Improving the representation of ozone in a chemical transport model

Peter D. Ivatt

Doctor of Philosophy

University of York

Chemistry

September 2020

Abstract

Ozone is a compound of interest in the troposphere due to its detrimental effect on human health and the environment and is of particular concern in Beijing. Here the GEOS-Chem chemical transport model was run with improved chemistry regionally over east China and evaluated against the 2017 Atmospheric Pollution and Human Health in a Chinese Megacity campaign. The primary emissions are found to require significant scaling to match observation. While the model was able to reproduce the O_3 concentration, the secondary compound performance was lacking. A system to produce O_3 isopleths was developed to investigate the photochemical regime, and it was found that during the campaign, Beijing was on average in a VOC limited regime, but there was a degree of day to day variability. The VOC evaluation found that the most significant reduction in O_3 would result from a reduction in aromatic compounds. The significance of reactive uptake of HO_2 onto aerosol was investigated and found to profoundly affect the O_3 concentration with the rate of NO emissions affecting the sensitivity. The noontime radical loss pathways were investigated, and it is found that while NO_2 + OH was the dominant radical loss pathway in Beijing, the contribution of radical-radical reactions and HO_2 uptake varies throughout the campaign period. Finally, a machine learning approach was developed to correct the O_3 bias within GEOS-Chem. The XGBoost algorithm was trained on six years (2010-2015) of global observational data and tested on two years (2016-2017). The model was able to predict and thus improve the model performance at the nine testing sites and identify known published biases globally. The approach was robust to the removal of training data, both spatially and temporally and could be a useful tool for performance prioritized applications such as air quality forecasts.

Contents

List of Figures

List of Tables

Authors Declaration

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as References.

Chapter 1

Introduction

1.1 Introduction

This thesis will focus on improving our understanding of tropospheric ozone (O_3), a pollutant which impacts human health, reduces crop yields and acts as a greenhouse gas. O_3 lies at the centre of the chemistry of the troposphere due to its ability to form the OH radical, which constitutes the most important tropospheric chemical sink. Understanding the processes controlling its concentrations requires understanding a wide range of chemical and physical processes.

A significant body of previous work has been undertaken to understand O_3. This thesis aims to fill three primary gaps in our understanding. The first two chapters focus on summertime O_3 in Beijing. Firstly, to improve our understanding of the optimal route to control O_3 in the city, and secondly, to explore aerosols' role in controlling O_3 in that city. The final chapter looks more broadly to provide a new method of being able to simulate O_3 around the world. All of these chapters use chemical transport models as a tool to support this understanding.

Improvement of chemical transport models is usually "process" based meaning that additional scientific understanding is programmed into the model to improve it. A process-based approach is taken in the first two chapters where the motivation is to maximise understanding in the model. New science is explicitly added to the model, and new techniques are developed for diagnosing model behaviour. A process-based approach to model improvement is contrasted with the third chapter, where an algorithmic approach is used to maximise performance. The algorithm is then investigated for its potential to aid process based improvement.

Given Beijing's focus, this work is motivated by the potential health impacts of O_3 and air pollution more widely. Thus in this chapter, the impact of air pollution on human health is first described (Section 1.3). The sources of atmospheric pollutants (Section 1.4) and the tropospheric gas chemistry involved in understanding O_3 will be outlined (Section 1.5). Finally, this work's chemical transport modelling will be introduced (Section 1.6).

1.2 The history of air quality

Elevated concentration levels of O_3 results in poor air quality and is keystone in our modern understanding of air pollution however air quality has long been a concern for humanity. Much of this stems from the overt nature of acrid smoke and foul odour. As far back as the ancient Greeks there was an association of foul odour (particularly concerning stagnant swamps) with "unhealthiness", while this was more reflective of water quality it created the accepted concept of "bad air". The earliest documented case of air pollution in the UK occurred in 1257 when Queen Eleanor visited Nottingham castle which was under repairs at the time. The Queen was forced to leave the site "to preserve her health" due to the smoke from the "new fuel", coal (Brimblecombe, 1976).

The fear of air pollution was common throughout medieval times, but it was not until the industrial revolution that it become a prominent social issue. The widespread burning of coal in dense urban areas created a perfect storm for air quality issues. During this time, scientists identified that chimney sweeps were developing occupation specific cancer, and the connection was made that coal soot was to blame (Waldron, 1983). This was the first time a chemical substance had been identified as causing cancer. Between 1870 and 1897 periods of smog (the simultaneous occurrences of "smoke" and "fog") and low temperature were blamed for a sharp rise in deaths in the London borough of Islington (Russell, 1924). Unfortunately, no action was taken and in winter of 1952 London experienced a significant pollution event in which the smog was so thick that traffic in the city came to a standstill and a substantial increase in the mortality rate followed (M. L. Bell, D. L. Davis, and Fletcher, 2004). These deaths were found to be the result of respiratory and cardiovascular complications (W. P. D. Logan, 1953). Up until this point pollution was primarily seen as an unavoidable consequence of urbanisation, however in the wake of the London smog event, it was deemed that this was no longer acceptable, and in 1956 the Clean Air Act came into effect. This was the first significant piece of legalisation to protect human health concerning air pollution, this act put in place controls on coal combustion.

During this time, Los Angeles was also experiencing smog events. However, this smog had a yellow colour and was not associated with cold temperatures. This was identified to be a "photochemical smog". Photochemical smog is where volatile organic compounds (VOCs) and nitrogen oxides (NO_x) react to form ozone (O_3), reactive gases and particulate. While all of the species associated with photochemical smog are detrimental to human health and the environment, it is O_3 that is of greatest interest. O_3 is of particular importance due to its problematically high concentration in polluted regions and long lifetime, allowing it to be transported over large regions.

While this thesis will focus on O_3, the atmosphere is a highly coupled system, and thus it is impossible to isolate one compound. Policies will often look to tackle high concentrations of multiple classes of harmful compounds at once. Furthermore, some policy decisions can even make certain species worse in the process. The effect on human health and ecosystem damage for the most problematic atmospheric compounds will now be discussed.

1.3 Health and ecosystem impact of air pollution

The degradation of air quality has repercussions for society and the natural environment (Manisalidis et al., 2020). The impact of pollution can occur throughout the atmosphere, such as chlorofluorocarbon damage to the O_3 layer in the stratosphere (Farman, Gardiner, and Shanklin, 1985). Additionally, the effects of indoor air pollution have also been investigated (Spengler and Sexton, 1983; Samet, Marbury, and Spengler, 1987; Bernstein et al., 2008). In this thesis, the focus is on the repercussions at the bottom-most 15km of the atmosphere, the troposphere, in outdoor air.

Many compounds in the troposphere can have a negative impact under certain conditions. Here they are split into two classes and discussed: particulate consisting of tropospheric aerosols (solid and liquid particles suspended in the air) and gases.

1.3.1 Particulate

Particulate matter (PM) consists of suspended particles in the air, and their negative impact on health has been extensively studied. They are classified by their maximum diameter in microns with the two main classes being PM2.5 (2.5 μm) and PM10 (10 μm). While all particle sizes pose a level risk it is the smaller PM2.5 that is of most concern due to its ability to most effectively infiltrate deep into the respiratory system (Kelishadi and Poursafa, 2010; Lingli Zhang et al., 2019). These small particles can deposit on the bronchioles and alveoli (Londahl et al., 2006) where they obstruct gas exchange and can penetrate the bloodstream (Guaita et al., 2011). This process puts considerable stress on both the respiratory and cardiovascular systems (Halonen et al., 2009; Perez et al., 2012; Samoli et al., 2008). As a result, PM2.5 is predicted to be the cause of approximately 4 million deaths per year (~7.6% of all deaths) (A. J. Cohen et al., 2017).

The composition of particulate can cause additional issues. In urban and industrial regions, PM may also contain heavy metals (Popoola, Adebanjo, and Adeoye, 2018). Heavy metals such as lead, cadmium and zinc emitted from industry and vehicular emissions (Viard et al., 2004) can bioaccumulate in the body, causing severe long term damage to various bodily systems in both humans and animals (Goyer, 1990; Jaishankar et al., 2014).

Polycyclic aromatic hydrocarbons are formed of a combination of two to six aromatic rings via low heat combustion such as during biomass burning or domestic solid fuel use (D. Liu et al., 2017). The larger of these compounds (greater than four rings) have low volatility due to their size and are thus found exclusively in the particulate phase (Jariyasopit et al., 2014). PAH's primary health risk is their carcinogenicity and mutagenicity (S. J. C. Wei et al., 1991), the mechanism for this comes not from the PAH compounds themselves but from the metabolites formed within the body. The larger PAH are of the most significant risk due to their metabolites having a structure that allows them to interact with DNA (Huberman et al., 1976).

When particulate undergoes deposition via moisture or precipitation, nitric and sulfuric acid components of the aerosol can cause a drop in the pH of the water droplet (Parungo, Nagamoto, and Madel, 1987). The formed acid rain can then cause ecological harm (A. Singh and Agrawal, 2008) as well as economic damage via the destruction of food crops (Irving, 1983). PM can cause further damage to plant life due to a reduction in photosynthesis brought about by decreased visibility and changes to leaf surface temperature due to the dry deposition resulting in a coating of particulate.

1.3.2 Gases

Nitrogen oxides (NO_x) consisting of NO and NO_2 are emitted predominantly from vehicular exhaust (Richmond-Bryant et al., 2017) and are gas-phase compounds of concern for human health. Nitrogen oxides have been found to increase the risk of cardiovascular and respiratory diseases (Brunekreef et al., 2009) as well as causing irritation to the eyes, throat, and nose (T.-M. Chen et al., 2007).

Volatile Organic Compounds (VOCs) are hydrocarbons (or compounds based on hydrocarbon molecule skeletons) that are found predominantly in the gas phase. VOCs have some associated health risks such as carcinogenic compounds benzene, toluene, xylene (Dehghani et al., 2018) and formaldehyde (Pilidis et al., 2009). However, the most problematic consequences of NO_x and VOC emissions is the subsequent production of reactive gases in the presence of sunlight.

The secondary compounds formed by the photochemical reactions include; peroxyacetyl nitrates (PANs), aldehydes and nitric acid and O_3. The combination of these secondaries with VOC and NO_x is known as photochemical smog. These reactive compounds are known to irritate eyes (Altshuller, 1977) and lungs (Jaffe, 1967) as well as the damage they can cause to vegetation (Taylor, 1969). However, of these secondary gases, it is O_3 with its potentially high concentration and relatively long tropospheric lifetime (weeks in the summer) that is of most significant concern.

