values in UM Global are on the order of  $10^{-3}$  ms<sup>-1</sup>) as well as having larger  $\sigma_w$ . Observations of mean  $w_s$  can be difficult to measure accurately, as aircraft measurements often have a mean bias larger than the magnitude of  $w_s$ . However, some published experimental aircraft observations measure  $w_s$  to be on the order of  $\pm 0.2$  cm s<sup>-1</sup> (Dodson and Small Griswold, 2021). The distributions in figure 5.5 are therefore of the correct magnitude, although  $\sigma$  of the UKV values is large. In the NAME entrainment parametrisation scheme,  $w_s$  is an important parameter in that its value determines whether a particle is transmitted or reflected from the boundary layer top interface. The clear difference between the two models will have implications for how much tracer is entrained through the boundary layer top.

#### 5.4.4 Surface sensible heat fluxes

Having discussed the role of  $Q_H$  in determining boundary layer stability in section 2.2.1.4, this section studies the direct influence of  $Q_H$  on boundary layer depth within the present evaluation. Figure 5.6 shows distributions of modelled  $Q_H$  from 53 urban locations in the UK for the study period as a function of hour of day. Average diurnal cycles of  $Q_H$  from UKV and UM Global are shown to have differences in time of daytime maximum, rate of morning increase and sign at night. Between 00 - 06 UTC, median values in UM Global are negative, whereas the UKV medians remain positive. This means that during the night, different criteria are used to diagnose boundary layer depth in the two model configurations at some locations: local  $Ri_c$  while  $Q_H \leq 0$  (UM Global) and adiabatic parcel while  $Q_H \geq 0$  (UKV). There is a significant difference in  $Q_H$  between the two models at 06 UTC, when the boundary layer starts to grow: while  $Q_H$  in UM G remains negative, there is no buoyant production of turbulence, and consequently no thermally-driven boundary layer growth. Meanwhile, UKV fluxes are positive and buoyant generation of turbulence is already contributing to boundary layer growth, as evident in figure 5.3a. Furthermore, since the boundary layer remains neutral in this scenario, the nocturnal inversion will be weaker, thus thermals are more likely to break down the inversion and entrain  $O_3$ -rich air from the residual layer reservoir aloft. Both models'  $Q_H$  values increase from 06 - 12 UTC as solar radiation acts to increase the energy at the surface. But while UKV reaches a daytime 3-hourly maximum at 12 UTC,  $Q_H$  in UM Global continues to rise until 15 UTC. Both the mean and median daytime  $Q_H$  values are higher in UKV than UM Global throughout the day until 18 UTC, when solar radiation weakens. This implies that the land-surface parametrisation in UKV (MORUSES) is not only able to store more urban heat throughout the night, but also that this continues throughout the daytime. The mismatched timing of  $Q_H$  increase and peak seen in figure 5.6 is consistent with the findings of by [Bohnenstengel and Hendry \(2016\)](#). Their report evaluates sensitivity tests on the performance of the MORUSES land tile against the previous JULES urban single-tile scheme, which is also the surface tile used in both AQUM and UM Global ([Best et al., 2011](#)). The differences between the two tile schemes was already introduced in section 5.3.5, but the main difference is that MORUSES employs a variable bulk representation of sub-grid parameters such as heat capacity due to buildings. The MORUSES tile scheme thus models  $Q_H$  values which are generally positive in urban areas (shown in figure 5.6). The positive night-time values of SSHF from these UKV simulations match the findings of [King \(2015\)](#), who



**Figure 5.6:** Distributions of modelled  $Q_H$  between 01-06-2017 - 30-06-2017 at 53 urban locations in the UK. Data are hourly from UKV (red) and from UM Global (“UMG”, blue). Boxes represent the 25th - 75th percentile range, and whiskers are 5th and 95th percentiles. Average values are represented by horizontal line in each box, while the median is a circle (UKV) or a cross (UMG).

demonstrate the similarity of the diurnal cycle in observed SSHF and UKV simulations using the MORUSES scheme, comparing it also to the urban-1t scheme, at one London-based observations site (Kings College London). The report finds that the improvement in representation of SSHF with using MORUSES in place of urban-1t is clearest during the daytime: the mean maximum values reached are better captured with MORUSES than urban-1t when compared against observations, and the timing of the peak occurs 1 - 2 hours later in MORUSES, which is more representative of what happens in urban areas.

## 5.5 Methodology for idealised NAME experiments

Having discussed some of the main differences in the underlying meteorological parameters from UKV and UM Global, this section will describe the NAME experiments which use the above model data as input. In this section, the configuration of the semi-idealised experiment simulating downward entrainment of  $O_3$  from the free troposphere into the boundary layer is presented. The experiment requires archived data from the meteorological models, a definition of particle sources and a tagging mechanism to allow us to distinguish whether particles have spent any time in the free troposphere before being entrained into the boundary layer. Thus, the aim is to quantify the relative contribution to surface concentrations of tracer originating above the boundary layer, in relation to the local surface emissions.

An overview of the tracer sinks and sources in the simulations is provided in table 5.2, with supporting information described in more detail in sections 5.5.1 to 5.5.4.

### 5.5.1 NWP input

Instantaneous and time-averaged UM Global meteorological fields (e.g. for cloud and precipitation) are used at 3 h resolution as input to NAME. Each file input to NAME contains forecast data from successive forecast cycles valid over a 6 h window, i.e. cycles initialised at 00 UTC, 06 UTC, 12 UTC and 18 UTC. Instantaneous fields from a 4-d variational data assimilation analysis are supplied for T+0, as well as the subsequent T+3 forecast value (Jones, 2015).

### 5.5.2 Tracer sources

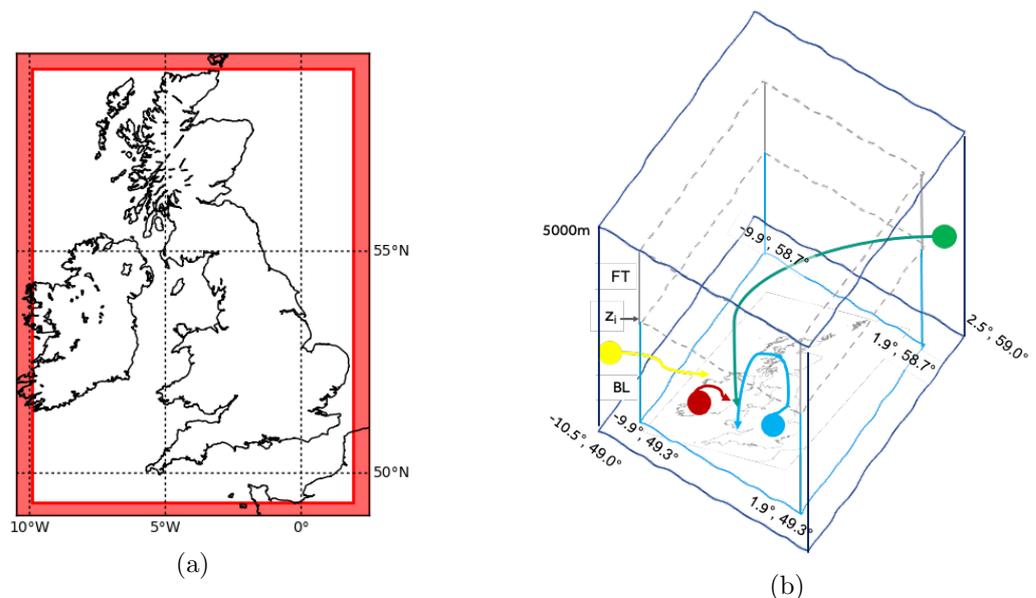
‘Particles’ can be regarded as moving air parcels within a Lagrangian framework to enable a study of dynamics within the boundary layer, rather than as physical molecules. Throughout this chapter they will simply be referred to as ‘tracers’. Tracer properties have been designed to simulate aspects of the behaviour of O<sub>3</sub> and NO<sub>2</sub> without the complexity of introducing chemical transformations. This has been achieved by (i) regarding the tracer to represent the sum of O<sub>3</sub> and NO<sub>2</sub> (O<sub>x</sub>), which is a conserved quantity in the photo-stationary state where rapid interchange between NO<sub>x</sub> and O<sub>3</sub> occurs; (ii) using a simple loss term at the surface, analogous to dry deposition with a defined deposition velocity - values of  $v_d = 1, 5$  or  $10 \text{ cm s}^{-1}$  are used - to represent loss of the tracer within the boundary layer; (iii) setting appropriate lateral boundary conditions which result in higher tracer concentrations above the boundary layer than below.

**Table 5.2:** Overview of tracer sources and sinks used in the simulations.

Sources		Sinks	
Surface	- Release height: 0 - 100m	dry deposition, $v_d$	1, 5 or $10 \text{ cm s}^{-1}$
	- Sinusoidal strength between 09:30 - 16:30, zero otherwise	wet deposition	by rainfall in underlying NWP
	- ‘SBL’ if within the BL	transport out of domain	by advection in underlying NWP
	- ‘SFT’ if above the BL		
Lateral	- Release height: 0 - 5000m		
	- Uniform strength between 0 - 800m		
	- Double release strength above 800m, decreasing with height		
	- ‘BGBL’ if within the BL		
	- ‘BGFT’ if above the BL		

Two types of tracer are released in the simulations: tracers originating at the surface, and tracers originating at the lateral boundaries. Surface tracers are uniformly released throughout 0 m - 100 m above ground level, over all model grid-points where land is present. This was achieved by creating an input file with independent tracer emissions sources spanning the UK landmass, created from the UM Global topography file. There is a diurnal variation in relative strength of release in order to simulate the daytime photochemical production of  $O_3$ , although no weekday variation is applied. The relative release strength varies sinusoidally, beginning at 09:30 UTC, reaching a maximum of 1 at 12:30 - 13:30 UTC, and ending at 16:30 UTC. There is no tracer released outside of these hours. The hours have been chosen to reflect  $NO_x$  emissions from road vehicles during a typical week-day in the UK, to a first-order. A more accurate method of achieving this effect would be to represent the diurnal variation in  $NO_x$  emissions from the NAEI inventory; however, for this idealised experiment a more approximate representation of surface emissions is adequate.

Lateral boundary tracers are released within a rectangular border around the domain. The dispersion domain has outer corners counting clockwise from the SE corner at coordinate locations of  $(-10.5^\circ, 49^\circ)$ ,  $(-10.5^\circ, 59^\circ)$ ,  $(2.5^\circ, 59^\circ)$  and  $(2.5^\circ, 49^\circ)$ . The dispersion domain width measures  $0.3^\circ$  in the southern- and northern-most part of the domain, while the east-west thickness is  $0.6^\circ$ . A schematic is shown in figure 5.7a. This



**Figure 5.7:** (a) Schematic representation of tracer dispersion domain, with vertical release at the red boundary up to an altitude of 5000 m. (b) demonstrates the tagging method with 4 tracer groups: surface-released ('SBL' [red] and 'SFT' [blue]) and tracers from lateral boundaries ('BGBL' [yellow] and 'BGFT' [green]).  $z_i$  represents boundary layer top. Not to scale.

was a convenient lateral boundary dispersion domain to configure in the simulation, as it is within the span of the UKV domain, whose meteorology is needed to advect tracers in from the configured boundaries. Moreover, the domain has a vertical height of up to 5000m above sea level, and tracers are released uniformly throughout the vertical span.

It is important to account for the reduction in air pressure with altitude, as this will impact the tracers released within the lateral boundaries. An idealised constant mass mixing ratio ( $\alpha = \frac{m_{tr}}{m_{air}}$ , where  $m_{tr}$  and  $m_{air}$  are the mass of tracers and air respectively) is used throughout the boundary layer ( $\alpha = 1 \text{ kg kg}^{-1}$ ) and in the free troposphere ( $\alpha = 2 \text{ kg kg}^{-1}$ ) in order to resemble the vertical profile of morning  $O_3$  concentrations as observed at the rural background site in Mace Head (Purvis et al., 2005). The 2x step-increase in  $\alpha$  is applied at a height of 800 m, which was chosen as a constant median boundary layer depth. Alternatively, a diurnally-varying boundary layer depth could be applied to mark where the step-change in  $\alpha$  occurs, but this was not tried in the present experiment.

In order to deduce the ratio of tracer density within the free troposphere ( $\rho_{tr,FT}$ ) to tracer density within the boundary layer ( $\rho_{tr,BL}$ ), we consider the equation of state, which can be expressed in terms of the mass mixing ratio  $\alpha$ , relating tracer and air mass densities as  $\rho_{tr} = \alpha\rho_a$ :

$$p(z) = \rho_a R' T(z) \qquad \rho_{tr} = \frac{\alpha p}{R' T(z)} \qquad (5.6)$$

where  $R' = 287 \text{ J K}^{-1} \text{ kg}^{-1}$  is the specific gas constant for air and  $T(z)$  is the temperature at height  $z$ .

Assuming that  $T(z)$  decreases linearly with altitude under adiabatic conditions,

$$T(z) = T_0 - \Gamma_d z \qquad (5.7)$$

$$= T_0 \left( 1 - \frac{\Gamma_d z}{T_0} \right) \qquad (5.8)$$

where  $\Gamma_d = 9.8 \times 10^{-3} \text{ K m}^{-1}$  is the dry adiabatic lapse rate. This means that the equation of state (equation 5.6) can be expressed in terms of equation 5.8 and the equation for the

approximation of atmospheric pressure  $p(z)$  dependence on height as:

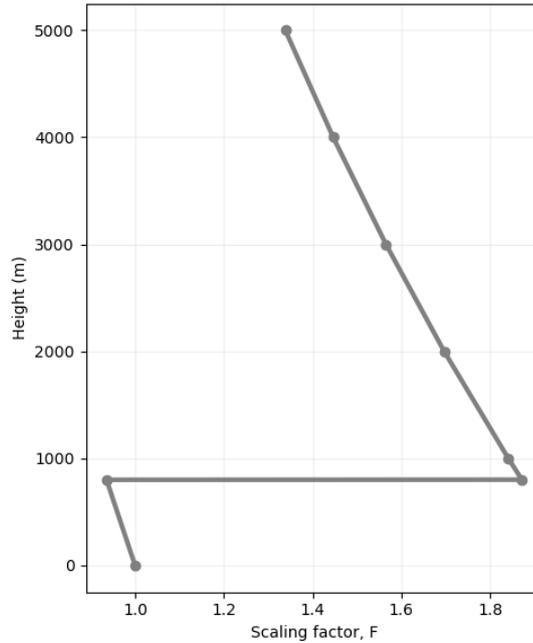
$$\rho_{\text{tr}} = \alpha \left( \frac{\rho_0 e^{-\frac{z}{H}}}{R'(T_0 - \Gamma_d z)} \right) \quad (5.9)$$

$$= \alpha \left( \frac{\rho_0}{R'T_0} \right) \left( \frac{e^{-\frac{z}{H}}}{1 - \frac{\Gamma_d z}{T_0}} \right) \quad (5.10)$$

$$= \alpha \left( \frac{\rho_0}{R'T_0} \right) F(z) \quad \text{where } F(z) = \left( \frac{e^{-\frac{z}{H}}}{1 - \frac{\Gamma_d z}{T_0}} \right) \quad (5.11)$$

where  $T_0 = 288$  K is the surface temperature and  $\rho_0$  is surface pressure. The scaling factor  $F(z)$  (plotted in figure 5.8) is the factor applied to tracer concentrations above the surface relative to surface concentrations in order to account for the decrease in atmospheric pressure while keeping  $\alpha$  constant within the boundary layer and in the free troposphere.

The tracers in the simulations represent the sum of  $\text{O}_3$  and  $\text{NO}_2$ , which involve photochemical production within the boundary layer. Some of the photochemically produced  $\text{O}_3$  in the preceding days will remain in the residual layer, where it undergoes long-range transport and subsequent re-entrainment into the boundary layer the following day. Thus equation 5.11 ensures that the concentrations of tracer advected into the domain from lateral boundaries are greater in the free troposphere than in the boundary layer, and  $\alpha$  is



**Figure 5.8:** Scaling factor  $F(z)$  i.e. second component of equation 5.11 which ensures constant mixing mass ratio  $\alpha$  of tracer to air as atmospheric pressure decreases with height.

kept constant. The vertical gradient sets up a configuration which is representative of the morning  $O_3$  gradient across the boundary layer top, thus air masses entrained from above 800m in the simulations should act to increase the mean boundary layer concentrations.

### 5.5.3 Tagging: tracking movement of tracers

Figure 5.7b is a simplified schematic of the different tracer groups in the NAME simulation, which also shows the domain size for the experiment. At the first time-step after release,  $t_0$ , tracers are prescribed with either a ‘boundary layer’ (BL) tag if their position is below the NWP boundary layer depth, or a ‘free troposphere’ (‘FT’) tag if they are released at a height above it. During the simulation, if at any time-step a ‘BL’ tracer is mixed to a height above the diagnosed boundary layer, it irreversibly switches its tag to ‘FT’. Conversely, any tracer initialised with ‘FT’ tag remains thus throughout all time-steps.

The tracers have a prefix of either ‘S’ if released over land surface, or ‘BG’ (background) if released within the lateral boundaries. Thus four families of tracer may be distinguished at any location or time: ‘SBL’ and ‘BGBL’ (boundary layer-bound tracer released at surface or at the lateral boundaries respectively); ‘SFT’ and ‘BGFT’ (tracers which have spent at least one model time-step above the boundary layer, released at surface or at lateral boundaries respectively). Tracers are prescribed a lifetime of infinity. After about 36 hours of simulation time, the domain-integrated mass of particles reaches a steady-state. Using  $v_d = 5\text{cm s}^{-1}$ , the domain-integrated mass of particles is around  $2 \times 10^8$  g.

It is important to note that the concentrations of tracer after partitioning into the 4 groups is affected by the boundary layer depth, and thus by the use of either NWP or NAME-diagnosed boundary layer depth in the simulation. The experiment is therefore also conducted with NAME’s own boundary layer scheme, results of which will be discussed in section 5.6.

### 5.5.4 Tracer sinks

It has already been mentioned that no chemical sinks exist within the experiment, tracers are not prescribed a decay rate and have infinite lifetime. Therefore the only tracer sink is through dry deposition or exit from the domain by lateral boundaries. The dry deposition velocity  $v_d$  is tuned in a way such that tracer concentrations within the boundary layer form a realistic diurnal cycle. The values of  $v_d$  used in the experiments

are  $1\text{cm s}^{-1}$ ,  $5\text{cm s}^{-1}$  and  $10\text{cm s}^{-1}$ . Section 5.6.1.1 will present results of the sensitivity tests of varying the value of  $v_d$  and its impact on the tracer diurnal cycle.

### 5.5.5 Boundary layer schemes in experiments

Instantaneous hourly boundary-layer-average tracer concentrations are analysed over a  $0.2^\circ$  latitude  $\times$   $0.2^\circ$  longitude (approx.  $20\text{ km}^2$ ,  $< 2$  AQUM grid boxes) area around 53 urban background locations from the AURN pollutant monitoring network. By considering boundary-layer averaged concentrations, the effect of entrainment from the free-troposphere to the boundary layer can be examined. The boundary layer depth directly from NWP is used in the simulations, rather than allowing NAME to re-calculate its own values based on the atmospheric profiles and Ri specifications. It will be shown later that there is little difference in boundary layer-average tracer concentrations between simulations where boundary layer depth is calculated with NAME's own scheme, or used directly from NWP.

Furthermore, the homogeneous boundary layer scheme is used in NAME for the simulations, which means that turbulence parameters do not vary with height and turbulence is fixed within the free troposphere (introduced in section 5.3.1.1). This configuration requires an entrainment parametrisation at the boundary layer top due to a step change in turbulence parameters. The homogeneous boundary layer scheme is less computationally-intensive than the inhomogeneous boundary layer scheme, which in these simulations could only be run for one week due to exceeding computation time.

## 5.6 Results of NAME tracer experiments

Firstly, presented in section 5.6.1 are differences in the diurnal cycle of all types of tracers, driven by UKV and UM Global meteorology. The mean diurnal cycle of tracers is of interest, as it is directly comparable with the diurnal cycle of forecast and observed  $\text{O}_3 + \text{NO}_2$  concentrations presented in section 3.5.1 (which is what the tracers are representing). All tracer families are considered in the initial evaluation in order for the reader to familiarise themselves with how the diurnal profiles of the different tracer families relate to one another.

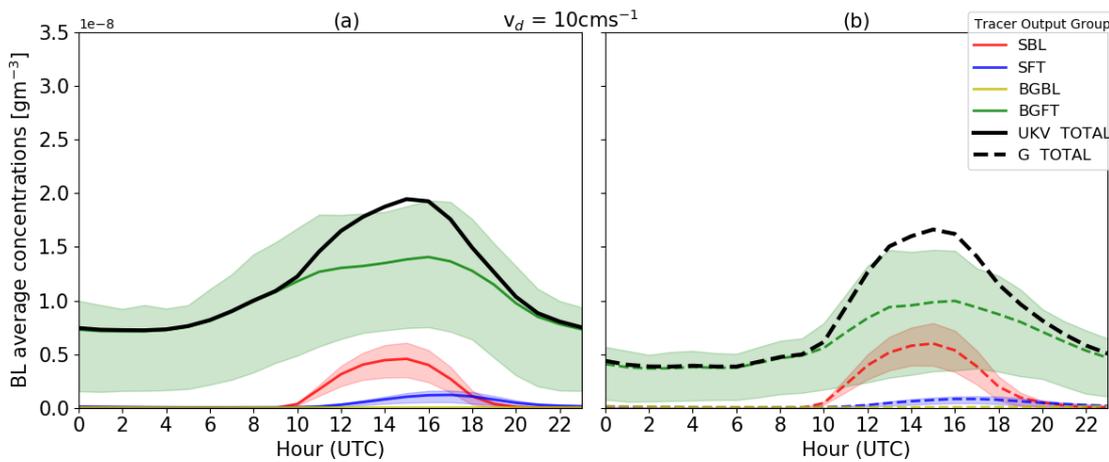
Then, in section 5.6.1.1 the focus is only on the background entrained tracer family, 'BGFT', as the purpose of this experiment is to evaluate the contribution of entrained tracer to boundary layer-averaged concentrations during the morning transition period.