Long term exposure to O_3 has been found to cause decreased respiratory and pulmonary function, airway inflammation, and aggravation of pre-existing diseases such as asthma (Schneidemesser et al., 2015; Kulle et al., 1985; Lippmann, 1989). Approximately 1.1 million premature deaths a year are believed to be a result of long-term O_3 exposure (Lim et al., 2012; R. A. Silva et al., 2013; Malley et al., 2017).

Exposure to O_3 can damage plant life via mechanisms such as accelerating leaf deterioration (Reich and Lassoie, 1985; Pell et al., 1999), altering carbon allocation (Friend and Tomlinson, 1992; Coleman et al., 1995) and predisposing trees to attack by pests (Stark et al., 1968; D. F. Karnosky et al., 2002). As a result, O_3 has caused long term detriment to forested regions (David F. Karnosky et al., 2007).

O_3 can severely damage crops leading to reduced crop yield and quality (Krupa, Nosal, and Legge, 1998). Total global annual economic loss as a result of O_3 is predicted to be between 14 and 26 billion dollars (Van Dingenen et al., 2009). Additionally, much of this crop damage occurs in developing countries impacting food security (Tai and Maria Val Martin, 2017).

1.4 Emissions

A large number of different gas phase molecules are emitted into the atmosphere (Duce, 1978). Their mass and composition is limited by their volatility (they need to be gases at surface temperatures) and their availability at the surface. This limits the range of possible molecules to, in general, those that contain H, O, C, N, S, P, Br, I, Cl and Si.

These compounds have differing impacts on the atmosphere. Some are inert in the lower part of the atmosphere and only influence the stratosphere (CFCs, HCFCs) due to the high energy required to break their chemical bonds (S. A. Montzka et al., 2018). From an urban health perspective there are two key group of compounds of for O_3 production. Firstly, VOCs and CO, these form the "fuel" for ozone production, and secondly, oxides of nitrogen, these form the catalyst for the production of ozone (Haagensmit, Bradley, and Fox, 1953).

1.4.1 Volatile organic compounds and carbon monoxide

Volatile organic compounds (VOCs) are hydrocarbons or hydrocarbon like compounds that are predominantly found in the gas phase at atmospherically relevant temperatures (R. Atkinson and Arey, 2003a). This is a diverse group of compounds including those from anthropogenic (derived from human activity) and biogenic (derived from natural activity in the biosphere) sources (R. Atkinson and Arey, 2003b), with lifetimes ranging from <1 hour in the case of unstable unsaturated branched compounds, to over ten years in the case of methane (Sonnemann and Grygalashvyly, 2014).

The long lifetime of methane coupled with substantial globally distributed emissions lead to it having the highest concentration of any of the VOCs in the atmosphere (Saunois et al., 2016). But it has little impact on the urban or regional spatial scale (A. M. Fiore, Jacob, et al., 2002). The other VOCs have a more profound influence (Derwent, Jenkin, and Saunders, 1996). Their anthropogenic source is dominated by release of unused hydrocarbon fuels or their incomplete combustion within engines (Hong-li et al., 2017). However, they are also used as industrial solvents, feed-stocks of industrial processes, and in domestic applications (Y. M. Kim, Harrad, and Harrison, 2001). All of these lead to intentional or unintentional release of these compounds into the atmosphere.

A significant amount of organic matter is tied up in the biosphere predominantly in plant life. This can be released into the atmosphere in two main ways. Plants naturally release VOCs into the atmosphere (Simpson et al., 1995). Vast quantities of compounds such as isoprene and monoterpenes are release into the atmosphere by plants (A. Guenther et al., 2006), the exact reason for doing this is unclear but as the emission increase with temperature it may be a mechanism for increasing heat tolerance. These emissions dwarf the anthropogenic source in many parts of the world (Jacob and S. C. Wofsy, 1988). Forest fires can also release substantial quantities of VOCs into the atmosphere as biomass is decomposed by the fire (Andreae and Merlet, 2001).

Carbon monoxide (CO) acts in a similar way in the atmosphere to VOCs (Zhong et al., 2017). It is emitted by the incomplete combustion of hydrocarbons, often due to a lack of oxygen during the combustion process (Khalil and Rasmussen, 1990). Well maintained and designed engines and furnaces should result in limited emissions of CO but poor maintenance or design can lead to substantial emissions (Horner, 2000). Similarly, the structure of the fire in forest fires lead to limited availability of oxygen which can lead to high CO emissions (Kasischke and Bruhwiler, 2002).

1.4.2 Oxides of nitrogen

In the troposphere the two prominent nitrogen oxides are NO and NO_2 collectively referred to as NO_x. NO_x is predominately emitted in the form of NO which are oxidised into NO_2 via reaction with an atmospheric oxidant, overwhelmingly O_3 (Chameides et al., 1992). This reaction forms the central part of the NO_x cycle (R. Atkinson, 2000). The nature of the cycle varies depending on atmospheric conditions (M. Val Martin et al., 2008).

NO is formed via the reactions of atmospheric O_2 and N_2, and this reaction takes considerable energy owing to the strong nitrogen triple bond on N_2 (Noxon, 1976). Thus, high temperature combustion is the dominant source of NO_x into the atmosphere$_2$ (Kavanaugh, 1987). Combustion engines produce temperatures greater than 1500 ^{o}C which facilitates the production of NO (P. F. Flynn et al., 2000), and this is particularly evident in diesel engines which run hotter than petrol engines Hoekman and C. Robbins, 2012. Large scale combustion from fossil fuel electrical power plants and industrial furnaces also reach temperatures high enough to produce NO (Foy et al., 2015). Thus these anthropogenic sources of NO_x dominate the global emissions (Delmas, Serca, and Jambert, 1997).

Natural sources of NO_x in the atmosphere also exist (Robinson and R. C. Robbins, 1970). Natural combustion within forest fires can emit substantial amounts of NO (H. B. Singh et al., 2010). Another large natural source of NO is via lightning (D. Z. Stockwell et al., 1999). Lighting has considerable energy and can thus break the O_2 bond (Schumann and Huntrieser, 2007). The oxygen atoms produced can then react with N_2 to produce NO (L. T. Murray, J. A. Logan, and Jacob, 2013). This represents the least quantified source of NO owing to the sparse and instantaneous nature of lighting strikes, as well as effecting the entire column below the storm system (Schumann and Huntrieser, 2007). Microbial activity in soil releases NO and due to its global distribution it is a prominent contributor to remote atmosphere emissions (Robinson and R. C. Robbins, 1970).

1.5 Gas phase chemistry

Once emitted into the atmosphere VOCs and oxides of nitrogen are oxidized (Ravishankara, 1988). This results in the VOCs producing carbon dioxide, water or soluble species which can be lost to clouds or wet surfaces. Oxides of nitrogen are similarly oxidized to give nitric acid or other water soluble species (A. Martin, 1984). This oxidation chemistry is central to controlling the amount of O_3 that the atmosphere can produce and so its health impacts.

VOC oxidation

OH is an extremely reactive radical that plays a central role in this oxidation chemistry (R. Atkinson and Arey, 2003a). It provides the largest removal pathway for VOCs in the atmosphere, and a significant route for the conversion of oxides of nitrogen into nitric acid (Brown, Talukdar, and Ravishankara, 1999).

Most OH in the atmosphere is formed via Reaction R 1.2 when an oxygen atom in its $O(^1D)$ excited state (typically following the photodissociation of ozone, (Reaction R 1.1) reacts with a water molecule (Michelsen et al., 1994).

OH is particularly reactive towards hydrocarbons due to the H-abstraction converting OH into H_2O which is kinetically favourable due to O-H bonds being stronger than C-H bonds (Kleinermanns and Luntz, 1982). The most common example of this reaction is the reaction of OH with methane (CH_4) (Vaghjiani and Ravishankara, 1991), Reaction R 1.3. The methyl radical rapidly reacts with the abundant O_2 to form a methyl peroxy radical (Reaction R 1.4). This radical can be more generally referred to as RO_2, where R is any length alkyl chain (King, Canosa-Mas, and Wayne, 2001). At this point, the RO_2 radical can react in one of two ways depending on the NO_x concentration.

When the concentration of nitric oxide (NO) is high, RO_2 will react with the NO to form an alkoxy radical (RO), in methane's case this is a methoxy radical (Reaction R 1.5). The RO will rapidly react with O_2 through a selection of pathways (Bofill et al., 1999) to form a hydrogenperoxy radical (HO_2) and a formaldehyde molecule (Reaction R 1.6). HO_2 is analagous to RO_2 and together make up the RO_x family of compounds along with hydrogen and OH which can react with O_2 and peroxy radicals respectively to form peroxy radicals (RO_x = H + HO + HO_2 + RO_2).

$$O_3 + hv \longrightarrow O(^1D)\cdot + O_2 \tag{R 1.1}$$

$$O(^1D)\cdot + H_2O \longrightarrow 2\,OH\cdot \tag{R 1.2}$$

$$CH_4 + OH\cdot \longrightarrow CH_3\cdot + H_2O \tag{R 1.3}$$

$$CH_3\cdot + O_2 \longrightarrow CH_3O_2\cdot \tag{R 1.4}$$

$$CH_3O_2\cdot + NO \longrightarrow CH_3O\cdot + NO_2 \tag{R 1.5}$$

$$CH_3O\cdot + O_2 \longrightarrow CH_2O + HO_2\cdot \tag{R 1.6}$$

At this stage looking at reactions 1-6 it can be seen that the reaction of OH with methane, produces two peroxy radicals. As such there has been a 2 fold amplification in radicals.

If the concentration of NO is low, the CH_3O_2 will react with itself to produce a peroxide species (Reactions R 1.7, R 1.8 and R 1.9). These peroxides are highly water soluble and so can be removed from the gas phase (Shallcross et al., 2005).

$$HO_2 + HO_2 \longrightarrow H_2O_2 + O_2 \tag{R 1.7}$$

$$CH_3O_2 + HO_2 \longrightarrow CH_3OOH + O_2 \tag{R 1.8}$$

$$CH_3O_2 + CH_3O_2 \longrightarrow CH_3OOCH_3 + O_2 \tag{R 1.9}$$

If the formaldehyde or peroxides produced are not wet deposited they can then go on to react with an OH or photolyse (Sumner and Shepson, 1999). Following the formaldehyde reaction with OH (Reaction R 1.10) or photolysis pathway (Reaction R 1.11) leads to the production of CHO which goes on to decompose and react with O_2 to form CO and HO_2 (Reaction R 1.12). There is another non-radical photolysis pathway that produces CO and H_2 directly (Reaction R 1.13) (Moortgat and Warneck, 1979). Both CO and H_2 are relatively stable but will ultimately react with OH to produce HO_2 (Reactions R 1.14 and R 1.15) (Meisner and Kaestner, 2016).

$$\mathrm{CH_2O + OH \longrightarrow CHO\cdot + H_2O} \tag{R 1.10}$$

$$\mathrm{CH_2O + hv + O_2 \longrightarrow CHO\cdot + HO_2\cdot} \tag{R 1.11}$$

$$\mathrm{CHO\cdot + O_2 \longrightarrow CO + HO_2\cdot} \tag{R 1.12}$$

$$\mathrm{CH_2O + hv \longrightarrow CO + H_2} \tag{R 1.13}$$

$$\mathrm{H_2 + OH\cdot \longrightarrow HO_2 + H_2O} \tag{R 1.14}$$

$$\mathrm{CO + OH\cdot + O_2 \longrightarrow HO_2 + CO_2} \tag{R 1.15}$$

Thus in this way hydrocarbons can be converted into CO_2 and H_2O. If the environment is wet, water soluble species such as CH_2O can dissolve into the liquid phase and be lost from the gas phase, resulting in the incomplete oxidation of these compounds (Rodhe, F. Dentener, and Schulz, 2002). Longer hydrocarbons may take slightly different reaction pathways and can form complicated molecule structures through the chemical addition of functional groups (peroxides, nitrates, aldehydes, ketones, enols etc.). However, this chain of oxidation, where a parent hydrocarbon is sequentially oxidized through radical chain reaction chemistry, is central to the chemistry of the troposphere (Jenkin, Saunders, and Pilling, 1997).

The oxidation of CO to CO_2 by OH (Reaction R 1.15), occurs via an HOCO adduct to ultimately form HO_2 (Reactions R 1.16, R 1.17 and R 1.18) (Ravishankara and R. L. Thompson, 1983). Through this reaction, CO represents the largest source of HO_2 in the troposphere as well as the biggest OH sink, especially in the remote atmosphere. Once the oxidation of VOCs has reached CO_2 no more oxidation can take place. CO_2 will go on to live in the order of 100 years (Archer, 2005) in the atmosphere and it forms part of the planets carbon cycle via photosynthesis where ultimately it will become incorporated into the lithosphere of the planet (Battin et al., 2009). It is also the most important climate gas (Cox et al., 2000).

$$\mathrm{CO + OH\cdot \longrightarrow HOCO} \tag{R 1.16}$$

$$\mathrm{HOCO \longrightarrow CO + OH} \tag{R 1.17}$$

$$\mathrm{HOCO \longrightarrow CO_2 + H\cdot} \tag{R 1.18}$$

$$\mathrm{H\cdot + O_2 \longrightarrow HO_2\cdot} \tag{R 1.19}$$

Nitrogen oxide chemistry

As discussed in Section 2.4.3, oxides of nitrogen are emitted into the atmosphere from a variety of sources, typically as NO. NO reacts predominantly with O_3 to produce NO_2, but it can also react with the HO_2 and RO_2 formed from the oxidation of VOCs or CO, to also produce NO_2 (R. Atkinson and Arey, 2003a). NO_2 can then go on to photolysis to produce an ground state oxygen atom ($O(^3P)$) and reform NO. Alternatively NO_2 can react with OH to from nitric acid (HNO_3) (Brown, Talukdar, and Ravishankara, 1999). Nitric acid is very soluble and will quickly wet deposit out of the atmosphere, or become bound up in aerosol as aerosol phase NO_3^-. This is an important removal pathway for NO_x, and NO_3^- is an important component of atmospheric aerosol in urban areas (Ravishankara and R. L. Thompson, 1983).

The $O(^3P)$ atom will react with oxygen to produce (O_3), this is a highly energetic reaction and requires a third body (usually N_2 due to its atmospheric abundance) to act as a collision partner and carry away the excess energy to stabilise the reaction (Monks, 2005).

During the day NO_2 also has the potential to react with RO_2 to form organic nitrates. The most important of these forms from the reaction of the peroxy-acetyl radical (CH_3CO_3) with NO_2 to form peroxy-acetyl nitrate (PAN) (Moxim, Levy, and Kasibhatla, 1996). PAN's decomposition rate is highly temperature dependent. In hot condition (>30 ^oC) its lifetime is short (minutes). However, in cold conditions (-20^oC) its lifetime can be weeks. Therefore if the PAN can be produced downwind of urban centres and then transported into the cold upper troposphere it can effectively transport NO_x away from polluted regions (Jaegle, Jacob, et al., 1998). If the air mass descends and warms, the PAN will decompose to release the NO_x back into the air producing higher background concentrations of NO than expected(Bottenheim, Brice, and Anlauf, 1984)

During the night NO_2 reacts with O_3 to form the NO_3 (nitrate) radical, NO_3 can then go on to react with another NO2 to from N_2O_5. The nitrate radical is dominant nigh time tropospheric oxidant. These highly photolabile compound decompose rapidly at sunrise to reform NO and NO_2.

1.5.1 Ozone production and loss

During the day time O_3 is photolysed into $O(^1D)$ or $O(^3P)$ and then rapidly converted back into O_3 as these species react with O_2 (Demore and Raper, 1966). This forms a number of null cycles. Another, daytime null cycle, forms from the reaction of NO with O_3 to give NO_2 and the subsequent photolysis of the NO_2.

$$\mathrm{O_3 + NO \longrightarrow O_2 + NO_2} \tag{R 1.20}$$

$$\mathrm{NO_2 + hv \longrightarrow NO + O(^3P)} \tag{R 1.21}$$

$$\mathrm{O(^3P) + O_2 + M \longrightarrow O_3 + M} \tag{R 1.22}$$

Thus, it is possible to form an odd oxygen (O_x) family composed of O_3, NO_2, $O(^1D)$ and $O(^3P)$. In most regions the O_3 concentration is significantly higher than the concentration of the other members of the family (Bates and Jacob, 2020).

O_x production does not occur during NO_2 photolysis, as this just represents conversion of one member of the family into another. However, it does occur when a peroxy radical reacts with an NO molecule to generate an NO_2 (Reaction R 1.13, R 1.15). As O_3 is the dominant member of the O_x family, the reaction of peroxy radicals with NO leads to the net production of O_3 (Kirchner and W. R. Stockwell, 1996).

O_x loss can occur in a number of ways. The formation of OH from the reaction of $O(^1D)$ with H_2O leads to a loss of O_x and so a loss of O_3. The reaction of O_3 with OH and HO_2 also forms a loss cycle for O_x.

1.5.2 Photochemical regimes

O_3 production is therefore due to the oxidation of VOCs in the presence of sunlight and nitrogen oxides. The VOCs act as a fuel, and the oxides of nitrogen catalyse the process. However, the relationship between the O_3 production and emitted componds is complex (Sillman, 1999). Two different chemical regimes: NOx limited and VOC limited (Blanchard and Fairley, 2001). In a NO_x limited regime increasing the NO_x concentration increases the O_3 production, whereas increasing the VOCs has little impact. In a VOC limited regime the opposite is true, increasing VOC concentration leads to increased O_3 production, whereas increasing the NO_x concentration does not (Thielmann et al., 2001).

These different regimes lie in the different sinks for the RO_x radicals in the atmosphere. Under low NO_x conditions the reactions between RO_x and other RO_x is the dominant sink for radicals (Reactions R 1.7, R 1.8 and R 1.9). Under high NO_x conditions the reaction between OH and NO_2 is the dominant sink (Monks, 2005).

Thus as NO_x concentrations increase the concentrations of radicals remains roughly constant in the regime where the self reaction of RO_x is the dominant sink. As the NO_x concentration increases yet further the dominant sink switches to be the reaction between OH and NO_2 and the radical concentration then drops. In the first regime the radical concentration remains roughly constant but the NO_x concentration increases, this leads to increased O_3 production. This is knows as a NO_x limited environment (Hamer et al., 2015). In the second regime the NO_x concentration continues to rise but the radical concentration drops, so no more O_3 can be produced, and the O_3 production become insensitive to the NO_x concentration. However, in this environment increasing the VOC concentration leads to increased RO_2 concentrations, so O_3 concentrations grow. This regime is knows as VOC limited environment (K. Lu et al., 2010).

At high emissions of NO, nearly all of the O_3 can be forced to react to give NO_2. The O_3 concentration can become very low. This is described as being O_3 titration (Imhoff et al., 1995).

Night-time chemistry

During the night-time the short lived photochemical components (OH, HO_2 and RO_2) are quickly removed as their dominant source has been switch off. NO will now react solely with O_3 to produce NO_2 (Brown, J. A. Neuman, et al., 2006). However NO_2 can no longer photolyse to produce O_3 and will instead react with O_3 to produce the radical NO_3 (Aliwell and R. L. Jones, 1998). As O_3 is no longer being produced and reacts with NO and NO_2 the O_3 concentration will now steadily reduce, however NO_x emissions are usually substantially lower than during the day and thus O_3 doesn't titrate out as quickly as in the daytime VOC limited environment(Ripperton and F. M. Vukovich, 1971).

The NO_3 radical is an important species during the night (Platt et al., 1990). However, it is extremely photolabile and thus cannot exist in any appreciable amount during the day. However at night in the absence of OH it becomes the dominant oxidant in the atmosphere. It may also go to react with NO_2 to form the reservoir species N_2O_5 (R. Atkinson, Winer, and Pitts, 1986). N_2O_5 hydrolyses rapidly on aerosols to form aerosol phase nitrate (Riemer et al., 2003). This represents the dominant night time sink for NO_x species (Ying Chen et al., 2018).

1.6 Atmospheric Modelling

Atmospheric models are used in both research and commercial applications for understanding and predicting the weather, climate, and air quality. They are usually based on numerical solution of the fundamental equations governing atmospheric behaviour, although in recent years, machine learning algorithm approaches have also found utility.