The boundary layer-average concentrations are compared against AQUM forecast and observed  $O_x$  concentrations near the surface, as well as the rate of change of their diurnal cycle. This is presented in terms of  $v_d$  sensitivity, as the results are variable depending on how strong the surface deposition is. This makes sense in light of an Eulerian tracer budget equation, where the boundary layer flux gradient of tracer concentrations is in part dependent on the deposition flux. Therefore, the Lagrangian NAME simulations are compared with an analogous study using an Eulerian budget closure method for calculating the contribution of entrainment to boundary layer-average  $O_3$  concentrations.

Finally, it is explored how the above differences are affected by high and low pressure days in June 2017, which leads to a discussion on the influence of the turbulence parameters and entrainment parametrisation on the tracers within the boundary layer scheme.

### 5.6.1 Diurnal variability

Firstly, let us consider the diurnal evolution in boundary layer-averaged concentrations of all types of tracer in the simulation. Figure 5.9 shows the mean diurnal cycle at 53 locations for all 4 tracer families: surface, BL (SBL, red); surface, entrained from FT (SFT, blue); background, BL (BGBL, yellow) and background, entrained from FT (BGFT, green). Black lines are the total tracer boundary layer-averaged concentrations. Sub-plot (a) is for the simulations using UKV meteorology, and (b) is the UM Global simulation. Dry deposition velocity here is  $10\text{cm}^{-1}$ , the highest value in all experiments.



**Figure 5.9:** Diurnal cycle of boundary layer average concentrations for all 4 tracer types, averaged over 53 locations for simulations using meteorology from UKV (a; solid lines) and UM Global (b; dashed lines). Dry deposition velocity  $v_d = 10\text{cm}^{-1}$ . Note that the BGBL (yellow) contribution is very small and therefore negligible.

All tracer types in simulations with both the UKV and UM Global meteorology exhibit a diurnal cycle which resembles the  $O_x$  diurnal cycles presented in section 3.5.1. Total tracer concentrations shown in black begin to increase smoothly after around 4 UTC in the UKV simulations (figure 5.9a), while in the UM Global simulations the rise in tracer concentrations has a sharp increase point at around 10 UTC (figure 5.9b). Both simulations have a definite afternoon peak (15 - 16 UTC) and collapse again towards the evening and throughout the night.

The surface-released tracers ('SBL' and 'SFT') do not contribute to the total concentrations as much as the entrained tracer ('BGFT'). This is likely to be an artefact of the experiment set-up, where the relative source strength of tracer emitted over land varies sinusoidally between 0 - 1, but could instead be scaled up to match the order of magnitude of the BGFT tracer. As expected, the SBL tracer (red) follows its sinusoidal source strength prescribed in the experiment configuration, which peaks at 15 UTC. The SFT tracer (blue) is more interesting, as it represents the tracers emitted over land which end up being mixed up into the residual layer and are subsequently re-entrained back down into the boundary layer as it grows in the morning. Consequently, there will naturally be less SFT than SBL tracers because SFT tracers must undergo two processes to be included within the blue line of plot 5.9: firstly, they need to reach the altitude of the free troposphere to gain the 'FT' tag; secondly they need to be re-entrained into the boundary layer during the day. The first process requires significant vertical mixing and / or convective updraughts, which are only likely during neutral or unstable conditions. Because the SBL tracers simply remain within the boundary layer, they are more likely to remain within the domain (i.e. not deposited to the surface or advected out of the domain) to be included within the red line in figure 5.9 than the SFT tracers, hence the difference in magnitude between the red and blue lines in figure 5.9.

Differences between the simulations using UKV and UM Global meteorology are most evident in the BGFT tracer (green), i.e. background tracer released at the lateral boundaries and entrained in from above the boundary layer. The amount of BGFT tracer entrained with UKV meteorology is larger than the amount entrained with UM Global meteorology throughout the day, reaching a factor of 1.5x at 16 UTC. This result is the first indication that there are significant differences in the representation of the entrainment process between the NAME-UKV and NAME-UM Global simulations, which will be explored further in sections 5.6.3 and 5.6.4 by comparison with observations and AQUM forecast.

One reason for these differences could be that there are differences in horizontal resolution of the NWP input (UKV is 1.5 km while UM Global is 17 km); there are also differences in the vertical resolution. One possible method of evaluating the differences between the simulations with the two NWP inputs would be to average the UKV output, so that it matches the resolution of UM Global. Resolution of the underlying NWP is an important consideration, as it dictates the scale of resolved processes (e.g. advection and turbulence), and is used to define default values of  $\sigma_u$  and  $\tau_u$  for treating unresolved mesoscale motions. Additionally, we have already seen in figure 5.5 that the diurnal distributions of large-scale vertical motion  $w_s$  at boundary layer top is much larger in the UKV than in UM Global. For example, in the morning (e.g. 06 - 09 UTC),  $w_s = 3 \times 10^{-2} \pm 0.2 \text{ ms}^{-1}$  in UKV, but  $w_s = 2 \times 10^{-3} \pm 0.03 \text{ ms}^{-1}$ . This means that the entrainment parametrisation in NAME runs with a larger magnitude of  $w_s$  for the UKV simulations than UM Global throughout the day, which could be a cause of higher entrained tracer concentrations within the boundary layer in the UKV simulations. The influence of  $w_s$  on entrainment flux will be explored further in section 5.6.2.

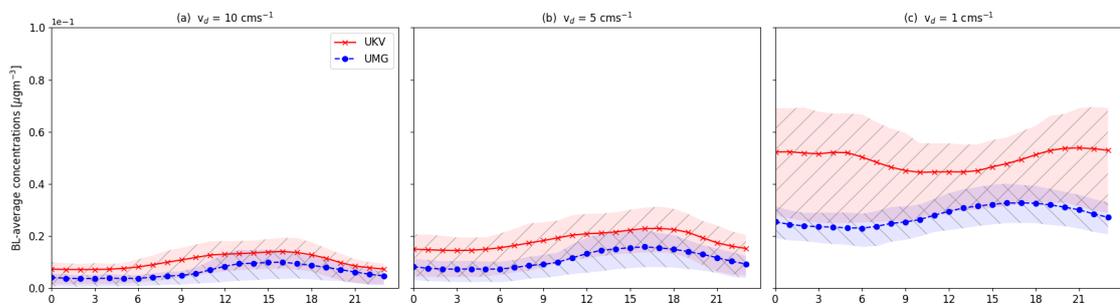
### 5.6.1.1 Dry deposition sensitivity tests

In order to account for all loss mechanisms, including chemical losses which are not explicitly represented in the simulations, the experiment uses high dry deposition values. These values are higher than observed, in order to re-create a tracer diurnal cycle which resembles the observed  $O_x$  concentrations. This will enable a pathway to calculate the contribution of entrained tracer to the surface concentrations, and translate the findings to observed  $O_x$ . It is important to note that the resultant diurnal tracer profile could match the observations and / or the forecast, but for the wrong reasons. I acknowledge this and make the assumption that the tracers represent real-world  $O_x$  as if it were a chemically inert quantity, whose boundary-layer concentrations were only dependent on meteorological conditions, surface deposition and horizontal transport out of the domain.

Figure 5.10 shows the mean diurnal cycle of BGFT tracer concentrations within the boundary layer when the value of  $v_d$  is changed between  $10 \text{ cm s}^{-1}$ ,  $5 \text{ cm s}^{-1}$  and  $1 \text{ cm s}^{-1}$ . Note that a realistic range for vegetated surfaces would be between  $0.2\text{--}2 \text{ cm s}^{-1}$  (Hardacre et al., 2015). The lowest value used was  $v_d = 1 \text{ cm s}^{-1}$ , where the expected diurnal cycle was only re-created for the UM Global simulation, but not in the UKV, as shown in figure 5.10. Clearly, the value of  $v_d$  needed to be increased, thus the following sensitivity tests attempt to answer the question: how large should  $v_d$  be, and should it

be the same for both the UKV and UM Global simulations?

There are some clear differences in the diurnal cycle of tracers under the varying values of  $v_d$  in figure 5.10. In the two simulations with  $v_d \geq 5 \text{ cm s}^{-1}$  (panels a and b), the BGFT tracer has a diurnal cycle profile which has the ‘expected’ shape, i.e. low concentrations overnight and in the morning, then rise between 6 - 12 UTC and reach a maximum by the afternoon before declining again. The balance between boundary layer entrainment as tracer source and dry deposition as tracer sink is different when  $v_d = 1 \text{ cm s}^{-1}$ , i.e. the smallest value in this experiment, and the closest to a realistic one. Here, the average night-time UKV tracer concentrations are approximately 2x the UM Global concentrations (figure 5.10c). There is also a reversed diurnal profile in the simulations driven by UKV meteorology but not for UM Global meteorology. My hypothesis is that enhanced entrainment reverses the vertical gradient of tracer concentrations, resulting in an inverted diurnal cycle of boundary layer tracer concentrations. This would occur if entrainment was acting to dilute the boundary layer concentrations by bringing down air masses with lower tracer concentrations. Throughout the night, when the boundary layer is shallow, existing tracer accumulates and results in high average concentrations. Because those concentrations may now be higher below the inversion than above it, air masses entrained in the morning can dilute the concentrations within the boundary layer. This can happen because when deposition to surface is small, the rate of tracer depletion is smaller than the rate of replenishment through entrainment from the residual layer. The resultant reversal of the diurnal profile suggests that the UKV simulations are unphysical when  $v_d$  is small. The experiment was specifically set up to ensure a positive mixing mass ratio gradient across the boundary layer top, whereas the reversal of the diurnal cycle profile implies a negative gradient.



**Figure 5.10:** Diurnal mean hourly boundary-layer averaged background free-tropospheric (BGFT) tracer concentrations for the period of 01-06-2017 to 30-06-2017, driven by UKV and UM Global meteorology. (a)  $v_d = 10 \text{ cm s}^{-1}$ , (b)  $v_d = 5 \text{ cm s}^{-1}$  and (c)  $v_d = 1 \text{ cm s}^{-1}$ . Shading represents the 25th and 75th percentile range.

### 5.6.1.2 Normalised rate of change in diurnal tracer concentrations

In order to check whether the lag in the  $O_3 + NO_2$  forecast is replicated within the semi-idealised experiment, the rate of change of the BGFT tracers throughout the day is evaluated. This is shown in figure 5.11.

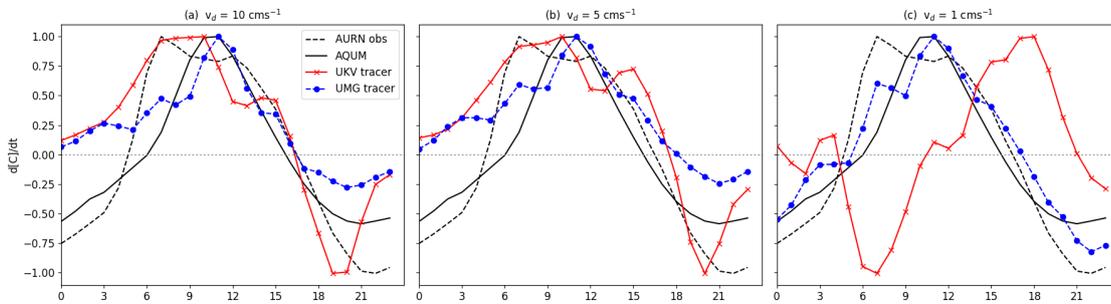
As before, each subplot refers to a value of  $v_d = 1, 5$  or  $10 \text{ cm s}^{-1}$ . The red and blue lines represent the average rate of change of BGFT tracer from NAME simulations using UKV and UM Global meteorology respectively. These have been normalised to their respective diurnal profile maximum  $\left(\frac{d[C]}{dt}\right)_{\max}$  i.e. at hour  $h$ ,

$$\text{normalised } \frac{d[C]}{dt} = \frac{\left(\frac{d[C]}{dt}\right)_h}{\left(\frac{d[C]}{dt}\right)_{\max}} \quad (5.12)$$

The black lines represent the rates of change of observed (AURN; dashed) and forecast (AQUM; solid) surface concentrations of  $O_3 + NO_2$  at urban background sites - these are the same throughout all three sub-plots of figure 5.11 and are used as reference for the simulated tracer.

Relative to the BGFT entrainment rate in UKV, the rate of change of UM Global tracer is largely unaffected by changes to  $v_d$ . In panels a, b of figure 5.11, peak increase in UM Global-driven tracer occurs around 11 - 12 UTC and consistently lags the peak in AQUM total oxidant by 1 hour. After around 8 UTC,  $\frac{d[C]}{dt}$  of tracer simulated with UM Global meteorology is comparable with the AQUM forecast of  $O_3 + NO_2$ , and in panel c the similarity is also consistent during the early morning.

The peak increase in UKV simulations for  $v_d > 5 \text{ cm s}^{-1}$  is comparable with AURN



**Figure 5.11:** Normalised rate of change of background free-tropospheric (BGFT) tracer concentration averaged over 53 locations for UKV (solid, red) and UM Global (dashed, blue), overlaid with rate of change of observed total oxidant (dashed, black) and forecast (black, solid), all normalised by their respective maxima. (a)  $v_d = 10 \text{ cm s}^{-1}$ , (b)  $v_d = 5 \text{ cm s}^{-1}$  and (c)  $v_d = 1 \text{ cm s}^{-1}$ .

observations, i.e. between 6 - 8 UTC in panels 5.11a, b. However throughout the early stages of the morning (0 - 5 UTC), both the UKV and UM Global tracer concentrations are increasing in panels (a,b) and have a smaller magnitude but opposite sign to to observed and forecast observations. This means that tracer concentrations in both the UKV and UM Global simulations are slowly increasing throughout 0 - 5 UTC, whereas the observed concentrations are decreasing until 4 UTC (6 UTC for AQUM forecast). This suggests that the balance between the sinks of tracer (i.e. deposition) and sources (i.e. entrainment, advection) - which may also include tracer accumulated within the boundary layer from the day before - is unequal and not enough tracer is depleted. Note also that total tracer deposition is proportional to average tracer concentrations within the boundary layer. Therefore using  $v_d > 5\text{cm s}^{-1}$  in UM Global simulations means that although there is potential to deplete more tracer to the surface, the boundary-layer average concentrations are less than with a smaller  $v_d$  (e.g. in figure 5.10) and so the vertical gradient in tracer concentrations across the boundary layer top is large. A large tracer gradient means that the entrainment flux into the boundary layer is also large, and potentially larger than the deposition flux (since the amount of tracer in the boundary layer is small), leading to accumulation and hence a positive value of  $\frac{d[C]}{dt}$ , despite the larger deposition value. Figure 5.11 thus suggests that in UM Global, using  $v_d > 5\text{cm s}^{-1}$  is unsuitable.

However, simulations with UKV meteorology show that the diurnal profile evolution is more sensitive to the choice of  $v_d$  than simulations with UM Global. This suggests that the balance between entrainment and deposition is different in the UKV simulations. For example, in UM Global simulations,  $\frac{d[C]}{dt}$  is negative between 18 - 00 UTC for all values of  $v_d$ , whereas for the UKV simulations this is only the case when  $v_d = 10\text{cm s}^{-1}$ . When  $v_d = 1\text{cm s}^{-1}$ , entrainment in the UKV simulation is larger than deposition in the evening (e.g. the tracer increase reaches a maximum at 18 UTC). This supports the hypothesis that the gradient in tracer concentrations reverses sign and results in an un-physical profile, consistent with the postulation that when  $v_d$  is small, morning entrainment acts to dilute boundary layer-average concentrations.

In the UKV simulations, the BGFT normalised rate of change at e.g. 05 UTC is 0.6 of its diurnal maximum, but only around 0.2 in UM Global (figures 5.11a,b). This is further evidence that entrainment of free-tropospheric air into the morning boundary layer occurs earlier in UKV than in UM Global.

### 5.6.1.3 Conclusions of sensitivity tests

Testing the sensitivity of tracer diurnal cycles from the UKV and UM Global simulations to the use of different values of  $v_d$  helps us explore the optimal balance between tracer sources and sinks within this semi-idealised experiment. Based on the timing of the afternoon peak and overall tracer concentration profile, the most ‘realistic’ diurnal cycle of tracers is given by using  $v_d = 1\text{cm s}^{-1}$  with UM Global meteorology, and  $v_d = 5\text{cm s}^{-1}$  with UKV meteorology. The fact that UM Global - but not UKV - maximum rate of increase in tracer concentrations lags the observations in all three  $v_d$  scenarios offers a direct link between the semi-idealised model set-up and the observations, showing that differences in the behaviour of tracers are a consequence of differences in the NWP meteorological fields used in the simulations.

One of the questions that this chapter aims to answer is: how does the diurnal evolution of  $\text{O}_3$  depend on the rate of growth in modelled boundary layer depth? The results presented thus far suggest that the shortcomings of the UM Global / AQUM boundary layer scheme is contributing to the delay in onset of forecast  $\text{O}_x$  increase in the morning. It was decided that in order to best answer the above question, the focus should instead be on  $\text{O}_x$  as a chemically-conserved quantity which can be represented by inert tracers. This eliminates fluctuations in  $\text{O}_3$  concentrations due to chemical reactions with  $\text{NO}_2$  in the photo-stationary state.

In order to re-create the observed and / or forecast  $\text{O}_x$  concentrations within the boundary layer, the tracer budget must be represented correctly. The tracer budget will be explored further in section 5.6.2, where an attempt to quantify the contribution of entrained tracer to the boundary layer-average concentrations will be analysed.

### 5.6.2 Comparison with Eulerian budget study

The study of Kaser et al. (2017) quantifies the contribution of entrained  $\text{O}_3$  (also  $\text{O}_x$ ) to the surface concentrations by calculating the  $\text{O}_3$  concentration budget within the boundary layer. They use simulations from the WRF-Chem model (Grell et al., 2005) and observations taken during a campaign in summer 2014, in the Colorado Front Range. The study concludes that morning entrainment of  $\text{O}_3$  from the residual layer contributes significantly to the morning rate of change of surface concentrations. In this section, the Eulerian budget method from Kaser et al. (2017) (henceforth referred to as ‘Kaser study’) is used to calculate the vertical tracer flux gradient from the Lagrangian NAME

experiments and subsequently the contribution of the entrained tracer to boundary layer-average concentrations.

### 5.6.2.1 Methodology

By conservation of a scalar quantity  $\psi$  - in this case  $\text{O}_3$  or  $\text{O}_x$  concentrations, or tracers in the NAME experiment - the rate of change is given by:

$$\frac{\partial \bar{\psi}}{\partial t} = -U_i \frac{\partial \bar{\psi}}{\partial x} - \frac{\partial \overline{w'\psi'}}{\partial z} + P_{\text{net}}(\psi) \quad (5.13)$$

([Lenschow et al., 1981](#)) where the first term on the right hand side represents mean horizontal advection, second represents the vertical flux gradient and the last term represents net sources and sinks. An assumption is made that the horizontal advection term is small, due to the horizontal homogeneity of  $\text{O}_3$  (discussed in section [4.5.1](#)). Only the ‘BGFT’ entrained tracer is considered (i.e. no boundary layer production); thus, the last term equals zero.

The middle flux term in equation [5.13](#) represents the net contribution of turbulent vertical transport of  $\psi$  to concentrations within the boundary layer and can be expressed as the  $\psi$  entrainment flux at boundary layer top minus deposition flux at the surface:

$$-\frac{\partial \overline{\psi'w'}}{\partial z} = - \left[ \frac{w\Delta\psi - v_d\psi}{z_i} \right] \quad (5.14)$$

$$= \frac{w\Delta\psi}{z_i} + \frac{v_d\psi}{z_i} \quad (5.15)$$

where  $v_d$  is the dry deposition velocity ( $v_d = 5 \text{ cm s}^{-1}$  is used in NAME simulations with UKV meteorology,  $v_d = 1 \text{ cm s}^{-1}$  with UM Global meteorology).  $\Delta\psi$  represents the difference in  $\psi$  concentrations between a defined layer below  $z_i$  and above. Depending on the boundary layer evolution profile, this layer can be as large as 300 m (as in the Kaser study), which estimates the proportion of the residual layer from which air will be entrained down into the boundary layer over a 2-hour period, assuming near-linear growth until 16 UTC. In the present case, a mean summertime morning boundary layer growth in the UKV and UM Global models is closer to 200 m over a 2-hour period ( $90 \pm 30 \text{ mh}^{-1}$ ). Because the boundary layer is considered as a box-model in this budget equation, the entrainment flux term and the dry deposition flux term act as a tracer source and sink respectively, thus work in opposite ways and should be treated with opposite signs.

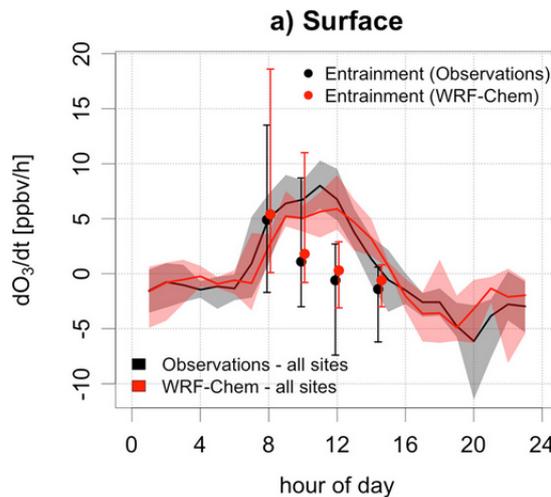
In equation [5.14](#), entrainment velocity  $w_e$  can be estimated by considering the growth

of the boundary layer to be a sum of contributions from (1) turbulent entrainment of air from the residual layer, and (2) large-scale vertical motion caused by convergence / divergence of tropospheric air masses ( $w_s$ ). From equation 5.4, it follows that

$$w_e = \frac{\partial z_i}{\partial t} - w_s \quad (5.16)$$

The Kaser et al. study do not use vertical velocity measurements, but their WRF-Chem simulations give a mean value of  $w_s = -1.3 \pm 1.0 \text{ cm s}^{-1}$ . They therefore acknowledge that their calculated  $w_e$  will be under-estimated by around 25% when represented only with  $\frac{\partial z}{\partial t}$ . While UKV and UM Global values for  $w_s$  are available, the values are instantaneous and are incompatible with the rest of the hourly-averaged quantities such as  $\Delta\psi$  or  $\frac{\partial z_i}{\partial t}$ . Furthermore, it is logical that the hourly-averaged  $w_s$  should be zero, as there is no reason why there should be more updrafts or downdrafts. We also know that the UKV resolves larger instantaneous values of  $w_s$  than UM Global, which is likely due to its significantly larger resolution - smaller grid-boxes are more likely to capture large instantaneous spikes in draft velocity than larger grid-boxes, where the velocities are unresolved and averaged over the entire grid-box. Therefore, the contribution of  $w_s$  to  $w_e$  in equation 5.16 is not included, and it is acknowledged that the entrainment values will be over-estimated due to average hourly  $w_s$  being non-zero between 00 - 15 UTC ( $2.5 \pm 0.5 \text{ cm s}^{-1}$  in UKV,  $0.3 \pm 0.2 \text{ cm s}^{-1}$  in UM Global).