Box models are the most basic form of atmospheric chemistry modelling. These zero-dimensional models treat a sample of air as a well-mixed box of molecules. Each time step the molecules in the box can undergo physical or chemical processes. Processes can include chemical and photolytic reactions, wet or dry deposition and heterogeneous reactions. The "simple" design of these models allows them to be used to investigate very complex systems without too much computational overhead.

Chemical transport models (CTMs) are Eulerian models where a three dimensional grid of box models are solved, and meteorology fields drive transport between the grid boxes. These meteorological fields can be calculated online (within the model) or supplied from a previous calculation with a different model (offline). Offline models can not represent the interaction between chemistry and meteorology. Whereas online models need to represent the coupling between the two systems. Online models are often used to understand climate processes, where changes in the composition of the atmosphere can lead to changes in the radiative balance which in turn can lead to changes in the meteorology and climate. Although these online models allow whole new classes of interactions to be explored they make analysing model output difficult. Has the concentration of a gas in the atmosphere changed between simulations because the chemistry of the atmosphere changed or because the meteorology has changed? Diagnosing changes in online models can prove to be surprisingly complicated. Online models can also be relatively slow to run computationally as they have to solve the equations describing the meteorology as well as the chemistry.

Offline models therefore have a place as they are relatively quick to run and remove some of the interactions which simplifies the analysis. It is such a model that is has been used in this thesis to gain more insight into the tropospheric chemistry of the atmosphere.

The model used here is GEOS-Chem.

1.6.1 GEOS-Chem

GEOS-Chem was formed initially as an offline chemistry model closely driven from the output of the NASA Goddard Earth Observing System (GEOS) Earth system model (Bey, Jacob, et al., 2001). However, over the last twenty years, GEOS-Chem has developed the capability of being driven by multiple different meteorological fields, whether from climate models or from regional meteorological model (H. P. Lin et al., 2020).

GEOS-Chem has been designed to be grid-independent, allowing it to be adapted to the available available meteorology. In this work GEOS-Chem "classic" is used. This is a conventional offline chemistry transport model which can be used either in a global or a nested regional configuration (X. Wang et al., 2019). The model has two vertical resolutions, either a "full" 72 layer grid (mainly used for stratospheric research) or a "reduced" 47 layer resolution with the reduction coming at the expense of stratospheric fidelity (mainly used for tropospheric chemistry).

The horizontal grid is user-defined and can be run either globally or nested, but to run with a nested grid a global model is needed first to produce boundary conditions. The spatial resolution is limited by the availability of met fields from NASA's GEOS system with 0.25°by 0.3125°currently the highest generally available resolution.

The different components of the GEOS-Chem model are now described.

Transport

In GEOS-Chem chemical species are either advected or non-advected. For advected species transport between grid boxes is calculated based on the (S. J. Lin and Rood, 1996) advection algorithm, using achieved meteorology (u,v,w components of the wind fields).

The NASA GEOS system produce two sets of meteorological data products that may be used with GEOS-Chem. The first of these is the Forward Processing (GEOS-FP) system which is a data assimilated product produced at near real-time with a spatial resolution of 0.25°by 0.3125°and a temporal resolution of 1 hour. The second Modern-Era Retrospective analysis for Research and Applications (MERRA-2) which is a reanalysis product (with more data assimilated than the GEOS-FP system) which has used a consistent physical model over many years, enabling the generation of a consistent set of meteorological fields back to the year 1980. MERRA-2 has a spatial resolution of 0.5°by 0.625°and temporal resolution of 3 hours (Molod et al., 2015).

Sub-gridscale processes such as convective transport and boundary layer mixing are calculated using the convective mass fluxes provided by the meteorological fields (X. Q. Wu et al., 2007), and by a non-local turbulence scheme (J. T. Lin and McElroy, 2010).

Emissions

Emissions in GEOS-Chem are handled via the Harvard-NASA Emissions Component (HEMCO), a dedicated emissions module (Keller et al., 2014).

GEOS-Chem includes two classes of emission, meteorologically dependant online emissions and meteorologically independent offline emissions. Online emissions include dust aerosol (Fairlie, Jacob, and R. J. Park, 2007), biogenic VOCs (A. B. Guenther et al., 2012), lightning NO_x (L. T. Murray, J. A. Logan, and Jacob, 2013), sea salt aerosol (Alexander et al., 2005) and soil NO_x (Hudman et al., 2012). The emission rates for these species are calculated online by GEOS-Chem based on meteorological fields. However, as of the recent v12.4 update, the archived output from a high-resolution run is used to keep emissions consistent between resolutions. Nevertheless, if research necessitates alteration to the emissions, they can be switched back online.

Three biomass burning inventories are available in GEOS-Chem; they include GFED (Giglio, Randerson, and Werf, 2013), FINN (Wiedinmyer et al., 2011) and GFAS (Kaiser et al., 2012). These inventories derive emissions from satellite-based burn area predictions, a knowledge of the forest time, and observed emissions ratios between different compounds.

Anthropogenic inventories are generated either globally or over select regions. The regional inventories are generally run as a mosaic on top of the most current global emissions. While GEOS-Chem does include default anthropogenic inventories, in most use cases, specific inventories are chosen based on the study being performed. As such, the default inventories will not be discussed here and specifically chosen and discussed in the model development sections of Chapters 2 and 4.

Chemistry

GEOS-Chem includes coupled aerosol-oxidant chemistry which is calculated via the Kinetic PreProcessor (KPP) chemical solver (Damian et al., 2002). KPP converts the chemical mechanism into FORTRAN code which is solved via numerical scheme at run time.

The tropospheric chemistry scheme includes HO_x, NO_x, VOC, O_3, halogens, aerosols. The scheme is a reduced chemistry scheme, meaning it is not explicit and involves combining products and skipping intermediates reactions in order to maximise computation expediency. While the complexity of the scheme has steadily increased with advances in computational power, some specific species mechanisms have been included; PAN (Fischer et al., 2014), isoprene (Bates and Jacob, 2019), methanol (X. Chen et al., 2019), Methyl, ethyl, and propyl nitrates (Fisher, Atlas, et al., 2018), halogens (X. Wang et al., 2019) and Criegee intermediates (D. B. Millet et al., 2015).

The model includes reactive uptake of NO_2, NO_3, and N_2O_5 (Holmes et al., 2019; McDuffie et al., 2018) as well as HO_2 uptake (Mao, Fan, et al., 2013).

Photolysis rates are calculated via the Fast-JX module (H. S. Bian and M. J. Prather, 2002; Mao, Jacob, et al., 2010). The solar spectrum is binned into 18 wavelengths spanning 177 to 850 nm for full chemistry (coupled stratosphere troposphere) simulation or 7-8 bins spanning 291 to 850 nm for tropospheric simulations Tian et al., 2019.

Aerosol

The SO_4^{2-}-NO_3^--NH_3^+ aerosol chemistry is coupled to the gas phase chemistry (R. J. Park et al., 2004), with partitioning being calculated via the GEOS-Chem implementation of ISORROPIAII (Fountoukis and Nenes, 2007).

Direct emission of non-volatile organic aerosol (Pai et al., 2020) and black carbon (Q. Q. Wang et al., 2014) use a simple irreversible scheme whereby emissions are apportioned between a hydrophilic and hydrophobic component. The hydrophobic aerosol is "aged" to hydrophilic with an e-folding time of 1.15 days. Secondary organic aerosol (SOA) uses the simplified Volatility Basis Set (VBS) scheme (Pye et al., 2010). The isoprene gas phase chemistry is coupled to the aqueous phase isoprene SOA scheme.

Mineral dust aerosol in GEOS-Chem uses a scheme that combines the mineral dust component of the GOCART chemical transport model and the Dust Entrainment And Deposition (DEAD) scheme (Fairlie, Jacob, and R. J. Park, 2007). Mineral dust is split into seven dust bins based on the effective radius ranging from 0.15 to 4.0 μm, which is condensed to four species for advection.

Sea salt aerosol consists of two bins; accumulation phase (0.01 to 0.5 μm) and course phase sea salt (0.5 to 8 μm) (Alexander et al., 2005; Jaegle, Quinn, et al., 2011).

Aerosol optical depth (AOD) is calculated via a humidity dependent scheme (Ridley, Heald, and Ford, 2012). GEOS-Chem calculates the resultant scattering and thus changes in photolysis rates (Latimer and R. V. Martin, 2019).

Stratosphere

GEOS-Chem contains two mechanisms to represent the composition of the stratosphere. The Unified tropospheric-stratospheric chemistry extension (UCX) (Eastham, Weisenstein, and Barrett, 2014) is a fully coupled chemistry scheme representing the chemistry of both the troposphere and the stratosphere. There is also a linear ozone approximation algorithm (LinOz) (McLinden et al., 2000) which calculates ozone and other species concentrations based on monthly mean archived production and loss frequencies of significant stratospheric species (L. T. Murray, Jacob, et al., 2012). Given the nature of this thesis, the LinOz linearised chemistry is used as it is computationally faster, and little of the research focused on stratospheric issues.

Deposition

Wet deposition is included in GEOS-Chem for both aerosol and gases in the forms of rainout and washout from large-scale and convective precipitation and scavenging in convective updrafts (H. Y. Liu et al., 2001; Selin and Jacob, 2008; Amos et al., 2012). Further wet deposition via snow and cold precipitation is also included (Q. Wang et al., 2011).

Dry deposition velocities are calculated via a resistance-in-series model whereby the velocity is inverse to the aerodynamic, boundary and canopy resistance (Y. H. Wang, Jacob, and J. A. Logan, 1998). Additionally, the dry deposition velocity accounts for gravitational settling and turbulent dry transfer of particles to the surface (Fairlie, Jacob, and R. J. Park, 2007). For aerosol deposition onto snow and ice a fixed dry deposition rate (0.03 cm s^{-1}) is used based on eddy co-variance flux measurements (Fisher, Jacob, et al., 2011).

1.7 Summary

The chemistry of the atmosphere plays an important role in determining the magnitude of O_3 air pollution. This chemistry is complex and non-linear, and as such the interactions between its different components are challenging to understand, making optimal strategies for reducing air pollution difficult.

1.7.1 Thesis aims

1. The first aim is to increase the understanding of the photochemistry occurring in the summertime for Beijing in China. Beijing is a location of paramount interest for the understanding of O_3 chemistry due to its problematically high summer concentration coupled with a high population density. This problem will be tackled via process based improvements in the model whereby further chemical understanding will be added to GEOS-Chem. In Chapter 2, the GEOS-Chem model's ability to simulate the city's secondary gas phase chemistry is analysed. The cities' photochemical regime is diagnosed using a new day-by-day O_3 isopleth technique. This chapter will conclude by aiming to explore the sensitivity of O_3 pollution to a selection of VOCs.