The diurnal cycle of  $\frac{dO_3}{dt}$  is aggregated over 6 sites in the Kaser et al. study from simulations and observations, and is shown in figure 5.12. The plot also shows the diurnal



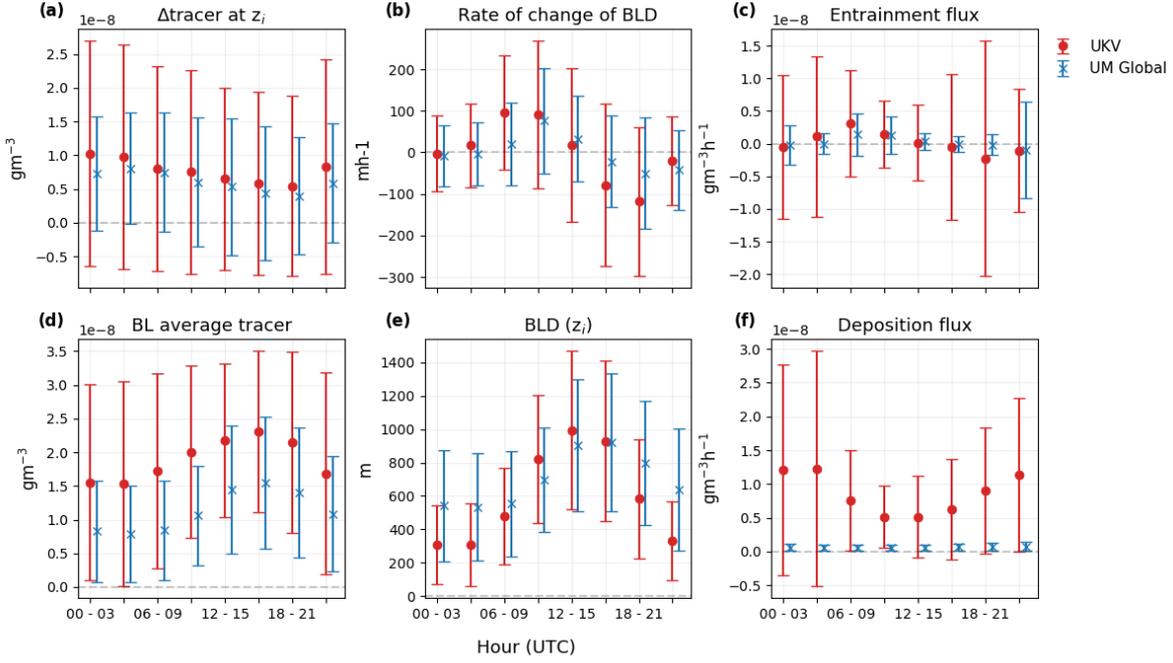
**Figure 5.12:** Diurnal rate of change of surface  $O_3$  at 6 sites from the Kaser et al. (2017) study. Red line represents WRF-Chem simulations, black are observations, with shading showing the minimum and maximum values. Circles represent entrained  $O_3$  through boundary layer growth, with error bars as minimum and maximum values.

cycle of  $O_3$  entrainment through boundary layer growth. At 7 - 9 LT, the entrained  $O_3$  in both model and observation matches the observed average surface  $O_3$  rate of increase. This implies that entrainment through boundary layer growth accounts for most, if not all, of the observed morning  $O_3$  increase and is the most important mechanism for surface  $O_3$  increase before 10 LT. After this time, other processes such as chemical production and loss, deposition or advection have a greater influence on the surface concentrations than entrainment.

### 5.6.2.2 Results of Eulerian comparison

First, let us consider the individual terms contributing to the vertical gradient flux in equation 5.14, which are shown in figure 5.13. Tracer gradient ( $\Delta$ tracer) at boundary layer top (5.13 a), rate of change of boundary layer depth (5.13 b) and boundary-layer average tracer concentrations (5.13 d) are all individual components of the entrainment flux (5.13 c). Boundary-layer average tracer concentrations are also used in the dry deposition flux term (5.13 f), alongside the boundary layer depth  $z_i$  (5.13 e). It is clear from the  $\Delta$ tracer distributions that tracer concentrations in the free troposphere are systematically greater than concentrations within the boundary layer. This is contrary to the Kaser et al. study, where the gradient is positive during the early morning but changes sign further into the afternoon due to accumulation of  $O_3$  following morning entrainment and larger source terms such as photochemical production within the boundary layer. In order to ensure a diurnal cycle of  $\Delta$ tracer which matches that in the Kaser et al. study, surface-released tracers could also be considered within the budget equation. However, even with mean positive  $\Delta$ tracer throughout the day, there is a negative entrainment flux term in the afternoon (figure 5.13 c) due to a decreasing boundary layer depth after 12 - 15 UTC in figure 5.13 b, whereas this remains positive until 16 LT in the Kaser et al. investigation. Consequently, tracer entrainment flux into the boundary layer is positive between 3 - 12 UTC in both UKV and UM Global simulations, as expected from comparison with figure 5.12.

The dry deposition flux in figure 5.13 f is interesting because not only are the UKV values much greater than UM Global, but they are also much larger than the UKV entrainment flux, nearing an order of magnitude difference between 0 - 6 UTC. The deposition term is larger in the UKV simulations than UM Global for two reasons: firstly, the boundary layer-average tracer is greater in UKV than UM Global simulations by a factor of 1.5 - 2 throughout the day (figure 5.13 d). Secondly, the value of dry deposition velocity

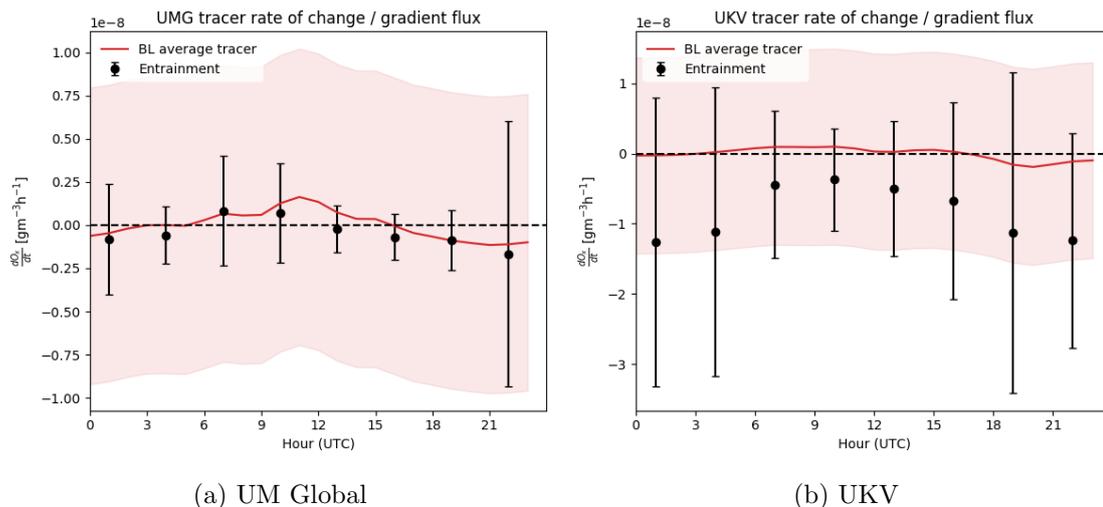


**Figure 5.13:** Mean and standard deviations in NAME simulations with UM Global (blue) and UKV (orange) simulations for (a) tracer gradient across the boundary layer top  $z_i$ , (b) rate of change of boundary layer ( $\frac{dz_i}{dt}$ ), (c) entrainment flux (i.e. first term of equation 5.15), (d) boundary-layer average tracer concentrations, (e) boundary layer depth  $z_i$  and (f) deposition flux term (i.e. second term of equation 5.15).

$v_d = 0.05ms^{-1}$  in the NAME simulations using UKV meteorology, while  $v_d = 0.01ms^{-1}$  for the UM Global simulations. The larger value of  $v_d$  for the UKV simulations is intended to compensate for the over-estimation of entrainment in UKV, as discussed earlier in the chapter. However, even by forcing the tracer concentrations in UKV simulations to have a realistic diurnal profile with a larger value of  $v_d$ , the problem of budget closure in the UKV simulations persists. The assumption of  $w_s = 0$  is invalid for the UKV simulations, because mean  $w_s$  at boundary layer top is non-zero and positive (see section 5.4.3).

The vertical flux gradient (i.e. equation 5.15, difference between entrainment flux and deposition flux) from NAME simulations using both UKV and UM Global meteorology is directly compared against the rate of change of boundary-layer averaged tracer concentrations from the simulations in figure 5.14, as a like-for-like comparison with figure 5.12.

For the NAME tracer simulation using UM Global meteorology, the relationship between rate of change of tracer (red line) to the vertical flux gradient (black dots) resembles the relationship of observed  $\frac{dO_3}{dt}$  and entrainment in the Kaser et al. study. This is reflected in the maintenance of a realistic diurnal cycle of boundary layer tracer within the NAME simulation. In the morning (6 - 9 UTC), the Eulerian budget calculation coincides



**Figure 5.14:** Average rate of change of boundary layer-average tracer in NAME simulations with (a) UM Global meteorology and (b) UKV meteorology, over 53 locations. Shaded area is the standard deviation. Black circles represent the Eulerian budget representation of entrainment at the same locations, with error bars representing standard deviation.

with the Lagrangian tracer rate of change, confirming that entrainment via boundary layer growth dominates the changes in boundary layer-average tracer concentrations during the morning. For 9 - 12 UTC, entrainment still contributes to at least 50% of the changes in boundary layer-average concentrations, but now other processes not accounted for in the budget (e.g. horizontal advection) are likely to contribute to the tracer concentrations. Note that there are no chemical transformation in the simulations, so unlike in the Kaser et al. study, chemical sources and sinks are not contributing to the total rate of change. In the afternoon (from 12 - 15 UTC onward), the gradient flux becomes negative due to gradient reversal of the boundary layer growth seen in figure 5.13b.

Conversely, the UKV simulation shows a constantly negative vertical gradient flux in figure 5.14, as it is dominated by the much larger deposition (figure 5.13f) than entrainment flux term (figure 5.13c). This means that the tracer budget is not closed for the UKV simulation. However, its boundary layer-average tracer diurnal profile is similar in shape to the UM Global simulation (figure 5.13d), which could mean that there is a missing term in the budget equation for the UKV simulation, or an invalid assumption (other than  $w_s = 0$ ) was made.

### 5.6.2.3 Conclusions of Eulerian budget comparison

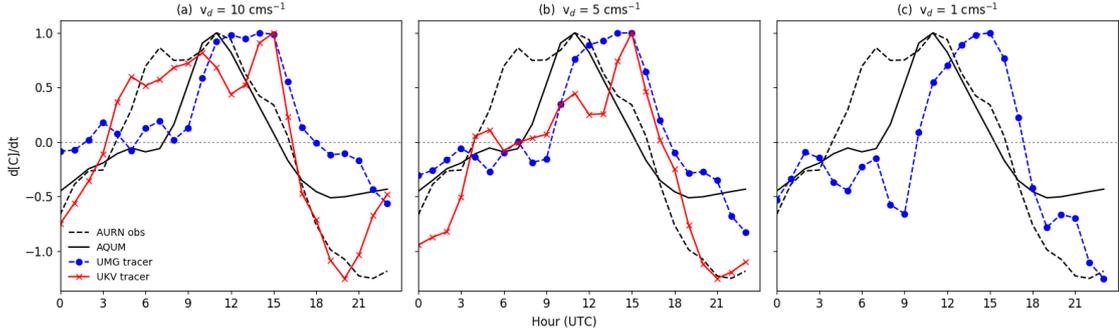
In the NAME simulations with UM Global meteorology, entrainment of free-tropospheric tracer via boundary layer growth is a dominant mechanism in the morning increase of boundary layer-averaged concentrations (and therefore also the surface

concentrations, due to the tracer being well-mixed). Lagrangian NAME simulations were compared against the  $O_3$  closure equation used in an equivalent  $O_3$  budget study of [Kaser et al. \(2017\)](#), who also conclude that entrainment is a dominant source for surface  $O_3$  concentrations before 11 LT. Unfortunately, this also reveals inconsistencies with the NAME simulation using UKV meteorology, for which the value of  $v_d$  already had to be multiplied by 5x more than in the UM Global simulations, with the intention of reproducing the diurnal cycle. The tracer budget equation for UKV simulations cannot be closed because the assumption that  $w_s = 0$  is invalid, but including this term results in unphysically large entrainment velocity  $w_e$ . Therefore, comparing the NAME tracer study with the Kaser study has enabled us to better understand why the UKV tracer simulations result in much larger values of entrained tracer than UM Global simulations. The implication is that large values of resolved vertical motion  $w_s$  at the boundary layer top are dominating entrainment in UKV simulations with NAME.

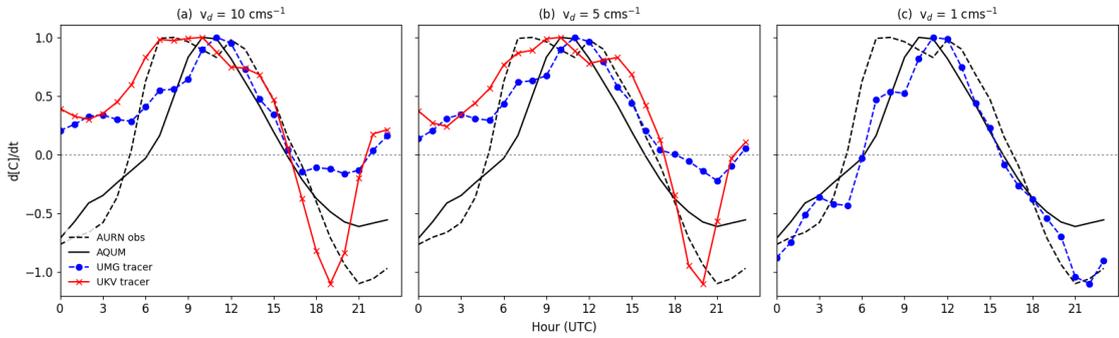
In this section, results were presented to show that there is a problem with using large values of  $w_s$  in the NAME tracer simulations. One might expect large  $w_s$  to be associated with unstable or neutral boundary layers (due to strong buoyant generation of turbulence, or strong wind shear). Higher wind speeds are likely to result in shear-dominated generation of turbulence, therefore stronger TKE and larger vertical motion across the boundary layer top than during low wind speed conditions. Stronger winds are generally associated with low-pressure synoptic conditions. Weaker winds are often present during high-pressure synoptic conditions, which are also associated with high incoming solar radiation (and thus buoyant production of turbulence). The following section applies a synoptic low- and high-pressure partitioning to the analysis of NAME tracer simulation diurnal cycles and examines their differences within the different synoptic regimes.

### 5.6.3 Synoptic variability: diurnal cycles

In order to further characterise the role of meteorology on tracer behaviour, the study was sub-sampled by low and high pressure days through a visual inspection of the Met Office surface pressure analysis charts, as described in section [5.4.1](#). The UKV evaluation for  $v_d = 1 \text{ cm s}^{-1}$  was not included due to it being un-physical as discussed in the previous section. Here, June 2017 is sub-sampled into days of generally high pressure and anti-cyclonic conditions (11 days), and low pressure with cyclonic conditions and / or stronger winds based on isobars from surface charts (9 days). Some days (10) are undetermined and therefore not included in the analysis.



**Figure 5.15:** Rate of change of boundary layer-averaged background free-tropospheric (BGFT) tracer concentrations for low pressure days in June 2017, using UKV and UM Global meteorology and averaged over 53 sites. (a)  $v_d = 10 \text{ cm s}^{-1}$ , (b)  $v_d = 5 \text{ cm s}^{-1}$  and (c)  $v_d = 1 \text{ cm s}^{-1}$ .  $d[C]/dt$  is a dimensionless quantity due to normalisation factor, but native in  $\mu\text{gm}^{-3}\text{h}^{-1}$ .



**Figure 5.16:** Same as figure 5.15, but for high pressure days.

From figures 5.15 and 5.16, it is clear that the maximum rate of increase in observed  $O_x$  concentrations (dashed, black line) occurs later in the day (around 12 UTC; figure 5.15) for low-pressure days than for high-pressure days (around 7 UTC; figure 5.16). The rate of change in AQUM forecast lags the observations in both pressure regimes. However, peak rate of increase occurs at the same time in forecast and observations (11 UTC) during the low pressure days (figure 5.15), despite the forecast lagging the observations throughout the morning (4 - 10 UTC).

The maximum rate of increase for tracers with UKV tracer is sensitive to both the pressure regime and choice of  $v_d$ . During the low pressure days and with  $v_d = 10 \text{ cm s}^{-1}$  (panel a), there is similarity between the morning (00 - 10 UTC) profile shape of the UKV tracer and observed  $O_x$  concentrations; likewise, the UM Global profile resembles the AQUM concentrations between 4 - 12 UTC. The similarities break down when  $v_d = 5 \text{ cm s}^{-1}$  is used, more so for the UKV than UM global simulations as the UKV  $\frac{d[C]}{dt}$  profile has a sharp peak at 15 UTC. The UKV profile is non-comparable by the time  $v_d$  is decreased to  $1 \text{ cm s}^{-1}$  and is therefore not shown in 5.15c. Panel b suggests that

in low-pressure conditions, even when  $v_d = 5\text{ cm s}^{-1}$ , the UKV still entrains too much tracer-rich air and produces an unrealistic diurnal cycle.

For  $v_d = 10$  or  $5\text{ cm s}^{-1}$ , the UKV and UM Global tracer profiles are more alike during the high-pressure than the low-pressure sub-sample (compare figure 5.15 b,c to 5.16 b,c). A possible reason for this is because wind speeds are generally lower during high-pressure than low-pressure conditions, and wind shear is one method of generating turbulence. It would therefore be expected that in high-pressure conditions, there is less interchange of air masses across the boundary layer top than during low-pressure conditions. This would mean that  $w_s$  is more similar between the two models during high-pressure than low-pressure conditions. There is also a possibility that the sub-sample of low pressure days (9) is too small to provide robust results. One way to make the study more robust would be to increase the sample size by evaluating also July and August.

The high pressure days have a negative rate of change of forecast and observed pollutant between 17 UTC through to 5 UTC (6 UTC for the forecast), meanwhile the tracer experiments using both UM Global and UKV meteorology remain positive from 23 UTC and through into the morning and afternoon for  $v_d > 5\text{ cm s}^{-1}$ . The positive gradient throughout the night means that the tracer experiments experience too much night-time tracer entrainment from the residual layer relative to tracer loss, as clearly both the observations and the forecast experience the opposite effect. This only happens during the high-pressure sub-sample of the data. In fact, the average boundary layer concentrations of BGFT tracer during high pressure days are smaller than during low pressure days. This could be an artefact of the nature of the semi-idealised experiment configuration, whereby some tracers released at the lateral boundaries of the domain might never reach the land surface if transported by weak winds.

The fact that both tracer simulations behave similarly during high pressure conditions, but differently during low pressure days, could indicate that the regime - and in particular, vertical and horizontal wind speed - influences the amount of large-scale transport of air into and out of the boundary layer in the NAME simulations. This could be a physical problem due to the use of the entrainment parametrisation scheme within NAME when the homogeneous boundary layer scheme is used. This is tested through the use of both the inhomogeneous and homogeneous boundary layer schemes with both UKV and UM Global tracer simulations, over a period of one week, 19-06-2017 to 25-06-2017. This period was chosen because synoptic conditions over the UK are generally within a high pressure regime for the first part until 22-06, after which a front comes through and the

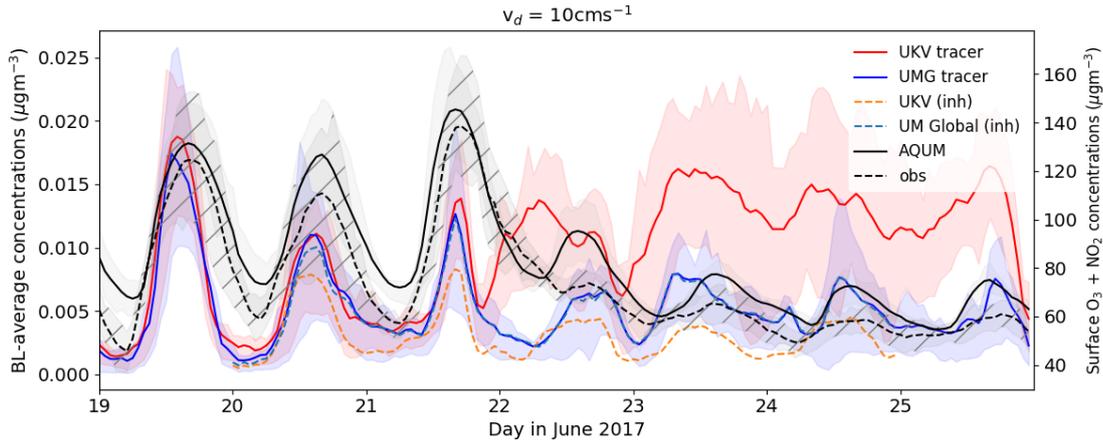
country is under a low-pressure regime from 23-06. The change in synoptic regime has implications for entrainment, where it is expected that stronger wind speeds within the low-pressure regime will enhance the entrainment process. The homogeneous and inhomogeneous simulations are also compared against  $O_x$  observations and AQUM forecast concentrations.

#### 5.6.4 Synoptic variability: case study of one week

The UK experienced an episode of widespread moderate surface  $O_3$  concentrations between 19 - 22nd June (see figure 5.17). On the 21st June, 6 separate AURN sites exceeded hourly  $O_3$  concentrations of  $\geq 180 \mu\text{gm}^{-3}$ , accompanied by anomalously elevated concentrations as measured throughout the AURN network. This event can be attributed to the prolonged stagnant, high pressure conditions (P. Agnew, personal communication). Concentrations throughout the country began to reduce in the evening of the 21st June with the influx of colder air from the Atlantic, as depicted in figure 5.17. The time-series simulated with  $v_d = 5 \text{ cm s}^{-1}$  looks very similar but is not shown.

For the same week, superimposed on figure 5.17 is the BGFT tracer profile from UKV and UM Global simulations with the homogeneous boundary layer scheme. Concentrations are averaged over 53 locations, beginning at 0 UTC on 19th June 2017 and the simulations use  $v_d = 10 \text{ cm s}^{-1}$ . For the first 66 hours, the variability of mean BGFT tracer concentrations throughout the boundary layer using both the UKV and UM Global meteorology is comparable to the AQUM forecast and surface observations. As the synoptic regime changes from high- to low-pressure, the mean tracer profiles diverge and become significantly different from 0 UTC on 22-06 until 22 UTC on 25-06.

The UKV tracer profile in figure 5.17 suggests that during low pressure conditions, greater amounts of BGFT tracer contribute to boundary layer-average concentrations than during high pressure conditions. On the other hand, the UM Global tracer profile during the second half of the week is similar to the AQUM forecast, although it over-estimates the observations after 23-06. The entrainment parametrisation scheme in NAME was discussed in section 5.3.1.1, where it was shown that the scheme uses resolved vertical winds  $w_s$  to determine whether particles are transmitted through the boundary layer top. There are large differences in resolved vertical winds  $w_s$  between UKV and UM Global (shown in section 5.4.3), and those differences are likely to be exacerbated during strong wind conditions.



**Figure 5.17:** 7-day time-series (19 / 06 / 2017 to 26 / 06 / 2017) of boundary layer-average concentrations of entrained background free-tropospheric tracer, averaged over 53 urban locations using  $v_d = 10\text{cm s}^{-1}$ . Homogeneous boundary layer scheme averages using UKV meteorology (red) and UM Global meteorology (dark blue), plotted along inhomogeneous results using UKV meteorology (orange, dashed) and UM Global meteorology (light blue, dashed) - all corresponding to left y-axis. AQUM forecast of  $O_x$  (solid, black) and the AURN surface observations (black, dashed, hatched shading) averaged over the same locations, both corresponding to the right y-axis. Shading represents inter-quartile range.

### 5.6.5 Inhomogeneous boundary layer simulations

One way to confirm that the distribution of  $w_s$  at boundary layer top in UKV simulations is influencing entrainment is to repeat the experiment using the inhomogeneous boundary layer scheme in NAME (introduced in section 5.3.1.1), as it bypasses the requirement to parametrise entrainment at boundary layer top. This means that any differences in the representation of entrainment, directly related to the varying magnitudes of  $w_s$  between the UKV and UM Global simulations, are neglected.

The experiment is repeated over the period of 19-06-2017 to 25-06-2017, again with  $v_d = 10\text{cm s}^{-1}$  and shown with the dashed lines in figure 5.17. The model spin-up period (00:00 19-06 to 00:00 20-06) is omitted. For the UKV tracer, results of the inhomogeneous case are significantly different from the homogeneous case, especially during the low-pressure conditions. Where in the homogeneous experiment, UKV tracer was showing signs of un-physical rates of entrainment into the boundary layer from 22-06, the inhomogeneous case gives values which are comparable to the UM Global tracer. Inhomogeneous UKV tracer concentrations are smaller by around  $2 \times 10^{-3} \mu\text{gm}^{-3}$  than the UM Global tracer during low-pressure conditions. During the high-pressure conditions (20-06 to 22-06), the inhomogeneous UKV and UM Global simulations capture the larger diurnal variability in concentrations seen in the observations and AQUM forecast.

The boundary layer-average tracer concentrations profile for the UM Global simulations with the inhomogeneous scheme is very similar to the homogeneous profile throughout both synoptic regimes. This result suggests that using the cheaper homogeneous scheme in NAME is fine for low-resolution NWP input (i.e. UM Global), but will lead to significant differences for higher resolution NWP data (i.e. UKV).

Using the inhomogeneous scheme, which bypasses the requirement for an entrainment parametrisation at the boundary layer top, gives results for the UKV tracer which are much closer to UM Global tracer during windy conditions than when using the homogeneous scheme. This is evidence to support the postulate that the entrainment scheme in NAME is unsuitable to deal with large vertical velocities from high-resolution NWP (i.e. UKV) and should be used with caution. Therefore, when high resolution NWP data is used for dispersion modelling with NAME, it is best to use the inhomogeneous boundary layer scheme, although it is much more memory intensive than the homogeneous scheme. This has implications for the results of this chapter, where if the entire month's worth of UKV simulations were to be repeated, the contribution of entrainment to boundary layer-average tracer may not be as significant as it is when using the homogeneous scheme. This is an important result for future development of air quality and dispersion models, especially as NWP resolution increases and, as seen with the UKV data, could result in over-estimation of pollutants during windy conditions.

## 5.7 Summary of results, conclusions and discussion

The process-based method of forecast evaluation presented in this chapter was motivated by results found in chapters 3 and 4. It was suggested in chapter 3 that errors in the diurnal profile of modelled  $O_3$  concentrations are influenced by the diurnal profile of the boundary layer depth. Here, the hypothesis was tested that a temporal lag in the growth of urban boundary layer depth in the air quality forecast model AQUM results in a lag in the modelled  $O_x$  concentrations in the morning due to a delay in downward vertical transport of  $O_3$ -rich air from the residual layer. Using a free-tropospheric tagging technique in NAME, tracers originating from above the modelled boundary layer which are entrained into it were identified. The meteorological data used in the experiments are from UKV and UM Global configurations of the MetUM, which differ substantially in their modelled boundary layer depth in the morning hours and rate of change throughout the study period.

The second research question concerned the reflection of changes in synoptic pressure

conditions within the  $O_3$  and  $NO_2$  forecasts. Of particular interest were tracers transported from outside of the domain or entrained from the residual layer - i.e. no local land sources. A one-week case study was identified where synoptic conditions in the UK changed from predominantly anti-cyclonic to cyclonic, with an evaluation of the ability of the tracer simulations with UKV and UM Global meteorology to capture this change in regime for non-local tracer.

### 5.7.1 Summary of results

The main results from this study are as follows:

- Entrainment and deposition processes alone can re-create the diurnal evolution of  $O_x$  when UM Global meteorology is used.
- There is a lag of at least 1 hour in the rate of maximum increase of entrained tracer from lateral boundaries ('BGFT') between the UM Global and UKV simulations. This temporal lag is due to the delay in initiation of morning boundary layer growth in UM Global, compared to the UKV. Because AQUM is a similar model configuration to the UM Global, the lag in boundary growth in the latter could result in a delay in the morning increase of surface  $O_x$  within the AQUM forecast.
- By calculating the vertical gradient flux from an Eulerian perspective, entrainment accounts for all of the tracer increase between 6 - 9 UTC in UM Global simulations, which is consistent with the findings of another modelling study verified by observations (Kaser et al., 2017). The contribution of entrained tracer to boundary layer-average concentrations decreases in the afternoon. UKV simulations have an entrainment flux term which is much larger in amplitude than the rate of change of mean tracer concentrations. The implication of this is that the UM Global simulations realistically represent the process of entrainment and can be compared with forecast and observed  $O_x$ , whereas the UKV simulations have some imbalance. The entrainment flux term is negative throughout the day in the UKV simulations, which implies constant detrainment if resolved vertical wind component  $w_s$  is (incorrectly) assumed to be zero in the budget equation. Otherwise, including a non-zero  $w_s$  term makes the vertical gradient flux un-physical.
- There is a loss of diurnal cycle and larger differences between tracer concentrations from UM Global and UKV during windy or low-pressure conditions than during high-pressure conditions.

- During low-pressure conditions, NAME with UKV meteorology simulates boundary layer-average tracer concentrations which are significantly too large, when both  $v_d = 5$  and  $10 \text{ cm s}^{-1}$  are used. This happens when NAME's homogeneous boundary layer scheme is used, which is less computationally intensive to run than the inhomogeneous scheme.
- However, when the inhomogeneous scheme is used, the magnitude of UKV boundary-layer average tracer during windy conditions is much smaller than with the homogeneous scheme. The tracer amount is then comparable to that from UM Global, although both simulations now have less boundary-layer tracer than the AQUM forecast or observations.

### 5.7.2 Chapter conclusions and discussion

This chapter investigated the influence of different NWP input on  $O_x$ -like tracers within a NAME simulation. The 1 - 2 hour lag in  $O_x$  concentration increase in the AQUM forecast is mirrored by tracers within the NAME-UM Global simulations, which shares most of AQUM's parametrisation characteristics. The offset in initialisation of morning surface tracer increase is less evident in the NAME-UKV simulations, which is likely due to its more realistic representation of the surface energy budget with the MORUSES surface tile scheme, which results in more accurate timing of boundary layer development. This difference suggests that a better representation of the surface heat fluxes in AQUM could result in a more accurate timing of the forecast  $O_x$  morning increase, by improving the modelling of boundary layer development. Because the simulated tracers are directly comparable to  $O_x$  concentrations within AQUM, a main conclusion of this chapter is that the representation of boundary layer development directly influences the morning increase timing offset between the AQUM forecast and observations.

Morning boundary layer growth is important for surface pollutant concentrations for many reasons, including but not limited to the fact that the process of its growth is related to entrainment of air (and any molecules) from the free troposphere; and also its depth dictates the volume that the molecules can occupy. Entrainment is a process that could contribute significantly to the morning increase of boundary layer-average  $O_x$ , and thus also to surface concentrations (assuming a well-mixed daytime boundary layer).

One of the aims of this chapter was to quantify this contribution. Results from this Lagrangian framework were compared against an Eulerian budget evaluation, where it was found that the entrainment contribution to surface tracers within the NAME-UM

Global simulation dominated between 6 - 9 UTC, and still contributed to at least 50% between 9 - 12 UTC. Because UM Global shares most of its parametrisation characteristics with AQUM, it can be concluded that the underlying meteorology in AQUM - including the entrainment process - is an important factor in determining the diurnal profile of the forecast surface  $O_x$  concentrations, even in the absence of chemical production and losses. Entrainment should therefore not be overlooked within future air quality model development.

In contrast, the budget for the NAME-UKV simulations was not closed due to invalid assumptions about vertical velocities ( $w_s$ ) at the boundary layer top. The assumption that  $w_s$  could be set to zero was invalid, because UKV resolves larger values of  $w_s$  at the boundary layer top than UM Global, possibly as a direct result of smaller grid spacing. This resulted in too high an entrainment flux within NAME's entrainment parametrisation scheme. This effect was found to be enhanced in low-pressure synoptic conditions, consistent with the expectation that vertical (and horizontal) mixing could be greater than during high-pressure conditions, thus resulting in greater magnitudes of resolved vertical velocities, and hence enhanced entrainment. A lesson learnt from the above is that resolution of the underlying NWP, as well as whether boundary layer mesoscale processes are resolved, needs careful consideration if NAME is to replace AQUM as the UK's operational air quality forecast model, as well as for research case studies.

One way of bypassing the use of the parametrisation scheme is to set the turbulence parameters to vary with height within the boundary layer ('inhomogeneous' scheme), instead of taking the vertical average ('homogeneous' scheme). By considering a one-week case study, it was shown that the use of the more expensive inhomogeneous scheme resulted in more realistic entrainment values within the NAME-UKV simulations during both low- and high-pressure days. From this, it can be concluded that modelling studies using NAME should beware errors in the concentrations of atmospheric species arising from the use of the computationally cheaper homogeneous scheme. Alternatively, they could consider using a short simulation time-step and particle synchronisation time (e.g. 5 minutes instead of 15). This also applies for future development of air quality models in general, as the Met Office are currently investigating how to re-purpose NAME for operational air quality modelling.

The study presented in this chapter was mostly limited by computational power and available time. The question of how much influence does the difference in model resolution between the UM Global and UKV make to perceived tracer results has not

been addressed, and it may well turn out that resolution could have an important role to play. If more time were available, the resolution of the UKV input to NAME could be down-scaled to match that of the UM Global, before re-running the experiments. One way of further ensuring robustness of results would be to extend the study period from just 30 days to one or more summer seasons, thus vastly increasing the data sample - in particular when the study period is further partitioned into synoptic conditions. Furthermore, the inhomogeneous scheme could be used instead of the homogeneous scheme, in order to bypass the inconsistencies demonstrated with the use of higher resolution UKV data within NAME's entrainment scheme. The expected benefit of using the inhomogeneous scheme would be that the enhanced entrainment rates simulated with the UKV meteorology would be significantly reduced. However, the computational time that this would take would be greatly increased. A sensitivity study over a shorter period with the inhomogeneous scheme would suffice to demonstrate this. Thus, I acknowledge the possibility that the importance of entrainment to surface tracer concentrations may have been over-estimated by using the homogeneous boundary layer scheme in the experiments with UKV meteorology due to the use of NAME's entrainment parametrisation scheme.

Another consideration is the use of high values for dry deposition,  $v_d$ . This was used as a proxy for chemical losses, which are not represented explicitly in the simulations, in order to re-create a valid mean diurnal profile of the tracers. For example, in the UKV simulations,  $v_d$  was set to an unrealistically high value of  $5 \text{ cm s}^{-1}$  (where the observed range over land is  $0.5 - 1 \text{ cm s}^{-1}$ ) in order to correct the tracer concentration gradient caused by enhanced entrainment (relative to the UM Global simulations). In other words, a wrong value of  $v_d$  is used to correct for errors in the representation of entrainment. This presents an issue because the resultant diurnal profile may seem 'correct', but for the wrong reasons. The gradient reversal, which is corrected with the use of high  $v_d$ , is likely caused by the relatively large magnitude of resolved vertical velocities  $w_s$  in the UKV simulations, which result in inconsistencies within the entrainment scheme when the homogeneous boundary layer scheme is used. Once again, if the experiment were to be repeated using the inhomogeneous scheme, perhaps dry deposition rates could be set to a more realistic value in the UKV simulation, as there may not be a need to offset the reversed vertical tracer concentration gradient.

Consistent with the findings of e.g. [Travis et al. \(2016\)](#), and despite the outlined limitations, the present study provides evidence to support the hypothesis that diurnal errors in surface  $\text{O}_x$  forecasts are strongly related to errors in the evolution of boundary

layer depth, whose importance should not be overlooked within the realm of improving surface pollutant forecasts. Although single observation campaigns such as the one of [Trousdel et al. \(2016\)](#) use observed entrainment velocities to determine the contribution of photochemical production to surface  $O_3$  in rural areas, there does not currently exist a network of routine observations of entrainment velocities. Given that this would likely involve eddy-covariance methods, it is unlikely that such a network will exist in the future, and observation studies are likely to be campaigns over a specific area and period in time. This semi-idealised tracer experiment thus offers a simple but insightful technique for statistical analysis of the effect of modelled boundary layer entrainment on surface  $O_x$  concentrations. It could also be extended to work with other pollutants transported from farther afield.

## 6. Conclusions

The work in this thesis was completed in order to determine the influence of meteorological forecast errors on the forecast errors of regional AQ variables within AQUM, through implementing various categories of evaluation methods. Meteorology serves an important role in AQ forecasting, because boundary layer processes such as precipitation, dispersion, turbulence and entrainment are all responsible for transporting and / or removing pollutants on both small and large length-scales (Seaman, 1999). Errors in meteorological variables are likely to propagate to the pollutant forecast, as confirmed by modelling studies such as Zhang et al. (2007) which demonstrates the uncertainties in predictability of an elevated urban O<sub>3</sub> concentrations event by perturbing initial meteorological conditions, such as surface wind speed and temperature. Therefore, through identifying and understanding errors in wind speed, the representation of boundary layer stability and entrainment, or precipitation within a coupled AQ-meteorology forecast model, developers can begin to improve the AQ model so as to minimise pollutant forecast errors rooted in meteorology.

This thesis has demonstrated an initial evaluation of the suitability of neighbourhood-based (Ebert, 2009; Mittermaier, 2014) and process-based verification techniques in order to learn more about the relationships between the forecast errors than those demonstrated with traditional point-based verification metrics. For example, the SO-NF method (Ebert, 2008) allows some degree of displacement between the observation site and the grid-point forecast, accepting that the forecast can still be useful even if the feature (e.g. a precipitation front) was not predicted to be in quite the right place at the right time. It is proposed that through constructing a distribution of possible values from within a neighbourhood around the point observation, a deterministic forecast can be evaluated against the observation on a probabilistic basis, which can be beneficial for high resolution forecasts of certain meteorological variables (Mittermaier, 2014). By evaluating the forecast over a neighbourhood rather than at grid-scale only, the existence of grid-scale relationships between any forecast errors could be amplified when larger areas are considered.

For example, meteorological errors close to, but not at, the measurement site could be influencing pollution forecast error. The neighbourhood scale at which the relationships are strongest may therefore be used to better understand the nature of the relationships, furthering any knowledge gained about the forecast errors from point-based metrics. Recently, the SO-NF method was successfully applied in oceanography (Maksymczuk et al., 2020; Crocker et al., 2020) but to the best of my knowledge has not yet been applied to AQ. It was a unique opportunity to test the suitability of the SO-NF method to evaluate AQ forecasts directly within the HiRA framework.

The remainder of this chapter is structured as follows: in sections 6.1 through to 6.3, main conclusions arising from this thesis will be presented in the context of the original research questions from section 1.4. A summary of the research questions and answers follow in section 6.4, with a discussion and further questions yet unanswered in this thesis will be presented in section 6.5.

## 6.1 Impact of 10 m wind speed errors on O<sub>3</sub> and NO<sub>2</sub> forecast errors

### Point-based metrics

In chapter 3, the 10 m wind speed forecast was demonstrated to have the strongest positive bias during the night and early morning, similar to that of other European AQ models (Brunner et al., 2015). This coincides with an over-estimation of mean surface O<sub>3</sub>, where the morning increase in forecast concentrations lags the observations by 1 - 2 hours. The mean NO<sub>2</sub> forecast diurnal profile has a negative bias and is anti-correlated with the observations, most likely because of its close association with O<sub>3</sub> via chemical processes. It was hypothesised that the leading cause of the morning lag in the increase of O<sub>3</sub> and NO<sub>2</sub> is a physical process, rather than a chemical one. This hypothesis is consistent with Im et al. (2015b), where it is shown that differences in the chemical or meteorological configurations within the same model framework can lead to significant differences in the prediction of O<sub>3</sub> concentrations.

In order to address the first overarching research question, “How are the forecast errors in O<sub>3</sub> and NO<sub>2</sub> related to forecast errors in 10 m wind speed?”, a comparison between the top percentiles of the point-based error distributions was done. The comparison of O<sub>3</sub>, NO<sub>2</sub> and 10 m wind speed forecast errors directly revealed that the top 10th percentile of the wind speed errors coincided with above-average O<sub>3</sub> (positive) errors, and

conversely with above-average NO<sub>2</sub> (negative) errors. The strength of this relationship leads to the conclusion that, when evaluated at the grid-scale, over-estimating night-time 10 m wind speed in AQUM increases the negative forecast bias in NO<sub>2</sub> and positive bias in O<sub>3</sub>. Consequently, 10 m wind speed is one variable whose improved forecast accuracy within AQUM could lead to a reduction in the biases of O<sub>3</sub> and NO<sub>2</sub> summertime forecasts between 00 - 08 UTC.

One possible way of reducing the surface wind speed bias would be to consider the effect of using stability functions at night. Specifically, the ‘long-tailed’ stability functions are currently used within the boundary layer parameterisation scheme in order to increase mixing during stable night-time conditions, which reduces radiation cooling to keep the surface temperatures accurate.

## Neighbourhood-based metrics

One of the other research questions posed at the beginning of this thesis was, “Can a probabilistic neighbourhood evaluation provide insights into the nature of relationships between meteorological and pollutant forecast errors?”. Chapter 4 therefore explored whether the spatial footprint of the surface wind speed forecast error contributes to the enhanced forecast O<sub>3</sub> concentrations and greater dispersion of NO<sub>2</sub>. In order to evaluate the impact of 10 m wind speed forecast errors within the vicinity of an AQ monitoring site on errors at the location, the SO-NF method was applied to the wind speed, O<sub>3</sub> and NO<sub>2</sub> forecasts, using BS and CRPS error metrics. These scores were calculated by forming a pseudo-ensemble forecast from a square neighbourhood of grid-boxes around a point-observation in increasing size, i.e. 3x3, 5x5, 7x7 etc. The larger the neighbourhood over which error correlations appear the strongest, the larger the footprint of influence.

A statistically significant negative correlation was found between 10 m wind speed (4 - 6 ms<sup>-1</sup>) and NO<sub>2</sub> forecast errors during the morning hours at both urban and rural locations, for neighbourhoods of size < 3x3 (36<sup>2</sup>km). This means that over-estimating wind speed > 6 ms<sup>-1</sup> is related to larger negative NO<sub>2</sub> bias, perhaps due to too much dispersion in the model on local scales. No significant correlations were found for neighbourhoods larger than 3x3. However, while the O<sub>3</sub> and O<sub>x</sub> evaluation with wind speed revealed a positive correlation in the forecast errors at urban background sites during the morning, the strongest correlations were found at 5x5 (60<sup>2</sup>km) neighbourhoods for O<sub>x</sub>, and 7x7 (84<sup>2</sup>km) for O<sub>3</sub>. This finding supports the hypothesis that an over-estimation of moderate wind speed is related to increased error in O<sub>3</sub> and O<sub>x</sub> concentrations on regional

scales.