2. The second aim is to use the improved Beijing model via the development of a new diagnostic/visualisation technique to understand better the influence of heterogeneous uptake of HO_2 in Beijing. Reactive HO_2 uptake has become a prominent area of study, mainly due to the unexpected recent rise in the Beijing summertime O_3 concentration. In Chapter 3, the influence of HO_2 uptake is explored and how it interacts with the current understanding of radical loss.

3. Improvements in representing atmospheric O_3 concentrations rely on improved scientific understanding, facilitating process-based improvements in the model. the last aim of this thesis is to develop an algorithmic approach to improve models focusing on performance over chemical understanding. The algorithm's behaviour will then be examined to see if it can help understand model shortcoming and thus aid fu-

ture process based improvements. In Chapter 4, a machine learning-based approach to enhance the predictions of O_3 concentrations by process-driven models such as GEOS-Chem is created and tested.

Chapter 2

Gas phase constraints on ozone production over Beijing

2.1 Introduction

In this chapter, a chemical transport model (GEOS-Chem), run regionally over eastern China, is used to explore the photo-chemical control of summertime surface ozone (O_3) in Beijing. The evaluation was possible due to a substantial field campaign that took place in Beijing in the summer of 2017. Funded by the U.K. Natural Environment Research Council (NERC) in partnership with the National Science Foundation of China (NSFC), UK Medical Research Council (MRC) and UK-China Innovation Newton Fund, the Atmospheric Pollution and Human Health in a Chinese Mega-city project (APHH-Beijing) provided a comprehensive assessment of the chemical state of the atmosphere over Beijing (see Section 2.3) for 5 weeks in the summer (Shi et al., 2019). There were also similar campaign in the winter months but this is not explored due to the lower photochemical activity.

In this section, the overarching aim is to evaluate and where possible, improve the chemical performance of GEOS-Chem over Beijing with a focus on O_3. O_3 will be the primary focus due to its negative impact on human and ecological health. Additionally, due to the complexity of the chemical interactions new ways of diagnosing and visualisation, the chemical state will be explored.

The urban nature of the chemistry in Beijing required the GEOS-Chem chemistry scheme to be improved beyond its standard (global focused) chemistry by the inclusion of aromatic (Section 2.4.1) and ethene chemistry (Section 2.4.2). An initial simulation using the default emissions is compared to the observations (Section 2.5) and the emission inventories are then tuned to match observations of primary emitted species (Section 2.6) with the results analysed (Section 2.7).

The model's ability to simulate secondary compounds (OH, OH reactivity, CH_2O, HO_2, RO_2, O_3) is then evaluated (Section 2.8). Systematic perturbation experiments are performed on the NO and VOC emissions in order to identify Beijing's photo-chemical O_3 production regime (VOC or NO_x limited) (Section 2.9). To investigate the variability in the photochemical regime O_3 isopleths are developed and examined (Section 2.10). The importance of different hydrocarbon classes (aromatics, alkenes, alkanes, and isoprene) for the production of O_3 is then explored (Section 2.11). Finally conclusions are drawn as to merits of either NOx or VOC reductions to reduce O_3 in Beijing (Section 2.12).

2.2 Ozone in China

O_3 is a hazardous pollutant that poses a risk to both human health and vegetation (T. Wang et al., 2017). As described in the Introduction (Chapter 1), many regions have issues with excessive O_3 concentrations. Over the last decade, the focus of global O_3 research has moved from countries in Europe and North America to China. With its large population (1.4 billion), densely populated eastern and southern regions, and rapid economic development, the emissions of air pollutants (S. X. Wang and Hao, 2012) has increased rapidly over the last decades in China, and so it now has a severe O_3 problem (B. Zheng et al., 2018).

High concentrations of O_3 are particularly prevalent in three mega-city regions; the Yangtze River Delta (Shanghai), the Pearl River Delta (Hong Kong) and "Jing-Jin-Ji", the Beijing metropolitan region or Capital Economic Zone (T. Wang et al., 2017). Each of these regions has a population of approximately 100 million people. If they were standalone countries, all three would rank in the global top 20 most populous countries. This study focuses on the Beijing region.

The Chinese government has passed legislation that has successfully targeted primary pollutants such as particulate matter (PM), sulphur dioxide (SO_2), carbon monoxide (CO) and nitrogen oxides (NO_x) (Feng and W. J. Liao, 2016). In 2012, O_3 was added to the Chinese air quality control standards. However, reducing of O_3 concentrations is more complicated than for primary species, due to its non-linear chemistry (Sillman, J. A. Logan, and S. C. Wofsy, 1990).

Efforts to reduce O_3 concentrations typically rest on two pillars: reducing the emissions of oxides of nitrogen (NO_x) and reducing the emissions of volatile organic compounds (VOCs). Most emphasis for reducing O_3 concentrations in western countries (Los Angeles, London etc) has gone into reducing NO_x emissions (Farber, Welsing, and Rozzi, 1994). However, the atmospheric chemistry for a city such as Beijing is likely very different and the so the role of VOCs needs to be considered fully.

Within a city, the VOCs most likely to cause ozone production are likely those which are relatively short lived, have a high Photo-chemical Ozone Creation Potential (POCP)

(Derwent, Jenkin, and Saunders, 1996), and are emitted in high concentrations. Four classes of compounds that full-fill those criteria for Beijing are longer alkanes (butane and larger), alkenes, aromatics, and biogenics (Q. Q. Li et al., 2020).

While all alkanes' oxidation leads to O_3 production, shorter alkanes have long lifetimes (on the order of years, months and weeks for methane, ethane and propane respectively) and are not of particular concern for urban regions as they diffuse over a large geographical area. Once the carbon number reaches four (butane) or greater, the lifetime is short enough that the compound becomes of urban significance (Q. Q. Li et al., 2020). Larger alkanes have POCP value ranging from 35 to 40 increasing with carbon length (Jenkin, Saunders, and Pilling, 1997). The POCP values of alkanes are appreciably lower than other compound classes explored in this section. Nonetheless, their higher concentration leads to their incorporation in this analysis.

Alkenes (hydrocarbons with a double bond) are a class of VOC compound with comparatively short lifetimes (on the order hours (R. Atkinson and Arey, 2003a)) with respect to OH due to the presence of the double bond. Ethene is the reference compound for POCP calculations (Saunders et al., 2003), with a POCP value, by default, of 100. This increases to 160 for propene due to additional stabilisation of the radical intermediate. Larger alkenes such as cis-2-butene have been found to have even higher POCP values above 200 (H. R. Cheng et al., 2010).

Aromatic compounds (specifically one ring "monoaromatic" compounds) are also predominantly anthropogenic (K. E. Ho et al., 2004). Benzene, toluene, xylene and trimethylbenzene (TMB) have POCP values of 3, 36, 97 and 180 (isomer averaged) respectively based on Chinese (Hong Kong) conditions (H. R. Cheng et al., 2010). Aromatic compounds' propensity to create O_3 is also related to their chemical decomposition products. After the initial OH oxidation, aromatic compounds break down into dicarbonyl species (such as $CH_3C(O)CHO$) via ring-opening reactions (C. Bloss et al., 2005). These dicarbonyls are highly photolabile, rapidly photolysing to form two peroxy radicals during the day. This contrasts to attack by the hydroxy radical which results in the production of only one peroxy radical (G. d. Silva, 2010). This rapid photolysis of dicarbonyls can, therefore, cause "radical amplification," rapidly increasing the concentration of peroxy radicals and thus O_3.

Alkenes and Aromatic are emitted predominantly from vehicle fuel, both from the exhaust and evaporation (Schuetzle et al., 1994). Chinese fuel standard V (H. H. Wei, 2019) (which was in place during APHH campaign) allowed for up to 24% and 40% of the volume of gasoline to be aromatics and alkenes respectively. The Chinese fuel standard V is comparable to the EU and American fuel standards (Y. H. Wang, R. Zheng, et al., 2016).

Biogenic compounds are another class of VOCs with a high propensity to make ozone. Emitted by the biosphere (trees, plants, algae, etc.), the class includes compounds such as alpha pinene ($C_{10}H_{16}$), limonene ($C_{10}H_{16}$), 2-methyl-3-butene-2-ol (C_5H_9OH) and the ubiquitous isoprene (C_5H_8). Emitted from a broad range of plant species isoprene represents the majority of biosphere VOC emissions (A. Guenther et al., 2006). Isoprene was calculated by H. R. Cheng et al., 2010 to have a POCP value of 171 (based on Pearl River Delta conditions). Isoprene's high POCP value is due to the fast rate of reaction of its double bonds with OH (daytime lifetime of approximately 1 hour) resulting in peroxy radical production occurring close to the emission source and thus in the presence of plenty of NO_x.

Assuming Beijing's O_3 shows some sensitivity to VOC emissions, it is likely that these four classes (alkanes, alkenes, aromatics, biogenics) of VOCs dominate its production. This chapter thus explores two questions: is Beijing currently in a NO_x or VOC sensitive O_3 regime, and which VOCs lead to the most O_3 production?

Answering these questions requires using a numerical model of Beijing's atmospheric composition, and more widely China's atmospheric chemistry. However, the model alone will not answer the question. The real-life emissions may have diverged from those estimated when the emission inventory was produced. Out of date inventories are particularly problematic in rapidly developing regions such as China. Further to this, inventories are not perfect, and many assumptions need to be made when developing the inventory which can later be found to be unreliable. Thus observations of the concentration of pollutants are needed to constrain the model to give results relevant to the actual conditions in Beijing experienced during the field campaign.

2.3 Beijing Observations

The NERC / Chinese funded Atmospheric Pollution and Human Health in a Chinese Megacity (APHH-Beijing) summer campaign (Shi et al., 2019) took place from May 21st to June 22nd 2017. It made extensive measurements of atmospheric composition and physics at the Institute of Atmospheric Physics (IAP) observational site in Beijing (39.976°N 116.377°E). These observations were designed to provide a complete set of observations of the composition of the atmosphere and are valuable for providing constraint on an atmospheric chemistry transport model investigating O_3 in Beijing.