A reason for this could be that for 10 m wind speeds exceeding a threshold (noted by [Sun et al. \(2012\)](#) to be around  $5 \text{ ms}^{-1}$  during stable conditions), turbulent kinetic energy can still result in exchange of air masses across the boundary layer top, i.e. entrainment. As previously noted, the residual layer can have higher concentrations of  $\text{O}_3$  than the boundary layer below, as the previous day's  $\text{O}_3$  would have accumulated and become decoupled upon the emergence of the shallow and stable nocturnal boundary layer during the evening. This results in a positive gradient in  $\text{O}_3$  across the boundary layer top in the early morning hours, which means that air entrained from the residual layer would lead to enhanced concentrations within the boundary layer. Because of the positive bias in wind speed over a large neighbourhood shown in [chapter 4](#), entrainment could be contributing more  $\text{O}_3$  than is observed, thus resulting in larger  $\text{O}_3$  forecast error.

## Limitations

Both the point-based and the neighbourhood approaches to this evaluation spanned short study periods of just one season. This means that the sample size was small, and very particular conditions within each season imply that the study cannot be generalised to other years. A simple way of ensuring robustness of the study would be to increase the sample size by considering e.g. winters or summers spanning an appropriate number of years - 3 or 4 would suffice. A further limitation to this, however, is that AQUM has been only been operational since 2013 and has since then undergone many fundamental changes in configuration, which would introduce a further complexity in the analysis of comparing like-for-like.

Another limitation, perhaps more specific to the neighbourhood approach than the point-based approach, is that independence of sites has not been ensured (apart from the removal of all London sites but one). There may be other overlapping regions of nearby sites, especially when larger neighbourhoods are evaluated. Independence of variables is important for a robust statistical evaluation, therefore the experiment could be improved by ensuring that the sites are far enough apart to represent unique regions within the domain.

## 6.2 Impact of entrainment representation on $O_3$ and $NO_2$ forecast errors

The contribution of boundary layer entrainment to surface  $O_3$  and  $NO_2$  forecast error was examined by designing a tracer dispersion experiment within NAME (Jones et al., 2007). This evaluation examined the impact of boundary layer development and entrainment on the pollutants, therefore addressing another science question outlined at the beginning of this thesis: “Can process-based evaluation be used to inform us about the relationship between meteorological and pollution forecast errors?”. The point based methods of chapter 3 and neighbourhood verification methods explored in chapter 4 provide two complementary techniques of analysing the relationship between the representation of boundary layer processes (i.e. wind speed, boundary layer development and entrainment) with  $O_3$  and  $NO_2$  forecast errors.

In order to minimise potential errors due to emissions or chemistry,  $O_x$  was represented in the simulation with inert tracers, using different representations of boundary layer development from the UKV and UM Global models. The simulation was set up such that there were no sources within the domain, no chemical reactions and no sources / sinks other than dry deposition to the Earth’s surface and advection in / out of the domain. The resulting average diurnal profile of the simulated tracers closely resembled the observations in June 2017, as in chapter 3 where a lag in the morning increase of forecast  $O_3$  concentrations was identified.

It was concluded from the month-long simulations that the delayed rise in  $O_x$  concentration levels in the morning is due to the delayed onset of efficient mixing and subsequent entrainment of air from the night-time residual layer to the growing daytime mixed layer. This finding is consistent with e.g. Hanna and Yang (2001), where too much mixing of  $O_3$  from aloft was shown to be a result of weak night-time temperature inversions at modelled boundary layer top. The evaluation in section 5.4.2 showed that boundary layer develops later in AQUM and UM Global than it does in the UKV, coinciding with the timing of the  $O_3$  error lag. The late boundary layer development in UM Global (and therefore AQUM) relative to UKV is likely due to the representation of the surface energy balance. AQUM and UM Global both use a single-tile approach to parametrise urban regions (Best, 2005), which has been shown to be less accurate than the 2-tile approach of MORUSES (Porson et al., 2010) used in e.g. the UKV. The representation of the surface sensible heat flux directly influences the model’s diagnosis of boundary layer stability, and hence the

amount of vertical advection and turbulent mixing, thus boundary layer development.

A one-week case study was selected to demonstrate that during a synoptic high-pressure period, the large diurnal variability in forecast and observed  $O_x$  was replicated with tracers from both UM Global and UKV simulations. This confirmed that under calm, stable conditions, the meteorology of the tracer simulations alone is able to replicate the forecast and observed cycles, thus emphasising the importance of meteorology on the transport of pollutants in the model. However, during the latter half of the week, synoptic conditions were more unsettled upon the passage of a weather front. During this time, the forecast and observed diurnal cycles reduced in amplitude and were replicated only with the UM Global tracer simulations. Tracer concentrations from the UKV simulations were over-predicted and required further investigation.

Through using the Eulerian budget approach of [Kaser et al. \(2017\)](#), it was concluded that entrainment is a process that contributes significantly to the morning increase of boundary layer-average  $O_x$ , and therefore also to surface concentrations, if the day-time boundary layer is well-mixed. The results show that entrainment is the leading source of tracer concentration increase in the morning (between 6 - 12 UTC) in the UM Global simulations, which is consistent with Kaser et al. However, because of failing to close the budget within the UKV simulation due to inconsistencies within the entrainment scheme when the ‘homogeneous’ boundary layer scheme is used, the hypothesis that entrainment is a dominant source of morning  $O_x$  is not supported with the UKV simulations presented in section 5.6. Instead, it should be noted that any future work with NAME should consider the underlying NWP resolution and whether it explicitly represents or parametrises turbulent parameters, in order to avoid double-counting of turbulent parameters across the boundary layer top.

## Limitations

It was assumed that the mean resolved vertical velocity ( $w_s$ ) was negligible, which could result in under-estimation of the entrainment velocity. This is a fine assumption to make for the UM Global simulation as mean  $w_s$  is small. However UKV resolves higher values of  $w_s$  than UM Global, which directly feed into the entrainment parametrisation scheme in NAME, used within the homogeneous boundary layer scheme. Therefore if the resolved vertical velocities  $w_s$  are large, the entrainment scheme could result in too much exchange of air masses, as in the case of the UKV simulations. This is likely because the horizontal grid-spacing of UKV (1.5 km) is much smaller than UM Global

(17 km in this dataset, although currently it stands at 10 km), therefore UKV can resolve smaller-scale structures and processes. This could lead to double-counting of turbulent motions - ones which are already resolved in UKV, and ones which are parametrised by the NAME homogeneous boundary layer scheme - thus resulting in an un-physically high rate of entrainment, as shown by the mean UKV tracer profile during unstable conditions. Indeed, switching the NAME boundary layer scheme to the inhomogeneous scheme (without the requirement of the entrainment parametrisation scheme) results in smaller tracer concentrations being entrained through the boundary layer top, which are closer to the observations and forecasts than when using the homogeneous scheme. However, the inhomogeneous scheme is computationally more expensive and takes around 5 times as long in real-time for each model time-step due to retaining the particles' memory. There is currently on-going work at the Met Office to adapt NAME's entrainment scheme to work with higher resolution UKV data in the homogeneous boundary layer framework (P. Agnew, *personal communication*).

It is important to be aware of the advantages and disadvantages of using a computationally cheaper boundary layer scheme, especially in light of using NWP data of increasing resolution in both off-line and on-line AQ models. Current on-going work at the Met Office points to the possibility of adapting NAME to work more efficiently with higher resolution meteorology data in order to provide a routine AQ forecast in place of AQUM. It is evident from results presented in chapter 5 that the representation of boundary layer processes and especially entrainment in subsequent iterations of air quality forecast models - whether that is AQUM, a version of NAME or anything else - should not be overlooked.

### 6.3 Impact of precipitation errors on $PM_{2.5}$ and $PM_{10}$

#### Point-based metrics

The other relationship explored in chapter 3 aims to answer one of the remaining research questions: "How are forecast errors in  $PM_{10}$  and  $PM_{2.5}$  related to forecast errors in precipitation?". Using point-based metrics, it was demonstrated that when precipitation is under-estimated, the PM forecast error is reduced. In other words if the precipitation bias was removed, the PM forecast error would be worse. This is because  $PM_{2.5}$  and  $PM_{10}$  both have negative bias in AQUM, therefore the under-prediction of rain has the effect of not removing as much PM as is observed, and thus the negative PM

bias appears smaller. Although it might be expected that this effect should be immediate due to the efficiency of precipitation as a PM sink, the composite error analysis showed that the expected reduction in PM error is delayed by a few hours. The delay could be caused by the fact that the precipitation forecasts are not verified at the same locations as the PM forecasts, since the AURN and WMO sites are not co-located. This means that although care has been taken to ensure that the PM forecast error is compared to the precipitation forecast error at the nearest available WMO site location, there does exist a spatial discrepancy.

Ebert (2009) states that indeed a biased forecast could be more useful than an unbiased one if the errors are compensating. A ‘bad’ forecast for a modeller may actually be a ‘good’ forecast for the end user, if it produces accurate results for the wrong reason. For example, in chapter 3 it was found that if the precipitation forecast were to be improved, the PM<sub>2.5</sub> error would increase. Thus, in the absence of fixing any discrepancies in primary emissions input, the AQ forecaster might actually prefer the underlying meteorology to have biases (e.g. in the precipitation) in order for the PM errors to compensate.

### Neighbourhood-based metrics

Because of spatial discrepancies associated with precipitation forecasts in particular, it is of interest to the scientific community to test the suitability of a neighbourhood forecast verification technique, in order to minimise the double error penalty of the precipitation forecast skill caused by the differences in location. The SO-NF evaluation presented in section 4.6.4 revealed that where relationships between precipitation BS and PM<sub>2.5</sub> exist, they are weak and their sign is site-dependent. For PM<sub>10</sub>, there were no statistically significant results. It can therefore be concluded that statistical neighbourhood-based evaluation methods are not useful for characterising sources of PM forecast error related to precipitation forecast error.

### Limitations

This study faced a number of limitations, some of which were discussed in section 4.6.5. Firstly, seen as though precipitation error itself was shown to not be large in relation to hourly accumulation, there may be other compounding factors which impact the precipitation and PM relationship - namely wind speed, wind direction and emissions, which were not evaluated in conjunction with precipitation but could potentially strengthen some of the relationships presented. Therefore one way of establishing the

influence of meteorological factors on PM could be to include a wind speed / direction analysis alongside the precipitation, as was also proposed for the O<sub>3</sub> and NO<sub>2</sub> work.

Secondly, because this was a statistical evaluation over numerous sites, no case studies were conducted to evaluate specific error patterns at individual sites or even specific meteorological conditions. It was shown that PM forecast errors were unique to each site (e.g. due to local emissions, geographical location, elevation, etc), therefore perhaps it would be appropriate to evaluate some case studies of impact of a particularly poor precipitation forecast on the PM forecast errors in relation to a baseline reference forecast (e.g. persistence). Especially in the context of using a neighbourhood approach, individual site analysis would ensure that the act of increasing the neighbourhood does not deteriorate the dataset by e.g. falsely including parts of the sea within a large neighbourhood, thereby potentially leading to more conclusive results about the strength of these relationships at different neighbourhood sizes.

Finally, the sites included in the study were not necessarily independent of each other, which could lead to inaccuracies of the results within overlapping neighbourhoods. If the experiment was repeated, it should be ensured that sites are independent.

## 6.4 Summary

This thesis has explored numerous avenues of identifying the impact of meteorological forecast errors on AQ forecast accuracy. In particular, these 4 main were questions addressed:

1. **How are forecast errors in O<sub>3</sub> and NO<sub>2</sub> related to forecast errors in 10 m wind speed?**

There is a positive correlation between the top wind speed error percentiles and both O<sub>3</sub> and NO<sub>2</sub> forecast errors at the grid-scale in AQUM.

2. **How are forecast errors in PM<sub>10</sub> and PM<sub>2.5</sub> related to forecast errors in precipitation?**

An error composite evaluation showed that under-estimating precipitation results in a reduced PM forecast error within 6 hours of the error 'event' - this relationship is thus because the PM forecast in AQUM already has a negative bias.

3. **Can a probabilistic neighbourhood evaluation provide insights into the nature of relationships between meteorological and pollutant forecast**

**errors?**

It can, although it is more insightful for establishing the spatial footprint of influence of wind speed forecast error on  $O_3$  and  $NO_2$ , than it is for determining relationships between precipitation and  $PM_{2.5}$  or  $PM_{10}$ .

**4. Can process-based evaluation be used to inform us about the relationship between meteorological and pollution forecast errors?**

Process-based evaluation can certainly help us understand how modelled boundary layer processes within coupled AQ - meteorological forecasts could result in pollution forecast errors. In particular, the impact of entrainment of  $O_3$  into the boundary layer from aloft could be evaluated with an idealised simulation, whereas determining this contribution with a standard point-based analysis would not be viable.

## 6.5 Further questions and discussion

Although the semi-idealised tracer study is a good proxy for analysing the meteorological influence on forecast  $O_x$  concentrations, it would enhance the robustness of the results to directly conduct sensitivity tests within a research configuration of AQUM. How would the AQUM pollutant forecast respond to a direct change in boundary layer development as a result of using the two-tile urban surface scheme in place of the current one-tile scheme? Another sensitivity study could be to vary the magnitude of night-time wind speed bias, which, if set to zero, would either provide further evidence to support, or disprove, the hypothesis that  $O_3$  and  $NO_2$  forecast errors are influenced by the wind speed bias.

One of the other remaining questions arising from the research presented in this thesis is whether the SO-NF technique to evaluate error relationships between precipitation and PM would be more insightful in the case of a high resolution forecast - as it was intended - rather than the regional resolution of AQUM. We might expect to see a stronger relationship between precipitation and PM when using SO-NF if the model gridspacing were smaller. Neighbourhood techniques are particularly beneficial to demonstrate the improvement of precipitation forecasts from a low to a higher resolution ([Mittermaier, 2014](#)), therefore it would be particularly interesting to see whether a future high-resolution development of AQUM could be evaluated against the current configuration using HiRA.

Zooming out further, it is possible that developing a higher resolution forecast than

AQUM's 12 km gridsampling would represent both the meteorology and pollution finer scales (i.e. urban areas) with more accuracy. There is currently on-going work at the Met Office to incorporate accurate convection modelling in city-scale NWP (Lean et al., 2019), whereby the improved representation of surface temperature and convective overturning could be useful to routine AQ forecasting, due to high detail of urban-scale representation of boundary layer processes. Such high resolution NWP is computationally expensive to run - and especially so if AQ forecasting is incorporated - thus they could be reserved for use during busy events with high people-density, e.g. the Olympics in Paris, 2024.

Currently, it is likely that off-line dispersion models such as NAME require further development to be able to make the best use of convection-resolving NWP, in light of the UKV tracer results presented in chapter 5. There is also the bigger question of whether off-line models would even be capable of operating efficiently at such high resolutions, since they would also require a small model time-step to match. However, fog forecasts within the 333 m scale London Model (Boutle et al., 2016) are already running routinely, therefore it is only a matter of time until routine on-line air quality forecasts reach similar city-scale resolutions. It was therefore a useful exercise to test the suitability of neighbourhood verification methods on air quality variables in chapter 4 of this thesis, as it is clear that moving forward, traditional verification metrics will be inadequate to provide an accurate assessment of the skill of a high-resolution forecast. Further work needs to be done on testing optimal neighbourhood sizes for different variables, testing which probabilistic scores are most useful (Brier score and CRPS used in this thesis are two of many), and testing which threshold values are most appropriate for each variable. Furthermore, probabilistic neighbourhood verification methods could also be applied not only to routine forecasts, but also to the type of evaluation conducted in chapter 5, which would provide more information about the type of spatial errors within a semi-idealised setting.

In terms of the future of air quality forecasting itself, there is a lot of rewarding work to be done. "Clean growth and Innovation" is one of the Met Office's current Research and Innovation Strategies <sup>1</sup> which includes air pollution, as it is widely recognised that air pollution is an area of high importance to public health and meeting international emissions incentives. A recent high-profile case of a child's death due to successive asthma attacks and respiratory problems has led to an inquest into the levels of pollutant exposure in London, as well as the public messaging surrounding air quality, such as the DAQI. Ella Adoo Kissi-Debrah is the first person to officially have had air pollution named as

---

<sup>1</sup><https://www.metoffice.gov.uk/research/approach/research-and-innovation-strategy> - last accessed on 14-01-2021

a cause of death<sup>2</sup>. Evaluating population exposure to PM<sub>2.5</sub> in particular is therefore a priority for the Office of Health Promotion (successor to Public Health England after October 2021).

Accurate short-term pollution forecasts (daily – weekly) at urban scales are therefore necessary in the interest of public health. Pollution concentrations in urban areas are not spatially homogeneous, as some regions of a city will be more highly polluted than others due to local sources and building characteristics dictating dispersion flows and surface heat fluxes. Therefore, in an ideal world with sufficient computational availability, a city-scale forecast of high accuracy could be possible, and it could be possible by relying on accurate representations of meteorological processes at fine scales. There is on-going work to advance the resolution capabilities of AQ forecasting, e.g. Wu et al. (2014) use the Community Multi-scale Air Quality Modeling (CMAQ) to forecast PM<sub>10</sub> in Beijing at 3km resolution; Žabkar et al. (2015) use the Weather Research and Forecasting model with chemistry (WRF-Chem) to forecast O<sub>3</sub> at 3.7 km resolution in Slovenia, with more skill than a statistical model. WRF-Chem forecasts of O<sub>3</sub>, PM<sub>2.5</sub>, PM<sub>10</sub> and other species are also evaluated against observations taken during a campaign in Australia, within nested domains at 81, 21, 9 and 3 km (Zhang et al., 2019). WRF-Chem PM<sub>2.5</sub> forecasting capability is evaluated at even higher resolutions by Jena et al. (2021), who evaluate the performance of the first 400 m WRF-Chem configuration as is used in Delhi to alert citizens of high pollution levels at the street-scale. It is apparent that city-scale AQ models are well on their way to becoming used in routine forecasting within the next decade.

Various international pollution incentives (e.g the Paris Agreement) are resulting in cities all over the UK pledging to reduce street pollution levels e.g. by reducing emissions from road vehicles, waste management and domestic heating. Therefore accurate long-term pollution predictions (e.g. seasonal – annual) based on anthropogenic scenarios are necessary to help policymakers reach realistic targets.

Both short- and long-term pollution forecasts need to be verified against a reliable truth. The current network of Automatic Urban and Rural (AURN) sites provides reliable quality surface observations; however, the irregularity of these sites and extreme sparseness in some rural regions of the UK (e.g. Scotland, North Wales, the Lake District and Northumberland to name a few) mean that the air quality forecast in these regions cannot be verified reliably without the use of spatial interpolation or neighbourhood ver-

---

<sup>2</sup><https://www.gov.uk/government/news/government-responds-to-coroner-after-ella-kissi-debrah-inquest>

ification methods, both of which introduce another element of measurement uncertainty. Therefore, there is a need for a higher density network of quality-controlled air quality observation sites in order to verify air quality forecasts reliably, especially as forecast model resolution increases. Higher spatial density of observations could also pave a way for incorporating data assimilation into the forecasts - either on small, urban-scales or regional scales. Data assimilation can improve the accuracy of a forecast by regularly incorporating near-real-time observations, in order to correct for any fluctuations from the truth in the forecast. The air quality monitoring sites should also be located near to existing WMO meteorological sites (of which there exist hundreds throughout the UK), which could improve scientists' understanding of direct links between air pollution and meteorology on scales smaller than is currently possible.



## REFERENCES

- Adams, S. V., R. W. Ford, M. Hambley, J. M. Hobson, I. Kavčić, C. M. Maynard, T. Melvin, E. H. Müller, S. Mullerworth, A. R. Porter, M. Rezny, B. J. Shipway, and R. Wong, 2019: LFRic: Meeting the challenges of scalability and performance portability in Weather and Climate models. *Journal of Parallel and Distributed Computing*, **132**, 383–396, doi:10.1016/j.jpdc.2019.02.007. (cited on page: 41)
- Ambaum, M. H., 2020: *Thermal physics of the atmosphere*. Developments in Weather and Climate Science Ser., Elsevier, 2nd ed. edition. (cited on page: 23)
- Ansmann, A., J. Bösenberg, A. Chiakovsky, A. Comerón, S. Eckhardt, R. Eixmann, V. Freudenthaler, P. Ginoux, L. Komguem, H. Linné, M. Lopez Marquez, V. Matthias, I. Mattis, V. Mitev, D. Müller, S. Music, S. Nickovic, J. Pelon, L. Sauvage, P. Sobolewsky, M. K. Srivastava, A. Stohl, O. Torres, G. Vaughan, U. Wandinger, and M. Wiegner, 2003: Long-range transport of Saharan dust to northern Europe: The 11-16 October 2001 outbreak observed with EARLINET. *Journal of Geophysical Research: Atmospheres*, **108**, doi:10.1029/2003jd003757. (cited on page: 14)
- Arakawa, A. and V. R. Lamb, 1977: Computational design of the basic dynamical processes of the UCLA general circulation model. *General circulation models of the atmosphere*, **17**, 173–265. (cited on page: 41)
- Baklanov, A., K. Schlünzen, P. Suppan, J. Baldasano, D. Brunner, S. Aksoyoglu, G. Carmichael, J. Douros, J. Flemming, R. Forkel, S. Galmarini, M. Gauss, G. Grell, M. Hirtl, S. Joffre, O. Jorba, E. Kaas, M. Kaasik, G. Kallos, X. Kong, U. Korsholm, A. Kurganskiy, J. Kushta, U. Lohmann, A. Mahura, A. Manders-Groot, A. Maurizi, N. Moussiopoulos, S. T. Rao, N. Savage, C. Seigneur, R. S. Sokhi, E. Solazzo, S. Solomos, B. Sørensen, G. Tsegas, E. Vignati, B. Vogel, and Y. Zhang, 2014: Online coupled regional meteorology chemistry models in Europe: Current status and prospects. *Atmospheric Chemistry and Physics*, **14**, 317–398, doi:10.5194/acp-14-317-2014. (cited on pages: 2, 3)
- Bangert, M., C. Kottmeier, B. Vogel, and H. Vogel, 2011: Regional scale effects of the aerosol cloud interaction simulated with an online coupled comprehensive chemistry model. *Atmospheric Chemistry and Physics*, **11**, 4411–4423, doi:10.5194/acp-11-4411-2011. (cited on page: 2)
- Best, M. J., 2005: Representing urban areas within operational numerical weather prediction models. *Boundary-Layer Meteorology*, **114**, 91–109, doi:10.1007/s10546-004-4834-5. (cited on pages: 128, 173)
- Best, M. J., C. S. B. Grimmond, and M. G. Villani, 2006: Evaluation of the urban tile in MOSES using surface energy balance observations. *Boundary-Layer Meteorology*, **118**, 503–525, doi:10.1007/s10546-005-9025-5. (cited on page: 24)
- Best, M. J., M. Pryor, D. B. Clark, G. G. Rooney, R. L. H. Essery, C. B. Menard, J. M. Edwards, M. A. Hendry, A. Porson, N. Gedney, L. M. Mercado, S. Sitch, E. Blyth, O. Boucher, P. M. Cox, C. S. B. Grimmond, and R. J. Harding, 2011: The Joint UK Land Environment Simulator (JULES), model description – Part 1: Energy and water fluxes. *Geosci. Model Dev.*, **4**, 677–699, doi:10.5194/gmd-4-677-2011. (cited on page: 138)
- Biedenkapp, D., L. G. Hartshorn, and E. J. Bair, 1970: The O(1D) + H<sub>2</sub>O reaction. *Chemical Physics Letters*, **5**, 379–381, doi:10.1016/0009-2614(70)85172-7. (cited on page: 17)

- Biggs, E. M. and P. M. Atkinson, 2011: A comparison of gauge and radar precipitation data for simulating an extreme hydrological event in the Severn Uplands, UK. *Hydrological Processes*, **25**, 795–810, doi:10.1002/hyp.7869. (cited on page: 65)
- Blumenthal, D. L., W. H. White, and T. B. Smith, 1977: Anatomy of a Los Angeles Smog Episode: Pollutant Transport in the Daytime Sea Breeze Regime. *Atmospheric Environment*, **12**, 893 – 907. (cited on page: 93)
- Bohnenstengel, S. I., S. E. Belcher, A. Aiken, J. D. Allan, G. Allen, A. Bacak, T. J. Bannan, J. F. Barlow, D. C. Beddows, W. J. Bloss, A. M. Booth, C. Chemel, O. Coceal, C. F. Di Marco, M. K. Dubey, K. H. Faloon, Z. L. Fleming, M. Furger, J. K. Gietl, R. R. Graves, D. C. Green, C. S. Grimmond, C. H. Halios, Hamilton J. F., R. M. Harrison, M. R. Heal, D. E. Heard, C. Helfter, S. C. Herndon, R. E. Holmes, J. R. Hopkins, A. M. Jones, F. J. Kelly, S. Kotthaus, B. Langford, J. D. Lee, R. J. Leigh, A. C. Lewis, R. T. Lidster, F. D. Lopez-Hilfiker, J. B. McQuaid, C. Mohr, P. S. Monks, E. Nemitz, N. L. Ng, C. J. Percival, A. S. H. Prévôt, H. M. A. Ricketts, R. Sokhi, D. Stone, J. A. Thornton, A. H. Tremper, A. C. Valach, S. Visser, L. K. Whalley, L. R. Williams, L. Xu, D. E. Young, and P. Zotter, 2015: Meteorology, air quality, and health in London: The ClearLo project. *Bulletin of the American Meteorological Society*, **96**, 779–804, doi:10.1175/BAMS-D-12-00245.1. (cited on page: 7)
- Bohnenstengel, S. I., S. Evans, P. A. Clark, and S. Belcher, 2011: Simulations of the London urban heat island. *Quarterly Journal of the Royal Meteorological Society*, **137**, 1625–1640, doi:10.1002/qj.855. (cited on page: 25)
- Bohnenstengel, S. I. and M. A. Hendry, 2016: Report on implementation and evaluation of MORUSES in the UKV (PS37). Technical report, Met Office. (cited on pages: 25, 129, and 138)
- Boutle, I. A., A. Finnenkoetter, A. P. Lock, and H. Wells, 2016: The London Model: forecasting fog at 333 m resolution. *Quarterly Journal of the Royal Meteorological Society*, **142**, 360–371, doi:10.1002/qj.2656. (cited on page: 179)
- Bower, J. S., G. F. Broughton, M. T. Dando, K. J. Stevenson, J. E. Lampert, B. P. Sweeney, V. J. Parker, G. S. Driver, A. G. Clark, C. J. Waddon, A. J. Wood, and M. L. Williams, 1989: Surface ozone concentrations in the U.K. in 1987–1988. *Atmospheric Environment (1967)*, **23**, 2003–2016, doi:10.1016/0004-6981(89)90527-1. (cited on page: 18)
- Brasseur, G. P. and D. J. Jacob, 2017: *Modeling of Atmospheric Chemistry*. Cambridge University Press, Cambridge, 1 edition, 606 pp. (cited on pages: 15, 25, and 26)
- Brier, G. W., 1950: Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, **78**, 1–3, doi:10.1175/1520-0493(1950)078;0001:VOFEIT;2.0.CO;2. (cited on pages: 10, 83, and 86)
- Brönnimann, S. and U. Neu, 1997: Weekend-weekday differences of near-surface ozone concentrations in Switzerland for different meteorological conditions. *Atmospheric Environment*, **31**, 1127–1135, doi:10.1016/S1352-2310(96)00311-1. (cited on page: 19)
- Brown, A., S. Milton, M. Cullen, B. Golding, J. Mitchell, and A. Shelly, 2012a: Unified modeling and prediction of weather and climate: A 25-year journey. *Bulletin of the American Meteorological Society*, **93**, 1865–1877, doi:10.1175/BAMS-D-12-00018.1. (cited on page: 41)

- Brown, B. G., E. Gilleland, and E. E. Ebert, 2012b: Forecasts of spatial fields. *Forecast verification: A practitioner's guide in atmospheric science*, I. T. Joliffe and D. B. Stephenson, eds., John Wiley & Sons, volume 2, chapter 6, 2 edition, 95–118. (cited on page: 9)
- Brown, S. S. and J. Stutz, 2012: Nighttime radical observations and chemistry. *Chemical Society Reviews*, **41**, 6405, doi:10.1039/c2cs35181a. (cited on page: 118)
- Brown-Steiner, B., P. Hess, and M. Lin, 2015: On the capabilities and limitations of GCM simulations of summertime regional air quality: A diagnostic analysis of ozone and temperature simulations in the US using CESM CAM-Chem. *Atmospheric Environment*, **101**, 134–148, doi:10.1016/j.atmosenv.2014.11.001. (cited on page: 3)
- Brunner, D., N. Savage, O. Jorba, B. Eder, L. Giordano, A. Badia, A. Balzarini, R. Baró, R. Bianconi, C. Chemel, G. Curci, R. Forkel, P. Jiménez-Guerrero, M. Hirtl, A. Hodzic, L. Honzak, U. Im, C. Knote, P. Makar, A. Manders-Groot, E. van Meijgaard, L. Neal, J. L. Pérez, G. Pirovano, R. San Jose, W. Schröder, R. S. Sokhi, D. Syrakov, A. Torian, P. Tuccella, J. Werhahn, R. Wolke, K. Yahya, R. Zabkar, Y. Zhang, C. Hogrefe, and S. Galmarini, 2015: Comparative analysis of meteorological performance of coupled chemistry-meteorology models in the context of AQMEII phase 2. *Atmospheric Environment*, **115**, 470–498, doi:10.1016/J.ATMOSENV.2014.12.032. (cited on pages: 2, 3, 5, 34, 36, 54, 77, 79, 127, 130, and 170)
- Casati, B., G. Ross, and D. B. Stephenson, 2004: A new intensity-scale approach for the verification of spatial precipitation forecasts. *Meteorological Applications*, **11**, 141–154, doi:10.1017/S1350482704001239. (cited on pages: 10, 82)
- Chapman, S., 1930: On ozone and atomic oxygen in the upper atmosphere. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, **10**, 369–383, doi:10.1080/14786443009461588. (cited on page: 15)
- Ciach, G. J. and W. F. Krajewski, 1999: On the estimation of radar rainfall error variance. *Advances in Water Resources*, **22**, 585–595, doi:10.1016/S0309-1708(98)00043-8. (cited on page: 105)
- Clark, P., N. Roberts, H. Lean, S. P. Ballard, and C. Charlton-Perez, 2016: Convection-permitting models: a step-change in rainfall forecasting. *Meteorological Applications*, **23**, 165–181, doi:10.1002/met.1538. (cited on page: 125)
- Coates, J., K. A. Mar, N. Ojha, and T. M. Butler, 2016: The influence of temperature on ozone production under varying NO<sub>x</sub> conditions - A modelling study. *Atmospheric Chemistry and Physics*, doi:10.5194/acp-16-11601-2016. (cited on page: 32)
- Cocks, A., 1993: *The Chemistry and Deposition of Nitrogen Species in the Troposphere*. The Royal Society of Chemistry, Cambridge, 133 pp. (cited on pages: 15, 18)
- COMEAP, 1998: Quantification of the effects of air pollution on health in the united kingdom: a report by the committee on the medical effects of air pollutants. Technical report. (cited on page: 14)
- 2011: Review of the UK Air Quality Index: A report by the Committee on the Medical Effects of Air Pollutants. Technical report, Health Protection Agency. (cited on pages: 8, 38)
- 2015: COMEAP: quantification of mortality and hospital admissions associated with ground-level ozone - GOV.UK. Technical report, Public Health England, London. (cited on page: 1)

- Connolly, E., G. Fuller, T. Baker, and P. Willis, 2013: Update on implementation of the daily air quality index. Available at [https://uk-air.defra.gov.uk/library/reports?report\\_id=750](https://uk-air.defra.gov.uk/library/reports?report_id=750), date last accessed: 2021-02-16. (cited on page: 8)
- Cox, R. A., A. E. Eggleton, R. G. Derwent, J. E. Lovelock, and D. H. Pack, 1975: Long-range transport of photochemical ozone in north-western Europe. *Nature*, **255**, 118–121, doi:10.1038/255118a0. (cited on page: 28)
- Crocker, R., J. Maksymczuk, M. Mittermaier, M. Tonani, and C. Pequignat, 2020: An approach to the verification of high-resolution ocean models using spatial methods. *Ocean Science*, **16**, 831–845, doi:10.5194/os-16-831-2020. (cited on pages: 83, 170)
- Crutzen, P. J., 1971: Ozone production rates in an oxygen-hydrogen-nitrogen oxide atmosphere. *Journal of Geophysical Research*, **76**, 7311–7327, doi:10.1029/jc076i030p07311. (cited on pages: 14, 15)
- Dabberdt, W. F., M. A. Carroll, D. Baumgardner, G. Carmichael, R. Cohen, T. Dye, J. Ellis, G. Grell, S. Grimmond, S. Hanna, J. Irwin, B. Lamb, S. Madronich, J. McQueen, J. Meagher, T. Odman, J. Pleim, H. P. Schmid, and D. L. Westphal, 2004: Meteorological research needs for improved air quality forecasting. *Bulletin of the American Meteorological Society*, **85**, 563–586, doi:10.1175/BAMS-85-4-563. (cited on page: 31)
- Dacre, H. F., 2011: A new method for evaluating regional air quality forecasts. *Atmospheric Environment*, **45**, 993–1002, doi:10.1016/j.atmosenv.2010.10.048. (cited on page: 82)
- Dacre, H. F., P. Bedwell, D. Hertwig, S. J. Leadbetter, P. Loizou, and H. N. Webster, 2020: Improved representation of particle size and solubility in model simulations of atmospheric dispersion and wet-deposition from Fukushima. *Journal of Environmental Radioactivity*, **217**, 106193, doi:10.1016/j.jenvrad.2020.106193. (cited on page: 75)
- Dacre, H. F., A. L. M. Grant, R. J. Hogan, S. E. Belcher, D. J. Thomson, B. J. Devenish, F. Marengo, M. C. Hort, J. M. Haywood, A. Ansmann, I. Mattis, and L. Clarisse, 2011: Evaluating the structure and magnitude of the ash plume during the initial phase of the 2010 Eyjafjallajökull eruption using lidar observations and NAME simulations. *Journal of Geophysical Research*, **116**, D00U03, doi:10.1029/2011JD015608. (cited on page: 14)
- Dai, A., K. E. Trenberth, and T. R. Karl, 1999: Effects of clouds, soil moisture, precipitation, and water vapor on diurnal temperature range. *Journal of Climate*, **12**, 2451–2473, doi:10.1175/1520-0442(1999)012<2451:eocsmpj2.0.co;2. (cited on page: 31)
- Damrath, U., 2004: Verification against precipitation observations of a high density network – what did we learn? Montreal, International Verification Methods Workshop, available at [https://www.cawcr.gov.au/projects/verification/Workshop2004/presentations/5.3\\_Damrath.pdf](https://www.cawcr.gov.au/projects/verification/Workshop2004/presentations/5.3_Damrath.pdf). (cited on page: 82)
- Davies, T., 2014: Lateral boundary conditions for limited area models. *Quarterly Journal of the Royal Meteorological Society*, doi:10.1002/qj.2127. (cited on page: 5)
- Davis, C. A. and F. Carr, 2000: Summary of the 1998 workshop on mesoscale model verification. *Bulletin of the American Meteorological Society*, **81**, 809–819, doi:10.1175/1520-0477(2000)081<0809:SOTWOMj2.3.CO;2. (cited on page: 81)
- de Arellano, J. V.-G., B. Gioli, F. Miglietta, H. J. J. Jonker, H. K. Baltink, R. W. A. Hutjes, and A. A. M. Holtslag, 2004: Entrainment process of carbon dioxide in the atmospheric boundary layer. *Journal of Geophysical Research*, **109**, D18110, doi:10.1029/2004JD004725. (cited on page: 118)

- de Leeuw, J., J. Methven, and M. Blackburn, 2015: Evaluation of ERA-Interim reanalysis precipitation products using England and Wales observations. *Quarterly Journal of the Royal Meteorological Society*, **141**, 798–806, doi:10.1002/qj.2395. (cited on page: 61)
- DEFRA, 2019: Air Pollution in the UK (2018). Technical report, DEFRA, available at [www.nationalarchives.gov.uk/doc/open-government-licence/version/3/](http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/), date last accessed: 2021-02-16. (cited on page: 38)
- Dentener, F., R. Derwent, E. Dlugokencky, E. Holland, I. Isaksen, J. Katima, V. Kirchhoff, P. Matson, P. Midgley, and M. Wang, 2001: Atmospheric Chemistry and Greenhouse Gases. *International Panel on Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge University Press, Cambridge. (cited on page: 14)
- Derwent, R., P. Simmonds, S. Seuring, and C. Dimmer, 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. *Atmospheric Environment*, **32**, 145–157, doi:10.1016/S1352-2310(97)00338-5. (cited on page: 28)
- Dodson, D. S. and J. D. Small Griswold, 2021: Turbulent and boundary layer characteristics during VOCALS-REx. *Atmospheric Chemistry and Physics*, **21**, 1937–1961, doi:10.5194/acp-21-1937-2021. (cited on page: 137)
- Eaton, S., 2016: The Environment Agency: QAQC Report for the Automatic Urban and Rural Network, January-March 2016. Technical report, Ricardo Energy & Environment, Didcot. (cited on pages: 38, 39)
- Ebert, E. E., 2008: Fuzzy verification of high-resolution gridded forecasts: a review and proposed framework. *Meteorological Applications*, **15**, 51–64, doi:10.1002/met.25. (cited on pages: 9, 82, and 169)
- 2009: Neighborhood Verification: A Strategy for Rewarding Close Forecasts. *Weather and Forecasting*, **24**, 1498–1510, doi:10.1175/2009WAF2222251.1. (cited on pages: 83, 169, and 176)
- Ebert, E. E. and J. L. McBride, 2000: Verification of precipitation in weather systems: Determination of systematic errors. *Journal of Hydrology*, doi:10.1016/S0022-1694(00)00343-7. (cited on page: 81)
- Ebert, E. E., L. Wilson, A. Weigel, M. Mittermaier, P. Nurmi, P. Gill, M. Göber, S. Joslyn, B. Brown, T. Fowler, and A. Watkins, 2013: Progress and challenges in forecast verification. *Meteorological Applications*, **20**, 130–139, doi:10.1002/met.1392. (cited on page: 35)
- Eder, B., D. Kang, R. Mathur, S. Yu, and K. Schere, 2006: An operational evaluation of the Eta-CMAQ air quality forecast model. *Atmospheric Environment*, **40**, 4894–4905, doi:10.1016/j.atmosenv.2005.12.062. (cited on pages: 3, 4, and 120)
- Essery, R. L. H., M. J. Best, R. A. Betts, P. M. Cox, and C. M. Taylor, 2003: Explicit Representation of Subgrid Heterogeneity in a GCM Land Surface Scheme. *Journal of Hydrometeorology*, **4**, 530–543, doi:10.1175/1525-7541(2003)004<0530:EROSHI>2.0.CO;2. (cited on page: 128)
- Ferro, C. A. T., D. S. Richardson, and A. P. Weigel, 2008: On the effect of ensemble size on the discrete and continuous ranked probability scores. *Meteorological Applications*, **15**, 19–24, doi:10.1002/met.45. (cited on page: 85)

- Foret, G., M. Eremenko, J. Cuesta, P. Sellitto, J. Barré, B. Gaubert, A. Coman, G. Dufour, X. Liu, M. Joly, C. Doche, and M. Beekmann, 2014: Ozone pollution: What can we see from space? A case study. *Journal of Geophysical Research: Atmospheres*, **119**, 8476–8499, doi:10.1002/2013JD021340. (cited on page: 6)
- Gandin, L. S. and A. H. Murphy, 1992: Equitable skill scores for categorical forecasts. *Monthly Weather Review*, **120**, 361–370, doi:10.1175/1520-0493(1992)120<0361:ESSFCF>2.0.CO;2. (cited on page: 9)
- Garratt, J. R., 1992: *The Atmospheric Boundary Layer*. Cambridge University Press, 316 pp. (cited on page: 20)
- Gilbert, G. K., 1884: Finley’s tornado predictions. *Amer. Met. J.* (cited on page: 9)
- Gilleland, E., D. Ahijevych, B. G. Brown, B. Casati, and E. E. Ebert, 2009: Intercomparison of Spatial Forecast Verification Methods. *Weather and Forecasting*, **24**, 1416–1430, doi:10.1175/2009WAF2222269.1. (cited on pages: 34, 35, and 81)
- Golding, B. W., 1998: Nimrod: A system for generating automated very short range forecasts. *Meteorological Applications*, **5**, S1350482798000577, doi:10.1017/S1350482798000577. (cited on page: 39)
- Gong, W., A. P. Dastoor, V. S. Bouchet, S. Gong, P. A. Makar, M. D. Moran, B. Pabla, S. Ménard, L.-P. Crevier, S. Cousineau, and S. Venkatesh, 2006: Cloud processing of gases and aerosols in a regional air quality model (AURAMS). *Atmospheric Research*, **82**, 248–275, doi:10.1016/j.atmosres.2005.10.012. (cited on page: 79)
- Gray, S. L., 2003: A case study of stratosphere to troposphere transport: The role of convective transport and the sensitivity to model resolution. *Journal of Geophysical Research*, **108**, 4590, doi:10.1029/2002JD003317. (cited on page: 14)
- Gregory, D. and P. R. Rowntree, 1990: A Mass Flux Convection Scheme with Representation of Cloud Ensemble Characteristics and Stability-Dependent Closure. *Monthly Weather Review*, **118**, 1483–1506, doi:10.1175/1520-0493(1990)118<1483:AMFCSW>2.0.CO;2. (cited on page: 126)
- Grell, G. A., S. E. Peckham, R. Schmitz, S. A. McKeen, G. Frost, W. C. Skamarock, and B. Eder, 2005: Fully coupled “online” chemistry within the WRF model. *Atmospheric Environment*, **39**, 6957–6975, doi:10.1016/j.atmosenv.2005.04.027. (cited on page: 152)
- Guenther, A. B., P. R. Zimmerman, P. C. Harley, R. K. Monson, and R. Fall, 1993: Isoprene and monoterpene emission rate variability: Model evaluations and sensitivity analyses. *Journal of Geophysical Research*, **98**, 12609, doi:10.1029/93JD00527. (cited on page: 31)
- Haiden, T., M. J. Rodwell, D. S. Richardson, A. Okagaki, T. Robinson, and T. Hewson, 2012: Intercomparison of global model precipitation forecast skill in 2010/11 using the SEEPS score. *Monthly Weather Review*, **140**, 2720–2733, doi:10.1175/MWR-D-11-00301.1. (cited on page: 35)
- Hanna, S. R. and R. Yang, 2001: Evaluations of mesoscale models’ simulations of near-surface winds, temperature gradients, and mixing depths. *Journal of Applied Meteorology*, **40**, 1095–1104, doi:10.1175/1520-0450(2001)040<1095:EOMMSO>2.0.CO;2. (cited on pages: 4, 173)