Trace gases O_3, NO, NO_2 and SO_2 were detected by a set of commercial trace gas analysers (Shi et al., 2019). VOC measurements for compounds C_2 to C_7 were measured using dual-channel gas chromatography with a flame ionisation detector (DC-GC-FID) (Hopkins, C. E. Jones, and Lewis, 2011). Large mono-aromatic compounds such as xylene and trimethylbenzene (TMB) fell outside the detection range of the DC-GC-FID and thus concentrations were obtained via Proton Transfer Reaction Time-Of-Flight Mass Spectrometry (PTR-TOF-MS) (Huang et al., 2016).

OH, HO_2 and RO_2 measurements were made using fluorescence assay by gas expansion (FAGE) (L. K. Whalley et al., 2010). OH reactivity and CH_2O was measured using laser-induced fluorescence (LIF) (Cryer, 2016). Photolysis rates were measured via a spectral radiometer (Bohn et al., 2016). Organic and inorganic (NH_4^+, SO_4^{2-} and NO_3^-) aerosol were measured via aerosol mass spectroscopy (AMS) (Sun et al., 2016), while black carbon was measured via a centrifugal particle mass analyser coupled to a single-particle soot photometer (CPMA-SP2) (D. T. Liu et al., 2017). Particle number and size distribution were measured via a scanning mobility particle sizer (SMPS) spectrometer (Z. J. Wu et al., 2016). Temperature and humidity were measured via a meteorological station (Jenkins, 2014).

Table 2.1: Overview of measurements from the APHH campaign

Instrument	Measurements
Commercial trace gas analysers	O_3, NO, NO_2 and SO_2 concentrations
DC-GC-FID	C_2 to C_7 VOC concentrations
PTR-TOF-MS	C_8 and C_9 aromatic compound concentrations
FAGE	OH, HO_2 and RO_2 concentrations
LIF	OH reactivity and CH_2O concentration
Spectral radiometer	Photolysis rates
AMS	NH_4^+, SO_4^{2-} and NO_3^- aerosol concentrations
CPMA-SP2	Black carbon concentration
SMPS	Particle number and distribution
Meteorological station	Temperature and humidity

All observational data were re-sampled (mean) onto 10 minute time-steps to harmonise the observations with the model time-step. Observations (other than radical species) outside two standard deviations were removed to eliminate high concentration peaks which could represent very localised sources such as a vehicle passing close to the measurement apparatus.

These observations represent one of the most comprehensive set of observations of the photochemical conditions within an urban Chinese area currently available. With the observational dataset described, the model used for reproducing Beijing atmospheric conditions is outlined.

2.4 Model development

The model used for this analysis was GEOS-Chem (V12-8), run regionally in a nest over Eastern China with the "tropchem" chemistry scheme. With the FlexGrid adaption the model can be run over any grid for which GEOS-Chem compatible meteorological data is available and so a smaller grid than is usually run for China (J. Zhu and H. Liao, 2016) was used (Figure 2.1). GEOS Forward Processing (GEOS-FP) meteorology (https://gmao.gsfc.nasa.gov/GMAO_ was used as it provided the highest resolution currently compatible with GEOS-Chem (0.25° x 0.3125°). Boundary conditions for the regional simulation were generated from a global GEOS-Chem simulation run at 4° by 5° resolution. Spin up of the global model was run from the 1st of May 2016 to the 25th of May 2017.

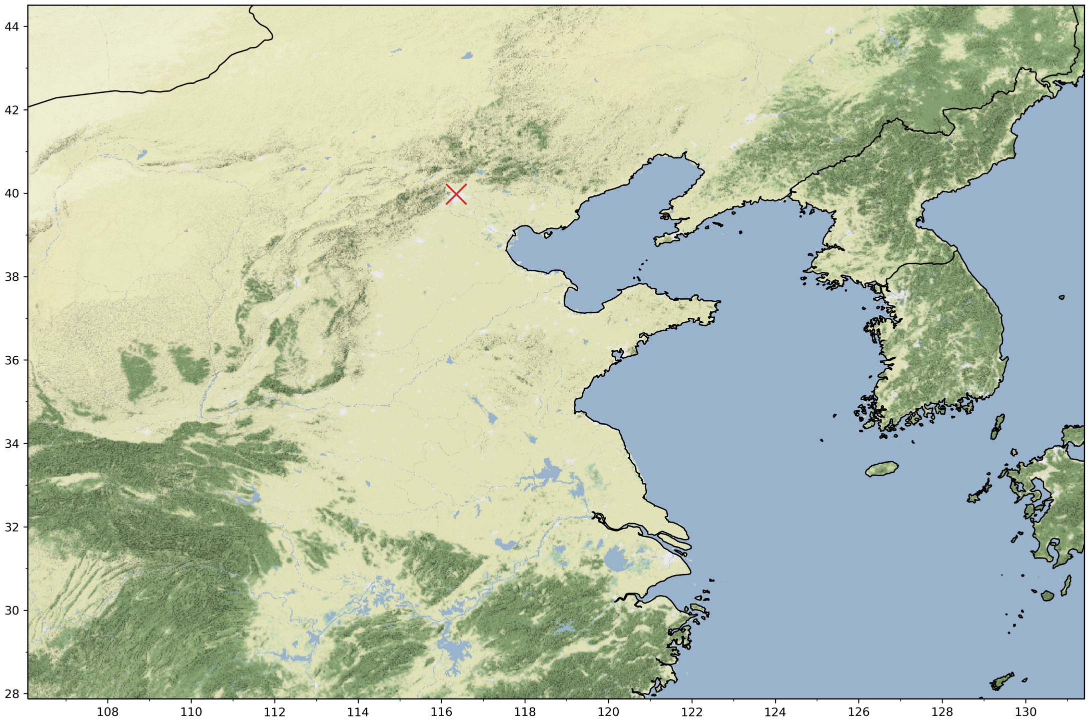

Figure 2.1: Domain of nested area used in this study. Red cross indicates the campaign location.

The standard chemistry scheme (https://github.com/geoschem/geos-chem/tree/master/KPP) used in GEOS-Chem has been developed with a global atmospheric chemistry focus. As such the chemistry development has been centred around long-lived species such as CH_4, alkanes and CO. Isoprene while short-lived has a substantial global emission and so is also included. The isoprene scheme was recently updated by Bates and Jacob, 2019.

Propene is also included in the scheme to represent all other short lived species. However, as identified earlier, for urban air pollution, aromatics and other alkenes play an important role in determining O_3 production. These are not in the standard chemistry of the model and so need to be added in.

2.4.1 Aromatic chemistry scheme

Benzene, toluene and xylene are included in GEOS-Chem V12.8, but only in a rudimentary form, to simulate the production of secondary organic aerosol (Henze, Seinfeld, Ng, et al., 2008). The model does not include any of their subsequent gas phase degradation chemistry that leads to the production of O_3.

The inclusion of an improved representation of aromatic chemistry scheme would thus seem necessary to model O_3 in urban environments. Adding additional chemical complexity, however, does increase the computation time it takes to run the model which could reduce its scientific usefulness. The chemistry time step is often the most substantial computational burden within chemical transport models (Long et al., 2015). Thus there is a trade-off between model complexity and the time it takes to run the model, adding too much complexity would make the model unusably slow.

Explicit or semi-explicit atmospheric chemical mechanisms such as the Master Chemical Mechanism (MCM) contain ~17,000 reactions (Sommariva et al., 2020) and are usually run within a single chemical box. This compares to the GEOS-Chem "tropchem" mechanism which is currently ~600 reactions. The 3-dimensional grid (latitude, longitude and level) used to simulate East China in this study consists of 247,455 boxes. Thus the inclusion of an explicit mechanism, such as the MCM, would lead to a large increase in computational burden. A "reduced" chemical mechanism aims to include a smaller number of reactions, balancing the need to represent the complexity of the chemistry with the need to keep the number of reactions to a minimum.

Two leading reduced regional atmospheric chemistry schemes are SAPRC (Carter, 1990) and RACMII (Goliff, W. R. Stockwell, and Lawson, 2013). SAPRC is frequently used for investigating aromatic compounds (Y. Z. Chen et al., 2015; Yan et al., 2019; Bey, Aumont, and Toupance, 2001) but implements "operators" (Carter, 2010) to handle radical production. This is inconsistent with the structure of GEOS-Chem, and so the RACM-II was chosen to be implemented.

The RACM-II's aromatic component (Goliff, W. R. Stockwell, and Lawson, 2013) consists of 91 reactions and 52 compounds. It achieves simplification by aggregating compounds with similar reactivity, structures and products together. Chemical species from RACM-II were mapped to existing GEOS-Chem species where possible, leading to 29 additional species. RACMII explicitly includes the degradation chemistry for benzene, toluene, o-xylene and a combined m and p-xylene species, but no higher carbon aromatics. Deposition of the hydrocarbon oxidation products followed the GEOS-Chem standard (H. Y. Liu et al., 2001; Y. H. Wang, Jacob, and J. A. Logan, 1998), with Henry's Law constants taken from Sander, 2015.

Trimethylbenzene (TMB) has a high POCP value (180) and has been identified as a possible significant contributor to Chinese O_3 production (Duan et al., 2008). Ambient summer concentrations of ~0.3 ppb have been previously measured in Beijing (B. Wang et al., 2010). Coupled with its high oxidation rate (~9 hour lifetime at $1x10^6$ molecules cm^{-3} OH), this would be indicative of a high rate of emission. A TMB specific emission inventory is also available, and observations of TMB were made during the campaign. Thus, finding a way of including this compound in the analysis seemed appropriate. However, TMB oxidation was not included in the RACMII chemistry scheme. To provide some analysis of its influence, a TMB tracer was included in the model with its initial oxidation by OH set at the MCM recommended rate. The resultant products were then made to be the same as the combined mp-xylene's degradation chemistry.

A large number of other aromatic compounds are emitted into the atmosphere. These include ethyl-benzene, styrene, 3,5-di-methyl ethyl-benzene, di-ethyl benzene etc. To represent these in the model an "other" aromatic tracer was included using ethylbenzene as their surrogate. Ethylbenzene was chosen as the surrogate species as it likely forms the most substantial proportion of the additional aromatics (Hampton et al., 1983). As with TMB, the MCM was used for a recommended ethylbenzene OH oxidation rate, and then this chemistry was also fed into the mp-xylene's degradation pathway.