- Hardacre, C., O. Wild, and L. Emberson, 2015: An evaluation of ozone dry deposition in global scale chemistry climate models. *Atmos. Chem. Phys. Discuss.*, **15**, 6419–6436, doi:10.5194/acp-15-6419-2015. (cited on pages: [29](#), [148](#))
- Harman, I. N. and S. E. Belcher, 2006: The surface energy balance and boundary layer over urban street canyons. *Quarterly Journal of the Royal Meteorological Society*, **132**, 2749–2768, doi:10.1256/qj.05.185. (cited on page: [129](#))
- Harrison, D. L., J. Driscoll, and M. Kitchen, 2000: Improving precipitation estimates from weather radar using quality control and correction techniques. **6**, 135–144. (cited on pages: [39](#), [40](#))
- Harrison, R. M. and H. M. McCartney, 1979: Ambient Air Quality at a Coastal Site in Rural North West England. *Atmospheric Environment*, **14**, 233 – 244. (cited on page: [93](#))
- Harvey, N. J., R. J. Hogan, and H. F. Dacre, 2013: A method to diagnose boundary-layer type using doppler lidar. *Quarterly Journal of the Royal Meteorological Society*, doi:10.1002/qj.2068. (cited on pages: [20](#), [46](#), [47](#), and [51](#))
- Hersbach, H., 2000: Decomposition of the Continuous Ranked Probability Score for Ensemble Prediction Systems. *Weather and Forecasting*, **15**, 559–570, doi:10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2. (cited on pages: [10](#), [88](#))
- Hess, G. D., K. J. Tory, M. E. Cope, S. Lee, K. Puri, P. C. Manins, and M. Young, 2004: The Australian air quality forecasting system. Part II: Case study of a Sydney 7-day photochemical smog event. *Journal of Applied Meteorology*, **43**, 663–679, doi:10.1175/2094.1. (cited on pages: [3](#), [5](#))
- Hobbs, P. V., 2000: *Basic Physical Chemistry for the Atmospheric Sciences*. Cambridge University Press, 1st edition. (cited on page: [15](#))
- Hoffman, R. N., Z. Liu, J.-F. Louis, and C. Grassoti, 1995: Distortion Representation of Forecast Errors. *Monthly Weather Review*, **123**, 2758–2770, doi:10.1175/1520-0493(1995)123<2758:drofe>2.0.co;2. (cited on page: [81](#))
- Hogan, R. J., C. A. Ferro, I. T. Jolliffe, and D. B. Stephenson, 2010: Equitability revisited: Why the “Equitable Threat Score” is not equitable. *Weather and Forecasting*, **25**, 710–726, doi:10.1175/2009WAF2222350.1. (cited on page: [9](#))
- Hogan, R. J. and I. B. Mason, 2012: Deterministic forecasts of binary events. *Forecast Verification: A Practitioner’s Guide in Atmospheric Science*, I. Jolliffe and D. Stephenson, eds., John Wiley & Sons, chapter 3, 2nd edition. (cited on page: [9](#))
- Hogrefe, C., W. Hao, K. Civerolo, J. Y. Ku, G. Sistla, R. S. Gaza, L. Sedefian, K. Schere, A. Gilliland, and R. Mathur, 2007: Daily simulation of ozone and fine particulates over New York State: Findings and challenges. *Journal of Applied Meteorology and Climatology*, **46**, 961–979, doi:10.1175/JAM2520.1. (cited on page: [3](#))
- Honoré, C., L. Rouïl, R. Vautard, M. Beekmann, B. Bessagnet, A. Dufour, C. Elichegaray, J.-M. Flaud, L. Malherbe, F. Meleux, L. Menut, D. Martin, A. Peuch, V.-H. Peuch, and N. Poisson, 2008: Predictability of European air quality: Assessment of 3 years of operational forecasts and analyses by the PREV’AIR system. *Journal of Geophysical Research*, **113**, D04301, doi:10.1029/2007JD008761. (cited on pages: [2](#), [5](#))

- House of Commons Environmental Audit Committee (HCEA), 2010: Air Quality: 5th report of session 2009 - 10, Vol 1. Technical report, available at [www.publications.parliament.uk/pa/cm200910/cmselect/cmenvaud/229/229i.pdf](http://www.publications.parliament.uk/pa/cm200910/cmselect/cmenvaud/229/229i.pdf), date last accessed: 2021-02-16. (cited on page: 1)
- Ilyinskaya, E., S. Mobbs, R. Burton, M. Burton, F. Pardini, M. A. Pfeffer, R. Purvis, J. Lee, S. Bauguitte, B. Brooks, I. Colfescu, G. N. Petersen, A. Wellpott, and B. Bergsson, 2018: Globally Significant CO<sub>2</sub> Emissions From Katla, a Subglacial Volcano in Iceland. *Geophysical Research Letters*, **45**, 332–10, doi:10.1029/2018GL079096. (cited on page: 7)
- Im, U., R. Bianconi, E. Solazzo, I. Kioutsioukis, A. Badia, A. Balzarini, R. Baró, R. Bellasio, D. Brunner, C. Chemel, G. Curci, H. Denier van der Gon, J. Flemming, R. Forkel, L. Giordano, P. Jiménez-Guerrero, M. Hirtl, A. Hodzic, L. Honzak, O. Jorba, C. Knote, P. A. Makar, A. Manders-Groot, L. Neal, J. L. Pérez, G. Pirovano, G. Pouliot, R. San Jose, N. Savage, W. Schroder, R. S. Sokhi, D. Syrakov, A. Torian, P. Tuccella, K. Wang, J. Werhahn, R. Wolke, R. Zabkar, Y. Zhang, J. Zhang, C. Hogrefe, and S. Galmarini, 2015a: Evaluation of operational online-coupled regional air quality models over Europe and North America in the context of AQMEII phase 2. Part II: Particulate matter. *Atmospheric Environment*, **115**, 421–441, doi:10.1016/j.atmosenv.2014.08.072. (cited on page: 10)
- Im, U., R. Bianconi, E. Solazzo, I. Kioutsioukis, A. Badia, A. Balzarini, R. Baró, R. Bellasio, D. Brunner, C. Chemel, G. Curci, J. Flemming, R. Forkel, L. Giordano, P. Jiménez-Guerrero, M. Hirtl, A. Hodzic, L. Honzak, O. Jorba, C. Knote, J. J. Kuenen, P. A. Makar, A. Manders-Groot, L. Neal, J. L. Pérez, G. Pirovano, G. Pouliot, R. San Jose, N. Savage, W. Schroder, R. S. Sokhi, D. Syrakov, A. Torian, P. Tuccella, J. Werhahn, R. Wolke, K. Yahya, R. Zabkar, Y. Zhang, J. Zhang, C. Hogrefe, and S. Galmarini, 2015b: Evaluation of operational on-line-coupled regional air quality models over Europe and North America in the context of AQMEII phase 2. Part I: Ozone. *Atmospheric Environment*, doi:10.1016/j.atmosenv.2014.09.042. (cited on pages: 3, 4, 10, and 170)
- Jaffe, D., 2011: Relationship between Surface and Free Tropospheric Ozone in the Western U.S. *Environmental Science & Technology*, **45**, 432–438, doi:10.1021/es1028102. (cited on page: 118)
- Jena, C., S. D. Ghude, R. Kumar, S. Debnath, G. Govardhan, V. K. Soni, S. H. Kulkarni, G. Beig, R. S. Nanjundiah, and M. Rajeevan, 2021: Performance of high resolution (400 m) PM<sub>2.5</sub> forecast over Delhi. *Scientific Reports*, **11**, 4104, doi:10.1038/s41598-021-83467-8. (cited on page: 180)
- Jenkin, M. E., T. J. Davies, and J. R. Stedman, 2002: The origin and day-of-week dependence of photochemical ozone episodes in the UK. *Atmospheric Environment*, **36**, 999–1012, doi:10.1016/S1352-2310(01)00360-0. (cited on pages: 19, 28)
- Jenkin, M. E. and G. D. Hayman, 1999: Photochemical ozone creation potentials for oxygenated volatile organic compounds: sensitivity to variations in kinetic and mechanistic parameters. *Atmospheric Environment*, **33**, 1275–1293, doi:10.1016/S1352-2310(98)00261-1. (cited on page: 17)
- Joliffe, I. T. and D. B. Stephenson, 2012: *Forecast Verification: A Practitioner's Guide in Atmospheric Science*. Wiley - Blackwell, 2nd edition. (cited on page: 33)
- Jones, A., 2017: NAME Technical Specification Document A01: User Guide for NAME v7.2. Technical report, Met Office. (cited on page: 123)

- Jones, A., C. E. Johnson, N. Bellouin, P. S. Davison, J. G. L. Rae, L. Roberts, D., M. J. Woodage, S. Woodward, C. Ordonez, N. H. Savage, and M. E. Brooks, 2019a: Unified Model Documentation Paper 020: CLASSIC Aerosol Scheme. Technical report, Met Office. (cited on pages: 30, 51)
- Jones, A., D. Thomson, M. Hort, and B. Devenish, 2007: The U.K. Met Office's Next-Generation Atmospheric Dispersion Model, NAME III. *Air Pollution Modeling and Its Application XVII*, Springer US, Boston, MA, 580–589. (cited on pages: 120, 122, and 173)
- Jones, A. R., 2015: Using NWP data in NAME - a practical guide (NAME III Document version 4). Technical report, UK Met Office. (cited on page: 140)
- Jones, L., R. Claxton, G. Thistlewaite, J. Raoult, J. Maccarthy, E. Karagianni, E. Kilroy, T. Misselbrook, A. Otto, N. Passant, B. Pearson, J. Richardson, D. Wakeling, and C. Walker, 2019b: Air Quality Pollutant Inventories for England, Scotland, Wales and Northern Ireland: 1990-2017. Technical report, NAEI, available at [https://uk-air.defra.gov.uk/assets/documents/reports/cat09/2004151035\\_DA\\_Air\\_Pollutant\\_Inventories\\_1990-2017\\_Issue\\_1.2.pdf](https://uk-air.defra.gov.uk/assets/documents/reports/cat09/2004151035_DA_Air_Pollutant_Inventories_1990-2017_Issue_1.2.pdf), date last accessed: 2020-02-16. (cited on page: 43)
- Joss, J. and A. Waldvogel, 1990: Precipitation measurements in hydrology. *Radar in Meteorology: Battan Memorial and 40th Anniversary Radar Meteorology Conference*, D. Atlas, ed., American Meteorological society, 577–606. (cited on page: 56)
- Kar, J., J. Fishman, J. K. Creilson, A. Richter, J. Ziemke, and S. Chandra, 2010: Are there urban signatures in the tropospheric ozone column products derived from satellite measurements? *Atmos. Chem. Phys.*, **10**, 5213–5222, doi:10.5194/acp-10-5213-2010. (cited on page: 6)
- Karnosky, D. F., J. M. Skelly, K. E. Percy, and A. H. Chappelka, 2007: Perspectives regarding 50 years of research on effects of tropospheric ozone air pollution on US forests. *Environmental Pollution*, **147**, 489–506, doi:10.1016/j.envpol.2006.08.043. (cited on page: 14)
- Kaser, L., E. G. Patton, G. G. Pfister, A. J. Weinheimer, D. D. Montzka, F. Flocke, A. M. Thompson, R. M. Stauffer, and H. S. Halliday, 2017: The effect of entrainment through atmospheric boundary layer growth on observed and modeled surface ozone in the Colorado Front Range. *Journal of Geophysical Research: Atmospheres*, **122**, 6075–6093, doi:10.1002/2016JD026245. (cited on pages: 119, 152, 154, 158, 164, and 174)
- Kavassalis, S. C. and J. G. Murphy, 2017: Understanding ozone-meteorology correlations: A role for dry deposition. *Geophysical Research Letters*, **44**, 2922–2931, doi:10.1002/2016GL071791. (cited on page: 31)
- King, R., 2015: Comparison of UKV with MORUSES and JULES in Urban Areas. Technical report, Met Office. (cited on page: 138)
- Lean, H. W., J. F. Barlow, and C. H. Halios, 2019: The impact of spin-up and resolution on the representation of a clear convective boundary layer over London in order 100m grid-length versions of the Met Office Unified Model. *Quarterly Journal of the Royal Meteorological Society*, **145**, 1674–1689, doi:10.1002/qj.3519. (cited on page: 179)
- Lee, J. D., W. S. Drysdale, D. P. Finch, S. E. Wilde, and P. I. Palmer, 2020: UK surface NO<sub>2</sub> levels dropped by 42% during the COVID-19 lockdown: Impact on surface O<sub>3</sub>. *Atmospheric Chemistry and Physics*, **20**, 15743–15759, doi:10.5194/acp-20-15743-2020. (cited on pages: 47, 50)

- Lee, J. D., C. Helfter, R. M. Purvis, S. D. Beevers, D. C. Carslaw, A. C. Lewis, S. J. Møller, A. Tremper, A. Vaughan, and E. G. Nemitz, 2015: Measurement of NO<sub>x</sub> fluxes from a tall tower in central London, UK and comparison with emissions inventories. *Environmental Science and Technology*, **49**, 1025–1034, doi:10.1021/es5049072. (cited on page: 48)
- Lee, P., F. Ngan, H. Lei, B. Baker, B. Dornblaser, G. McGauhey, and D. Tong, 2014: An Application for Improving Air Quality (a Houston Case Study). *Earthzine (IEEE)*. (cited on page: 5)
- Lee, S.-M., H. J. Fernando, M. Princevac, D. Zajic, M. Sinesi, J. L. McCulley, and J. Anderson, 2003: Transport and Diffusion of Ozone in the Nocturnal and Morning Planetary Boundary Layer of the Phoenix Valley. *Environmental Fluid Mechanics*, **3**, 331–362, doi:10.1023/A:1023680216173. (cited on page: 14)
- Leighton, P. A., 1961: *Photochemistry of Air Pollution*. Academic Press, 300 pp. (cited on page: 16)
- Lelieveld, J. and P. J. Crutzen, 1990: Influences of cloud photochemical processes on tropospheric ozone Gas-phase chemistry. Technical report. (cited on page: 31)
- Lelieveld, J. and F. J. Dentener, 2000: What controls tropospheric ozone? *Journal of Geophysical Research: Atmospheres*, **105**, 3531–3551, doi:10.1029/1999JD901011. (cited on pages: 14, 19, and 28)
- Lenschow, D. H., R. Pearson, and B. B. Stankov, 1981: Estimating the ozone budget in the boundary layer by use of aircraft measurements of ozone eddy flux and mean concentration. *Journal of Geophysical Research*, **86**, 7291, doi:10.1029/JC086iC08p07291. (cited on page: 153)
- Lock, A. P., A. R. Brown, M. R. Bush, G. M. Martin, and R. N. B. Smith, 2000: A New Boundary Layer Mixing Scheme. Part I: Scheme Description and Single-Column Model Tests. *Monthly Weather Review*, **128**, 3187–3199, doi:10.1175/1520-0493(2000)128;3187:ANBLMS;2.0.CO;2. (cited on pages: 47, 76, and 125)
- Lock, A. P., J. Edwards, and I. Boutle, 2017: The Parametrization of Boundary Layer Processes. *Unified Model Documentation Paper 24*. (cited on pages: 24, 126, and 127)
- Logan, J. A., M. J. Prather, S. C. Wofsy, and M. B. McElroy, 1981: Tropospheric chemistry: A global perspective. *Journal of Geophysical Research*, **86**, 7210, doi:10.1029/JC086iC08p07210. (cited on page: 118)
- Lorenz, E. N., 1969: Atmospheric Predictability as Revealed by Naturally Occurring Analogues. *Journal of the Atmospheric Sciences*, **26**, 636–646, doi:10.1175/1520-0469(1969)26;636:APARBN;2.0.CO;2. (cited on page: 83)
- Lyons, W. A. and L. E. Olson, 1973: Detailed Mesometeorological Studies of Air Pollution Dispersion in the Chicago Lake Breeze. *Monthly Weather Review*, **101**. (cited on page: 93)
- Maksymczuk, J., R. Crocker, M. Mittermaier, and C. Pequignet, 2020: High-resolution model Verification Evaluation (HiVE). Part 1: Using neighbourhood techniques for the assessment of ocean model forecast skill. *EGU General Assembly 2020, Online*, number EGU2020-8681. (cited on page: 170)

- Marécal, V., V.-H. Peuch, C. Andersson, S. Andersson, J. Arteta, M. Beekmann, A. Benedictow, R. Bergström, B. Bessagnet, A. Cansado, F. Chéroux, A. Colette, A. Coman, R. L. Curier, H. A. C. Denier van der Gon, A. Drouin, H. Elbern, E. Emili, R. J. Engelen, H. J. Eskes, G. Foret, E. Friese, M. Gauss, C. Giannaros, J. Guth, M. Joly, E. Jaumouillé, B. Josse, N. Kadygrov, J. W. Kaiser, K. Krajsek, J. Kuenen, U. Kumar, N. Liora, E. Lopez, L. Malherbe, I. Martinez, D. Melas, F. Meleux, L. Menut, P. Moinat, T. Morales, J. Parmentier, A. Piacentini, M. Plu, A. Poupkou, S. Queguiner, L. Robertson, L. Rouïl, M. Schaap, A. Segers, M. Sofiev, L. Tarasson, M. Thomas, R. Timmermans, Valdebenito, P. van Velthoven, R. van Versendaal, J. Vira, and A. Ung, 2015: A regional air quality forecasting system over Europe: the MACC-II daily ensemble production. *Geoscientific Model Development*, **8**, 2777–2813, doi:10.5194/gmd-8-2777-2015. (cited on page: 51)
- Marzban, C. and S. Sandgathe, 2006: Cluster analysis for verification of precipitation fields. *Weather and Forecasting*, **21**, 824–838, doi:10.1175/WAF948.1. (cited on page: 81)
- Matheson, J. E. and R. L. Winkler, 1976: Scoring Rules for Continuous Probability Distributions. *Management Science*, **22**, 1087–1096, doi:10.1287/mnsc.22.10.1087. (cited on page: 84)
- Mauritsen, T. and G. Svensson, 2007: Observations of Stably Stratified Shear-Driven Atmospheric Turbulence at Low and High Richardson Numbers. *Journal of the Atmospheric Sciences*, **64**, 645–655, doi:10.1175/JAS3856.1. (cited on page: 127)
- Mauzerall, D. L., B. Sultan, N. Kim, and D. F. Bradford, 2005: NO<sub>x</sub> emissions from large point sources: Variability in ozone production, resulting health damages and economic costs. *Atmospheric Environment*, **39**, 2851–2866, doi:10.1016/j.atmosenv.2004.12.041. (cited on page: 18)
- Mavrakou, T., K. Philippopoulos, and D. Deligiorgi, 2012: The impact of sea breeze under different synoptic patterns on air pollution within Athens basin. *Science of the Total Environment*, **433**, 31–43, doi:10.1016/j.scitotenv.2012.06.011. (cited on page: 93)
- McNair, L. A., R. A. Harley, and A. G. Russell, 1996: Spatial inhomogeneity in pollutant concentrations, and their implications for air quality model evaluation. *Atmospheric Environment*, **30**, 4291–4301, doi:10.1016/1352-2310(96)00098-2. (cited on page: 35)
- McNider, R. T. and A. Pour-Biazar, 2020: Meteorological modeling relevant to mesoscale and regional air quality applications: a review. *Journal of the Air and Waste Management Association*, **70**, 2–43, doi:10.1080/10962247.2019.1694602. (cited on pages: 4, 14, and 53)
- Met Office, 2003: Met Office Rain Radar Data from the NIMROD System. NCAS British Atmospheric Data Centre. <https://catalogue.ceda.ac.uk/uuid/82adec1f896af6169112d09cc1174499>, date last accessed: 2020-10-14. (cited on page: 39)
- 2012: Met Office Integrated Data Archive System (MIDAS) Land and Marine Surface Stations Data (1853-current). NCAS British Atmospheric Data Centre. <http://catalogue.ceda.ac.uk/uuid/220a65615218d5c9cc9e4785a3234bd0>, date last accessed: 2020-09-25. (cited on page: 39)
- 2020a: Climate Summaries. Available at <https://www.metoffice.gov.uk/climate/uk/summaries/2017/june>, date last accessed 2021-02-16. (cited on page: 131)

- 2020b: MIDAS Data User Guide for UK Land Observations. Available at <http://cedadocs.ceda.ac.uk/id/eprint/1488>, date last accessed: 2021-02-16. (cited on page: 56)
- Milczewska, K. and H. F. Dacre, 2021: PhD Thesis: Identifying Meteorological Inaccuracies in Air Quality Forecasts - primary data. <http://dx.doi.org/10.17864/1947.325>. (cited on page: 122)
- Mittermaier, M. P., 2008: Introducing uncertainty of radar-rainfall estimates to the verification of mesoscale model precipitation forecasts. *Natural Hazards and Earth System Sciences*, **8**, 445–460. (cited on page: 57)
- 2014: A Strategy for Verifying Near-Convection-Resolving Model Forecasts at Observing Sites. *Weather and Forecasting*, **29**, 185–204, doi:10.1175/WAF-D-12-00075.1. (cited on pages: 82, 84, 85, 169, and 178)
- Mittermaier, M. P. and G. Csima, 2017: Ensemble versus Deterministic Performance at the Kilometer Scale. *Weather and Forecasting*, **32**, 1697–1709, doi:10.1175/WAF-D-16-0164.1. (cited on pages: 40, 83, and 88)
- Monks, P. S., 2000: A review of the observations and origins of the spring ozone maximum. *Atmospheric Environment*, **34**, 3545–3561, doi:10.1016/S1352-2310(00)00129-1. (cited on page: 19)
- Mukammal, E. I., H. H. Neumann, and T. J. Gillespie, 1982: Meteorological conditions associated with ozone in southwestern Ontario, Canada. *Atmospheric Environment (1967)*, **16**, 2095–2106, doi:10.1016/0004-6981(82)90281-5. (cited on page: 119)
- Murphy, A. H., 1973: A New Vector Partition of the Probability Score. *Journal of Applied Meteorology*, **12**, 595–600, doi:10.1175/1520-0450(1973)012<0595:ANVPOT>2.0.CO;2. (cited on page: 86)
- Neal, L. S., P. Agnew, S. Moseley, C. Ordóñez, N. H. Savage, and M. Tilbee, 2014: Application of a statistical post-processing technique to a gridded, operational, air quality forecast. *Atmospheric Environment*, **98**, 385–393, doi:10.1016/j.atmosenv.2014.09.004. (cited on pages: 35, 45)
- Neal, L. S., M. Dalvi, G. Folberth, R. N. McInnes, P. Agnew, F. M. O ’connor, N. H. Savage, and M. Tilbee, 2017: A description and evaluation of an air quality model nested within global and regional composition-climate models using MetUM. *Geosci. Model Dev*, **10**, 3941–3962, doi:10.5194/gmd-10-3941-2017. (cited on page: 43)
- Newman, K., T. Jensen, B. Brown, R. Bullock, and T. Fowler, 2019: The Model Evaluation Tools v8.1.2 User’s Guide. Technical report, Developmental Testbed Center, available at [https://dtcenter.org/sites/default/files/community-code/met/docs/user-guide/MET\\_Users\\_Guide\\_v8.1.2.pdf](https://dtcenter.org/sites/default/files/community-code/met/docs/user-guide/MET_Users_Guide_v8.1.2.pdf), date last accessed: 2021-02-16. (cited on pages: 85, 87)
- Ngan, F., D. Byun, H. Kim, D. Lee, B. Rappenglück, and A. Pour-Biazar, 2012: Performance assessment of retrospective meteorological inputs for use in air quality modeling during TexAQS 2006. *Atmospheric Environment*, **54**, 86–96, doi:10.1016/j.atmosenv.2012.01.035. (cited on page: 5)
- Ngan, F., H. Kim, P. Lee, K. Al-Wali, and B. Dornblaser, 2013: A study of nocturnal surface wind speed overprediction by the WRF-ARW model in Southeastern Texas. *Journal of Applied Meteorology and Climatology*, **52**, 2638–2653, doi:10.1175/JAMC-D-13-060.1. (cited on page: 5)