2.4.2 Ethene chemistry scheme

Currently, alkenes are included in GEOS-Chem in the form of the single alkene species (labelled PRPE in GEOS-Chem). The species incorporates emissions of all alkenes with a carbon number of three or higher. Although consideration could be given to the explicit oxidation of the larger alkenes (butene etc), the most obvious compound missing from the model's oxidation scheme is ethene. Global emissions of anthropogenic ethene are approximately double that of propene (Rhew et al., 2017).

Ethene is the POCP reference compounds and has been identified as a possible substantial contributor to Chinese O_3 production (Q. Q. Li et al., 2020). In order to keep the new chemistry consistent with the new aromatic chemistry, the ethene mechanism from RACM-II were also incorporated into GEOS-Chem. As with the aromatic scheme, the species from RACM-II were mapped to existing GEOS-Chem species. The new scheme included 11 reactions and only necessitated the addition of two new chemical species (ethene itself and an ethene peroxy radical).

Table 2.2: Summary of new species added to GEOS-Chem for the improved aromatic and ethene schemes

GEOS-Chem Tracer	Species
DCB1	Unsaturated dicarbonyls $C_{4.5}$
DCB2	Unsaturated dicarbonyls C_7
DCB3	Unsaturated dicarbonyls C_4
BALD	Benzaldehyde and other aromatic aldehydes
BALPA	Peroxy radical from benzaldehyde oxidation
BALPB	Peroxy radical from benzaldehyde oxidation
BALPC	Peroxy radical from benzaldehyde oxidation
CSL	Cresol and other hydroxy substituded aromatics
CHO	Phenoxy readical formed from cresol
CSLOH	Aromatic-OH adduct from cresol
PHEN	Phenol
PHENNO3	Aromatic-NO3 adduct from phenol
MCT	Methyl catechol
TRO2A	Peroxy radical from toluene oxidation
TRO2B	Peroxy radical from toluene oxidation
TRO2C	Peroxy radical from toluene oxidation
PER1	Peroxy intermediate formed from toluene
PER2	Peroxy intermediate formed from toluene
EPX	Epoxide formed in toluene and xylene reactions
XYLO	o-Xylene
XRO2A	Peroxy radical from mp-xylene oxidation
XRO2B	Peroxy radical from mp-xylene oxidation
XRO2C	Peroxy radical from mp-xylene oxidation
XORO2A	Peroxy radical from o-Xylene oxidation
XORO2B	Peroxy radical from o-Xylene oxidation
XO2	Accounts for for addition NO to NO2 conversions
MCTO	Alkoxy radical formed from methyl catechol + HO and methyl catechol + NO3
MCTP	Alkoxy radical formed from methyl catechol + O3
TMB	Trimethylbenzene
AROM	Larger aromatics
C2H4	Ethene
EO2	Peroxy radical from C2H4
ETEG	Ethylene glycol

2.4.3 Emissions inventories

Emissions of primary compounds into the atmosphere are the driving force for the production of O_3 in the atmosphere. In this section the standard emissions configuration used in the model simulations are described and additions made to accommodate the new aromatic and alkene species described.

Anthropogenic

By default, GEOS-Chem uses the 2008 to 2010 MIX anthropogenic inventory (M. Li, Q. Zhang, et al., 2017) for the Asian region. This study continues to use the MIX inventory for NO, CO, NH_3 and SO_2. Despite the model being run for 2017, rather than 2010 (the last year of the MIX dataset), no additional scalings are used in the initial model simulations.

While MIX does include a total VOC emissions, no speciation for the VOC species is given. The CEDS global 2014 inventory (Hoesly et al., 2018) does provide emissions for individual VOCs including aromatic and alkene compounds, and so the CEDS inventory is used instead of the MIX inventory for the VOCs other than ethane and propane. Table 2.3 shows the mapping of CEDS to species in GEOS-Chem. CEDS is also used for the emission of black carbon and for primary organic carbon aerosol.

Table 2.3: CEDS species to GEOS-Chem tracer mapping list.

GEOS-Chem tracer	Compound name	CEDS species
ALK4	C4 Alkanes and larger	Butane Pentane Hexane and higher alkanes
C2H4	Ethene	Ethene
PRPE	C3 Alkenes and larger	Propene Other alkenes
BENZ	Benzene	Benzene
TOLU	Toluene	Toluene
XYLO	O-Xylene	Xylene x 0.33
XYLE	MP-Xylene	Xylene x 0.66
TMB	Trimethylbenzene	Trimethylbenzene
AROM	Other aromatics	Other aromatics
MOH	Methanol	Alcohol x 0.5
EOH	Ethanol	Alcohol x 0.375
ROH	C3 Alcohols and larger	Alcohol x 0.125
CH2O	Formaldehyde	Formaldehyde
ALD2	Acetaldehyde	Acetaldehyde
RCHO	C3 Aldehydes and larger	Larger aldehydes
BCPO	Hydrophobic Black carbon	Black carbon x 0.8
BCPI	Hydrophilic Black carbon	Black carbon x 0.2
OCPO	Hydrophobic Organic carbon	Organic carbon x 0.5
OCPI	Hydrophilic Organic carbon	Organic carbon x 0.5

Ethane and propane emissions still use the GEOS-Chem default Tzonmpa-Sosa (Tzompa-Sosa et al., 2017) and Xiao (Y. P. Xiao et al., 2008) inventories respectively .

Biomass burning

The Global Fire Emission Database (GFED4) is the default biomass burning emission inventory in GEOS-Chem (Giglio, Randerson, and Werf, 2013). GFED4 is implemented into GEOS-Chem in the form of an extension module rather than direct emission files. It uses input of the burnt area and tree type to calculate speciation and emission rate. This makes changing the emissions cumbersome as any scaling would need to be hardcoded into GEOS-Chem. Instead, the Global Fire Assimilation System (GFAS) was chosen for biomass burning emissions (Kaiser et al., 2012). GFAS is an offline inventory, with speciation and emission rate pre-calculated. This makes scaling the GFAS emissions straightforwards should it be necessary.

Biogenic

The Model of Emissions of Gases and Aerosols from Nature (MEGAN) inventory was used for biogenic emissions (A. B. Guenther et al., 2012). As with GFED4, MEGAN is an extension using global base emissions that are scaled based on a variety of variables including leaf area, temperature, soil wetness and solar flux. NASA Moderate Resolution Imaging Spectroradiometer (MODIS) data was used for leaf area index. There is no viable alternative to MEGAN for GEOS-Chem and as such perturbation studies were performed via alterations to the MEGAN source code.

Now the updates to the model chemistry scheme and emission inventories have been explained, a comparison between the modelled physical parameters and those observed is first undertaken and then a comparison between the composition calculated by the model in this "initial" model simulation is described.

2.5 Initial simulations

An initial simulation was performed to evaluate the performance of the model in its standard configuration. The model was run from May 18th to June 25th, with the first week considered spin-up, with the initial conditions taken from an up-scaling of the boundary condition run. The model run took approximately 30 hours to run using 20 CPU cores.

An initial comparison is made with the model's ability to simulate the physical conditions at the Beijing site, and then a comparison is made with the observed chemical composition of the air.

2.5.1 Physical parameters

Correctly representing the meteorological parameters is critical for getting the chemical transportation correct however, the physical parameter are also important for the chemistry. Tropospheric reactions that govern secondary organic compounds' production are temperature, solar radiation or humidity dependent. Therefore it is essential that before evaluating other aspects of model performance, the model's representation of these physical conditions is first evaluated. The time series and mean diurnal for the temperature, $J(NO_2)$, $J(O_3\text{->}O(^1D))$ and Relative humidity for the period the campaign can be seen in Figs. 2.2 and 2.3 respectively.

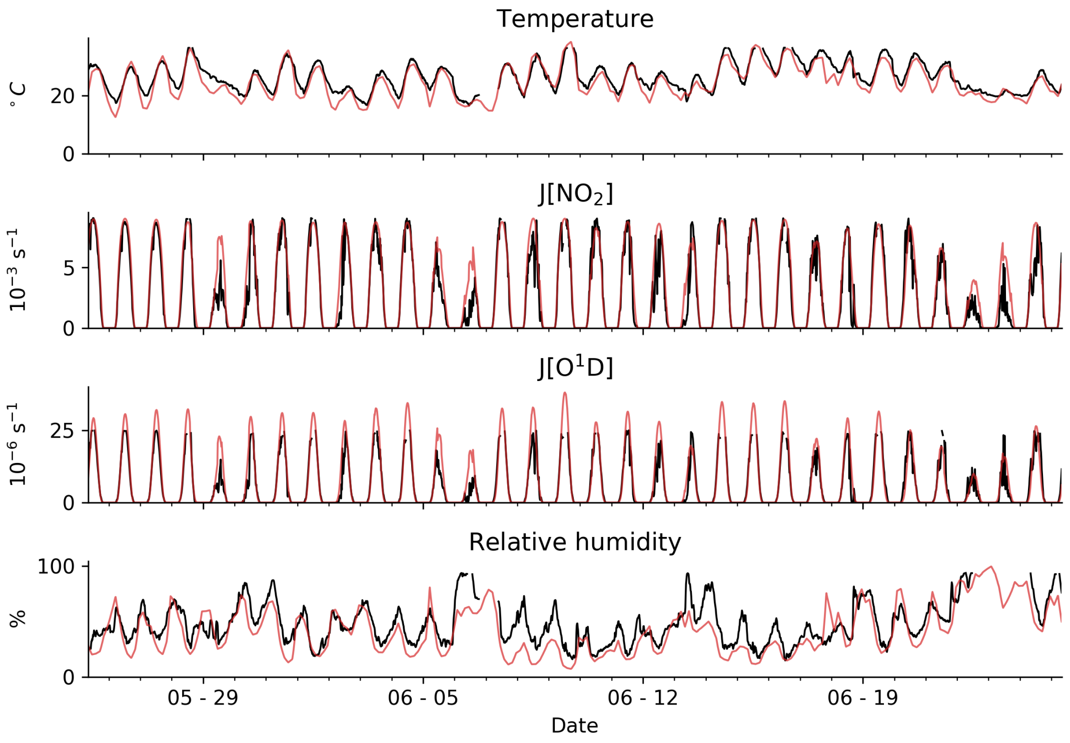

Figure 2.2: Time series of modelled (red) and observed (black) physical variables at the Institute of Atmospheric Physics (IAP)

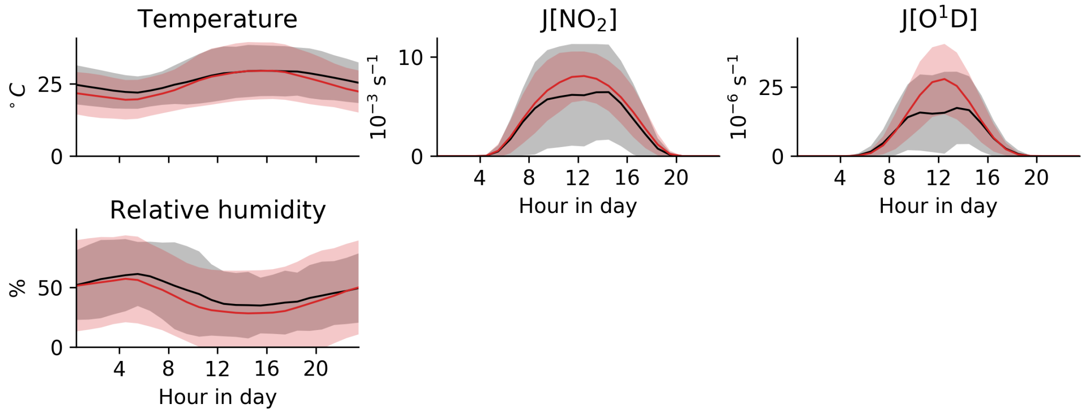

Figure 2.3: Mean diurnal variations of modelled (red) and observed (black) physical variables. Shaded area indicates one standard deviation.