- O'Connor, F. M., C. E. Johnson, O. Morgenstern, N. L. Abraham, P. Braesicke, M. Dalvi, G. A. Folberth, M. G. Sanderson, P. J. Telford, A. Voulgarakis, P. J. Young, G. Zeng, W. J. Collins, and J. A. Pyle, 2014: Evaluation of the new UKCA climate-composition model – Part 2: The Troposphere. *Geoscientific Model Development*, **7**, 41–91, doi:10.5194/gmd-7-41-2014. (cited on pages: 28, 29, and 42)
- Oke, T. R., 1982: The energetic basis of the urban heat island. *Quarterly Journal of the Royal Meteorological Society*, **108**, 1–24, doi:10.1002/qj.49710845502. (cited on page: 25)
- 1987: *Boundary layer climates*. Methuen, 2nd edition. (cited on page: 24)
- Ouyang, W., B. Guo, G. Cai, Q. Li, S. Han, B. Liu, and X. Liu, 2015: The washing effect of precipitation on particulate matter and the pollution dynamics of rainwater in downtown Beijing. *Science of the Total Environment*, **505**, 306–314, doi:10.1016/j.scitotenv.2014.09.062. (cited on page: 75)
- Parrish, D. D., K. C. Aikin, S. J. Oltmans, B. J. Johnson, M. Ives, and C. Sweeny, 2010: Impact of transported background ozone inflow on summertime air quality in a California ozone exceedance area. *Atmospheric Chemistry and Physics*, **10**, 10093–10109, doi:10.5194/acp-10-10093-2010. (cited on page: 118)
- Peake, D., 2012: Coastal influences on vertical pollutant partitioning - PhD Thesis, *University of Reading*. (cited on page: 121)
- Porson, A., P. A. Clark, I. N. Harman, M. J. Best, and S. E. Belcher, 2010: Implementation of a new urban energy budget scheme in the MetUM. Part I: Description and idealized simulations. *Quarterly Journal of the Royal Meteorological Society*, **136**, 1514–1529, doi:10.1002/qj.668. (cited on pages: 125, 128, 129, and 173)
- Pryor, S. C. and D. G. Steyn, 1995: Hebdomadal and diurnal cycles in ozone time series from the Lower Fraser Valley, B.C. *Atmospheric Environment*, **29**, 1007–1019, doi:10.1016/1352-2310(94)00365-R. (cited on page: 19)
- Purvis, R. M., J. B. McQuaid, A. C. Lewis, J. R. Hopkins, and P. Simmonds, 2005: Horizontal and vertical profiles of ozone, carbon monoxide, non-methane hydrocarbons and dimethyl sulphide near the Mace Head observatory, Ireland. *Atmospheric Chemistry and Physics Discussions*, **5**, 12505–12530, doi:10.5194/acpd-5-12505-2005. (cited on pages: 117, 142)
- Roberts, N., 2008: Assessing the spatial and temporal variation in the skill of precipitation forecasts from an NWP model. *Meteorological Applications*, **15**, 163–169, doi:10.1002/met.57. (cited on page: 10)
- Roberts, N. M. and H. W. Lean, 2008: Scale-Selective Verification of Rainfall Accumulations from High-Resolution Forecasts of Convective Events. *Monthly Weather Review*, **136**, 78–97, doi:10.1175/2007MWR2123.1. (cited on pages: 65, 82)
- Ryder, C. L., E. J. Highwood, A. Walser, P. Seibert, A. Philipp, and B. Weinzierl, 2019: Coarse and giant particles are ubiquitous in Saharan dust export regions and are radiatively significant over the Sahara. *Atmospheric Chemistry and Physics*, **19**, 15353–15376, doi:10.5194/acp-19-15353-2019. (cited on page: 7)
- Ryu, Y.-H., A. Hodzic, J. Barre, G. Descombes, and P. Minnis, 2018: Quantifying errors in surface ozone predictions associated with clouds over the CONUS: a WRF-Chem modeling study using satellite cloud retrievals. *Atmospheric Chemistry and Physics*, **18**, 7509–7525, doi:10.5194/acp-18-7509-2018. (cited on pages: 4, 31, and 47)

- Savage, N. H., P. Agnew, L. S. Davis, C. Ordonez, R. Thorpe, C. E. Johnson, F. M. O'Connor, and M. Dalvi, 2013: Air quality modelling using the Met Office Unified Model (AQUUM OS24-26): model description and initial evaluation. *Geoscientific Model Development*, **6**, 353–372, doi:10.5194/gmd-6-353-2013. (cited on pages: 4, 8, 30, 42, 43, 45, 47, 51, 52, 78, and 130)
- Schiff, H. I., 1969: Neutral reactions involving oxygen and nitrogen. *Canadian Journal of Chemistry*, **47**, 1916. (cited on page: 15)
- Schroedter-Homscheidt, M., A. Arola, N. Killius, M. Lefèvre, L. Saboret, W. Wandji, L. Wald, and E. Wey, 2016: The Copernicus Atmosphere Monitoring Service (CAMS) Radiation Service in a nutshell. doi:10.5194/amt-6-2403-2013. (cited on page: 6)
- Seaman, N. L., 1999: Meteorological modeling for air-quality assessments. *Atmospheric Environment*, 2231–2259. (cited on page: 169)
- Seinfeld, J. H. and S. N. Pandis, 2016: *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change, 3rd Edition*. Wiley, New Jersey, 3 edition, 1152 pp. (cited on pages: 18, 28, and 30)
- Shi, Z., T. Vu, S. Kotthaus, R. M. Harrison, S. Grimmond, S. Yue, T. Zhu, J. Lee, Y. Han, M. Demuzere, R. E. Dunmore, L. Ren, D. Liu, Y. Wang, O. Wild, J. Allan, W. Joe Acton, J. Barlow, B. Barratt, D. Beddows, W. J. Bloss, G. Calzolari, D. Carruthers, D. C. Carslaw, Q. Chan, L. Chatzidiakou, Y. Chen, L. Crilley, H. Coe, T. Dai, R. Doherty, F. Duan, P. Fu, B. Ge, M. Ge, D. Guan, J. F. Hamilton, K. He, M. Heal, D. Heard, C. Nicholas Hewitt, M. Hollaway, M. Hu, D. Ji, X. Jiang, R. Jones, M. Kalberer, F. J. Kelly, L. Kramer, B. Langford, C. Lin, A. C. Lewis, J. Li, W. Li, H. Liu, J. Liu, M. Loh, K. Lu, F. Lucarelli, G. Mann, G. McFiggans, M. R. Miller, G. Mills, P. Monk, E. Nemitz, F. O'Connor, B. Ouyang, P. I. Palmer, C. Percival, O. Popoola, C. Reeves, A. R. Rickard, L. Shao, G. Shi, D. Spracklen, D. Stevenson, Y. Sun, Z. Sun, S. Tao, S. Tong, Q. Wang, W. Wang, X. Wang, X. Wang, Z. Wang, L. Wei, L. Whalley, X. Wu, Z. Wu, P. Xie, F. Yang, Q. Zhang, Y. Zhang, Y. Zhang, and M. Zheng, 2019: Introduction to the special issue: “In-depth study of air pollution sources and processes within Beijing and its surrounding region (APHH-Beijing)”. *Atmospheric Chemistry and Physics*, **19**, 7519–7546, doi:10.5194/acp-19-7519-2019. (cited on page: 7)
- Sillman, S., J. A. Logan, and S. C. Wofsy, 1990: The sensitivity of ozone to nitrogen oxides and hydrocarbons in regional ozone episodes. *Journal of Geophysical Research*, **95**, 1837, doi:10.1029/JD095iD02p01837. (cited on page: 18)
- Sillman, S. and P. J. Samson, 1995: Impact of temperature on oxidant photochemistry in urban polluted rural and remote environments. *Journal of Geophysical Research*, **100**, 497–11, doi:10.1029/94jd02146. (cited on page: 53)
- Sinclair, V. A., S. E. Belcher, and S. L. Gray, 2010: Synoptic controls on boundary-layer characteristics. *Boundary-Layer Meteorology*, **134**, 387–409, doi:10.1007/s10546-009-9455-6. (cited on page: 119)
- Smith, R. N. B., 1990: A scheme for predicting layer clouds and their water content in a general circulation model. *Quarterly Journal of the Royal Meteorological Society*, **116**, 435–460, doi:10.1002/qj.49711649210. (cited on page: 126)
- Solazzo, E., R. Bianconi, G. Pirovano, M. D. Moran, R. Vautard, C. Hogrefe, K. W. Appel, V. Matthias, P. Grossi, B. Bessagnet, J. Brandt, C. Chemel, J. H. Christensen, R. Forkel, X. V. Francis, A. B. Hansen, S. McKeen, U. Nopmngcol, M. Prank, K. N. Sartelet, A. Segers, J. D. Silver, G. Yarwood, J. Werhahn, J. Zhang, S. T. Rao, and

- S. Galmarini, 2013: Evaluating the capability of regional-scale air quality models to capture the vertical distribution of pollutants. *Geoscientific Model Development*, **6**, 791–818, doi:10.5194/gmd-6-791-2013. (cited on page: 118)
- Solazzo, E. and S. Galmarini, 2016: Error apportionment for atmospheric chemistry-transport models - A new approach to model evaluation. *Atmospheric Chemistry and Physics*, **16**, doi:10.5194/acp-16-6263-2016. (cited on page: 118)
- Song, Y., D. Han, and J. Zhang, 2017: Radar and rain gauge rainfall discrepancies driven by changes in atmospheric conditions. *Geophysical Research Letters*, **44**, 7303–7309, doi:10.1002/2017GL074493. (cited on page: 65)
- Staniforth, A., 1997: Regional modeling: A theoretical discussion. *Meteorology and Atmospheric Physics*, **63**, 15–29, doi:10.1007/BF01025361. (cited on page: 5)
- Staniforth, A. and N. Wood, 2003: The deep-atmosphere Euler equations in a generalized vertical coordinate. *Monthly Weather Review*, **131**, 1931–1938, doi:10.1175//2564.1. (cited on page: 41)
- Stull, R. B., 1988: *An Introduction to Boundary Layer Meteorology*. Atmospheric sciences library., Kluwer Academic, Dordrecht. (cited on pages: 22, 127, and 136)
- Stull, R. B. and C. D. Ahrens, 2000: *Meteorology for Scientists and Engineers*. Brooks / Cole, 2nd edition, 502 pp. (cited on page: 118)
- Sun, J., L. Mahrt, R. M. Banta, and Y. L. Pichugina, 2012: Turbulence Regimes and Turbulence Intermittency in the Stable Boundary Layer during CASES-99. *Journal of the Atmospheric Sciences*, **69**, 338–351, doi:10.1175/JAS-D-11-082.1. (cited on pages: 26, 95, and 172)
- Tang, Y., P. Lee, M. Tsidulko, H.-C. Huang, J. T. McQueen, G. J. DiMego, L. K. Emmons, R. B. Pierce, A. M. Thompson, H.-M. Lin, D. Kang, D. Tong, S. Yu, R. Mathur, J. E. Pleim, T. L. Otte, G. Pouliot, J. O. Young, K. L. Schere, P. M. Davidson, and I. Stajner, 2009: The impact of chemical lateral boundary conditions on CMAQ predictions of tropospheric ozone over the continental United States. *Environmental Fluid Mechanics*, **9**, 43–58, doi:10.1007/s10652-008-9092-5. (cited on page: 5)
- Taylor, G. I., 1922: Diffusion by Continuous Movements. *Proceedings of the London Mathematical Society*, **S2-20**, 196–212, doi:10.1112/plms/s2-20.1.196. (cited on page: 124)
- Travis, K. R., D. J. Jacob, J. A. Fisher, P. S. Kim, E. A. Marais, L. Zhu, K. Yu, C. C. Miller, R. M. Yantosca, M. P. Sulprizio, A. M. Thompson, P. O. Wennberg, J. D. Crounse, J. M. St. Clair, R. C. Cohen, J. L. Laughner, J. E. Dibb, S. R. Hall, K. Ullmann, G. M. Wolfe, I. B. Pollack, J. Peischl, J. A. Neuman, and X. Zhou, 2016: Why do models overestimate surface ozone in the Southeast United States? *Atmospheric Chemistry and Physics*, **16**, 13561–13577, doi:10.5194/acp-16-13561-2016. (cited on pages: 3, 32, 119, and 167)
- Trousdell, J. F., S. A. Conley, A. Post, and I. C. Faloona, 2016: Observing entrainment mixing, photochemical ozone production, and regional methane emissions by aircraft using a simple mixed-layer framework. *Atmospheric Chemistry and Physics*, **16**, 15433–15450, doi:10.5194/acp-16-15433-2016. (cited on pages: 118, 168)
- Tsagatakis, I., M. Ruddy, J. Richardson, A. Otto, B. Pearson, and N. Passant, 2019: UK Emission Mapping Methodology 2017: A report of the National Atmospheric Emissions Inventory (NAEI). Technical report, NAEI, available at [https://naei.beis.gov.uk/reports/reports?report\\_id=997](https://naei.beis.gov.uk/reports/reports?report_id=997), date last accessed 2021-02-16. (cited on page: 16)

- UK Department for Transport, 2021: Transport and environment statistics 2021 annual report. Available at [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/984685/transport-and-environment-statistics-2021.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/984685/transport-and-environment-statistics-2021.pdf), last accessed 17-08-2021. (cited on page: 14)
- Vaughan, A. R., J. D. Lee, P. K. Misztal, S. Metzger, M. D. Shaw, A. C. Lewis, R. M. Purvis, D. C. Carslaw, A. H. Goldstein, C. N. Hewitt, B. Davison, S. D. Beevers, and T. G. Karl, 2016: Spatially resolved flux measurements of NO<sub>x</sub> from London suggest significantly higher emissions than predicted by inventories. *Faraday Discussions*, **189**, 455–472, doi:10.1039/c5fd00170f. (cited on pages: 48, 93)
- Vautard, R., M. D. Moran, E. Solazzo, R. C. Gilliam, V. Matthias, R. Bianconi, C. Chemel, J. Ferreira, B. Geyer, A. B. Hansen, A. Jericevic, M. Prank, A. Segers, J. D. Silver, J. Werhahn, R. Wolke, S. T. Rao, and S. Galmarini, 2012: Evaluation of the meteorological forcing used for the Air Quality Model Evaluation International Initiative (AQMEII) air quality simulations. *Atmospheric Environment*, doi:10.1016/j.atmosenv.2011.10.065. (cited on page: 3)
- Wayne, R. P., 2000: *Chemistry of Atmospheres: An Introduction to the Chemistry of the Atmospheres of Earth, the Planets, and Their Satellites*. Oxford University Press, 3rd edition, 775 pp. (cited on pages: 14, 28)
- Webster, H. N. and D. J. Thomson, 2011: Dry deposition modelling in a Lagrangian dispersion model. *International Journal of Environment and Pollution*, **47**, 1–9. (cited on page: 124)
- Wernli, H., M. Paulat, M. Hagen, and C. Frei, 2008: SAL—A Novel Quality Measure for the Verification of Quantitative Precipitation Forecasts. *Monthly Weather Review*, **136**, 4470–4487, doi:10.1175/2008MWR2415.1. (cited on pages: 65, 81)
- Wesely, M. L., 1989: Parameterization of surface resistances to gaseous dry deposition in regional-scale numerical models. *Atmospheric Environment*, **23**, 1293–1304. (cited on page: 29)
- Weygandt, S. S., A. F. Loughe, S. G. Benjamin, and J. L. Mahoney, 2004: Scale sensitivities in model precipitation skill scores during ihop. *22nd Conference on Severe Local Storms*, American Meteorological Society. (cited on page: 10)
- Whall, T., T. Scarbrough, A. Stravrakaki, C. Green, J. Squire, and R. Noden, 2010: DEFRA UK Ship Emissions Inventory. Technical report, Entec UK Ltd, London, available at [https://uk-air.defra.gov.uk/assets/documents/reports/cat15/1012131459\\_21897\\_Final\\_Report\\_291110.pdf](https://uk-air.defra.gov.uk/assets/documents/reports/cat15/1012131459_21897_Final_Report_291110.pdf), date last accessed 2021-02-16. (cited on page: 43)
- Wild, O., X. Zhu, and M. J. Prather, 2000: Fast-J: Accurate Simulation of In-and Below-Cloud Photolysis in Tropospheric Chemical Models. *Journal of Atmospheric Chemistry*, **37**, 245–282. (cited on page: 42)
- Wilks, D. S., 2020: Forecast Verification. *Statistical Methods in the Atmospheric Sciences*, Elsevier, chapter 9, 4th edition, 369–483. (cited on page: 33)
- Williams, R. S., M. I. Hegglin, B. J. Kerridge, P. Jöckel, B. G. Latter, and D. A. Plummer, 2019: Characterising the seasonal and geographical variability in tropospheric ozone, stratospheric influence and recent changes. *Atmospheric Chemistry and Physics*, **19**, 3589–3620, doi:10.5194/acp-19-3589-2019. (cited on page: 14)

- Wood, N., A. Staniforth, A. White, T. Allen, M. Diamantakis, M. Gross, T. Melvin, C. Smith, S. Vosper, M. Zerroukat, and J. Thuburn, 2014: An inherently mass-conserving semi-implicit semi-Lagrangian discretization of the deep-atmosphere global non-hydrostatic equations. *Quarterly Journal of the Royal Meteorological Society*, **140**, 1505–1520, doi:10.1002/qj.2235. (cited on page: 41)
- World Health Organization, 2006: Air Quality Guidelines Global Update (2005). Technical report, World Health Organisation, available at [http://euro.who.int/\\_\\_data/assets/pdf\\_file/0005/78638/E90038.pdf](http://euro.who.int/__data/assets/pdf_file/0005/78638/E90038.pdf), date last accessed: 2021-02-16. (cited on page: 1)
- 2020: Ambient (outdoor) air pollution - key facts. Available at [https://www.who.int/en/news-room/fact-sheets/detail/ambient-\(outdoor\)-air-quality-and-health](https://www.who.int/en/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health), date last accessed: 2021-02-16. (cited on page: 1)
- Wu, Q., W. Xu, A. Shi, Y. Li, X. Zhao, Z. Wang, J. Li, and L. Wang, 2014: Air quality forecast of PM 10 in Beijing with Community Multi-scale Air Quality Modeling (CMAQ) system: emission and improvement. *Geosci. Model Dev*, **7**, 2243–2259, doi:10.5194/gmd-7-2243-2014. (cited on page: 180)
- Wu, S., L. J. Mickley, D. J. Jacob, J. A. Logan, R. M. Yantosca, and D. Rind, 2007: Why are there large differences between models in global budgets of tropospheric ozone? *Journal of Geophysical Research*, **112**, D05302, doi:10.1029/2006JD007801. (cited on pages: 15, 28)
- Yates, E., S. Anquetin, V. Ducrocq, J. D. Creutin, D. Ricard, and K. Chancibault, 2006: Point and areal validation of forecast precipitation fields. *Meteorological Applications*, **13**, 1–20, doi:10.1017/S1350482705001921. (cited on page: 82)
- Žabkar, R., L. Honzak, G. Skok, R. Forkel, J. Rakovec, A. Cegljar, and N. Žagar, 2015: Evaluation of the high resolution WRF-Chem (v3.4.1) air quality forecast and its comparison with statistical ozone predictions. *Geoscientific Model Development*, **8**, 2119–2137, doi:10.5194/gmd-8-2119-2015. (cited on page: 180)
- Zepeda-Arce, J., E. Foufoula-Georgiou, and K. K. Droegemeier, 2000: Space-time rainfall organization and its role in validating quantitative precipitation forecasts. *Journal of Geophysical Research: Atmospheres*, **105**, 10129–10146, doi:10.1029/1999JD901087. (cited on pages: 10, 82)
- Zhang, F., N. Bei, J. W. Nielsen-Gammon, G. Li, R. Zhang, A. Stuart, and A. Aksoy, 2007: Impacts of meteorological uncertainties on ozone pollution predictability estimated through meteorological and photochemical ensemble forecasts. *Journal of Geophysical Research*, **112**, D04304, doi:10.1029/2006JD007429. (cited on page: 169)
- Zhang, R., A. T. White, A. Pour Biazar, R. T. McNider, and D. S. Cohan, 2018: Incorporating GOES Satellite Photosynthetically Active Radiation (PAR) Retrievals to Improve Biogenic Emission Estimates in Texas. *Journal of Geophysical Research: Atmospheres*, **123**, 1309–1324, doi:10.1002/2017JD026792. (cited on page: 31)
- Zhang, Y., 2008: Online-coupled meteorology and chemistry models: history, current status, and outlook. *Atmospheric Chemistry and Physics*, **8**, 2895–2932, doi:10.5194/acp-8-2895-2008. (cited on page: 2)
- Zhang, Y., M. Bocquet, V. Mallet, C. Seigneur, and A. Baklanov, 2012: Real-time air quality forecasting, part II: State of the science, current research needs, and future prospects. *Atmospheric Environment*, **60**, 656–676, doi:10.1016/J.ATMOENV.2012.02.041. (cited on page: 3)

- Zhang, Y., C. Jena, K. Wang, C. Paton-Walsh, -A. Guérette, S. Utembe, J. D. Silver, Keywood, and Melita, 2019: Multiscale Applications of Two Online-Coupled Meteorology-Chemistry Models during Recent Field Campaigns in Australia, Part I: Model Description and WRF/Chem-ROMS Evaluation Using Surface and Satellite Data and Sensitivity to Spatial Grid Resolutions. *Atmosphere*, **10**, 189, doi:10.3390/atmos10040189. (cited on page: 180)
- Zhang, Y., K. Sartelet, S. Zhu, W. Wang, S.-Y. Wu, X. Zhang, K. Wang, P. Tran, C. Seigneur, and Z.-F. Wang, 2013: Application of WRF/Chem-MADRID and WRF/Polyphemus in Europe – Part 2: Evaluation of chemical concentrations and sensitivity simulations. *Atmospheric Chemistry and Physics*, **13**, 6845–6875, doi:10.5194/acp-13-6845-2013. (cited on page: 3)