Temperature in the model is provided by the meteorological fields (GEOS-FP) and is in good agreement with observations. Daytime diurnal mean temperature are within 0.5°C of the observation. There does though appear to be a more substantial underestimation of the temperatures at night. This may reflect local conditions, compared to the more regional conditions indicated by the 25 km model grid. It is not surprising that the agreement here is relatively good as the GEOS-FP system assimilates observed temperatures into its products.

The j values are the first order rate of photolysis reactions and reflect a combination of the solar radiation falling on the top of the atmosphere, the concentration of absorbing and scattering elements within the atmosphere and the photochemical properties of the reacting compound. As such, there is potential for uncertainty, including quantum yield (Wuerth et al., 2011), solar flux (Shetter et al., 2003), the surface reflection (Y. Zhang, Z. Q. Li, and Macke, 2002) and scattering by both clouds and aerosol (Latimer and R. V. Martin, 2019). $j(NO_2)$ (NO_2 + hν -> NO + O) and $jO(^1D)$ (O_3 ->O^1D) represent two critical drivers of photochemistry in the atmosphere.

The noontime average $j(NO_2)$ values are overestimated by 15%. However this bias predominately arise from an overestimation on a few low jNO_2 days (May 29^{th}, June 6^{th} and June 22^{nd}). For the rest of the campaign the model performs well. This could reflect problems in the GEOS-FP meteorology in replicating cloud opacity, or in the simulation of aerosol scattering. Further work is necessary to evaluate why the model fails on these days.

The $jO(^1D)$ bias appears to be larger and more consistent than found for jNO_2, with a mean noon time overestimation of 45%. This overestimate occurs on most days. A number of explanations might exist for this. Aerosols are essential for light scattering. The model appears to over estimate the concentration of aerosol mass but underestimates aerosol surface area. Thus the model may also be under-representing scattering. The single bin, mass based aerosol scheme used in these simulations makes improving these calculations difficult. S. Li et al., 2019 show that GEOS-Chem's Aerosol Optical Depth is biased low over east China in GEOS-Chem, which would support this. $jO(^1D)$ can also be biased due to an incorrect O_3 column value, as this would increase or decrease absorption before the light reaches Beijing.

A number of different humidity parameters could be evaluated here (absolute humidity, relative humidity, vapour pressure etc). What is provided from the GEOS-FP assimilation is water mass mixing ratio (kg moisture kg^{-1} air), and it is this which is used for calculating gas phase reaction rates. However, for aerosol properties such as surface area and radius, the relative humidity is important, which is derived from water mass mixing ratio and temperature. The diurnal mean is similar to that observed but is biased low by 8% during the day.

2.5.2 Initial comparison to chemical observations

Figure 2.4 shows the comparison of the initial model to the campaign observations for some key species. The initial model performance was poor. NO_x, toluene, SO_2, and black carbon show varying degrees of over estimation, whereas ethane, propane, TMB, >C3 alkenes and isoprene are all underestimated. Previous studies have examined the correctness of the emissions inventories for Beijing (M. Li, H. Liu, et al., 2017; Deng et al., 2020; Jing et al., 2016; M. Wang et al., 2014). However, given the initial level of model performance, the model's utility for science would likely be questionable. Conclusions on the sensitivity of O_3 over Beijing to VOCs or NO_x using the initial model, would not be grounded in a realistic representation of the chemical state.

The emphasis of this study is on the sensitivity of O_3 to its precursor emissions rather than understanding the magnitude of the emissions themselves. Therefore the model emissions are "tuned" to fit the observation of the primary species and so provide better reproduction of the actual atmosphere in Beijing for the duration of the campaign.

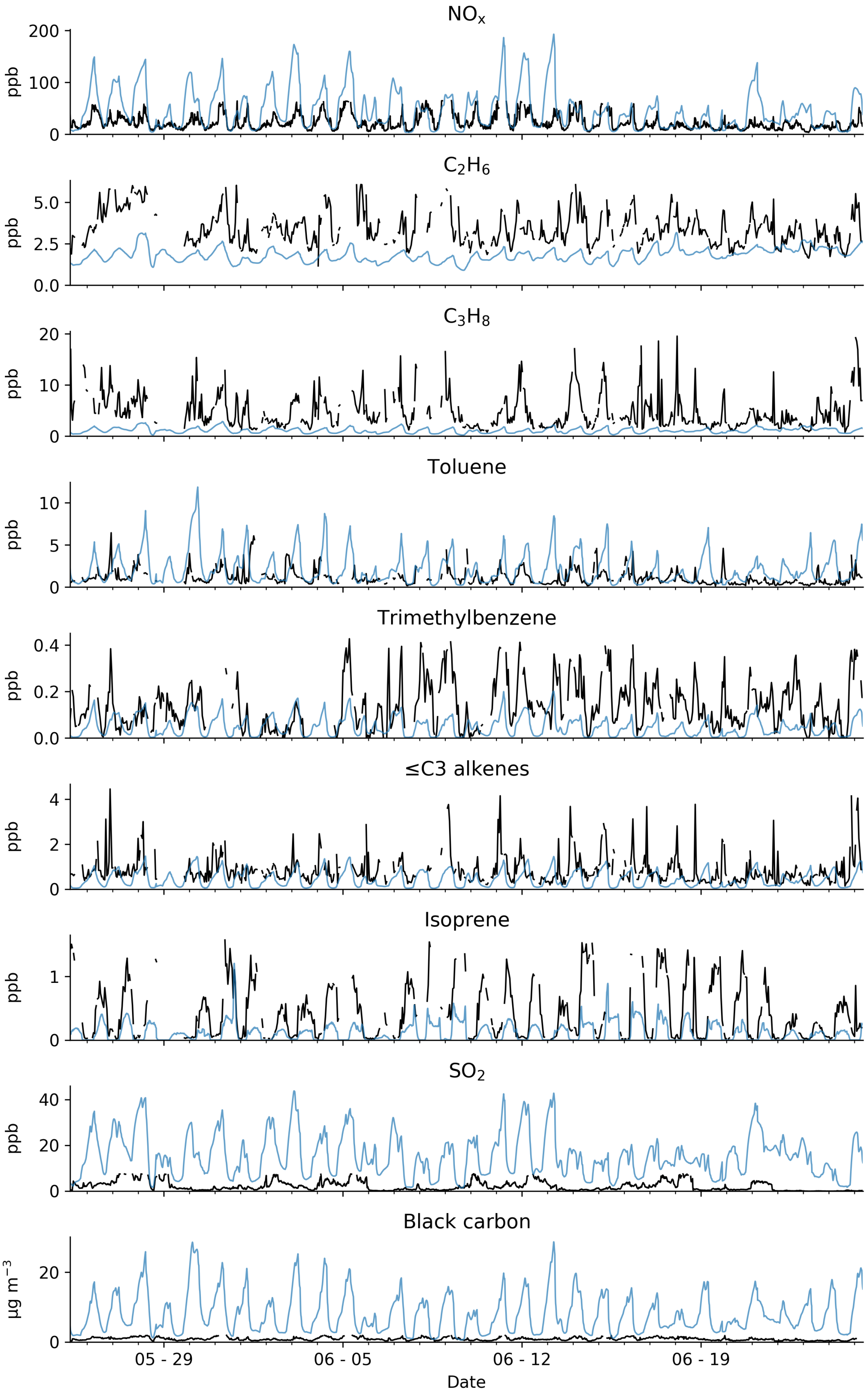

Figure 2.4: Time series of initial model (blue) and observed (black) primary species.

2.6 Primary emission tuning

Tuning the model to fit the observations could be achieved in many ways. It is possible to formally "invert" the emissions from the observations (Henze, Seinfeld, and Shindell, 2009; L. Zhu et al., 2013; Lee et al., 2011) or to "nudge" the model towards observations through data assimilation (Mertens et al., 2020; Astitha et al., 2012). Both of these are sophisticated approaches and are subject to significant ongoing research. A more straightforward approach was adopted here. The basis for this tuning was the mean diurnal cycle of the observations during the campaign. The objective was to get the model to fit the observations as tightly as possible by varying the anthropogenic and biogenic emissions within the model domain.

The model showed little sensitivity from varying the biomass burning emissions. Similarly, soil and lighting NO_x sources are small compared to the anthropogenic sourced within the domain. The two inventories that were therefore tuned were the anthropogenic emission and the biogenic emissions. Changes to boundary conditions were found to have little effect for most species but did show some sensitivity for certain long-lived species (CO, C_2H_6 and O_3). By only focusing on emissions, all differences between the model and measurement are attributed to failures in the emissions, model sinks are correctly simulated. The model's ability to simulate the OH radical is discussed in Section 2.8.4 but model shows a systematic underestimate in the OH concentration by around 30%, thus the corrections made here are likely on the conservative side.

The other critical component is the transport term which mixes emitted compounds out of model's surface box both horizontally and vertically. Significant uncertainties lie in the ability of models to simulate this correctly (Schuh et al., 2019), notably in the description of the boundary layer height. However, given that the emphasis of this chapter is on the concentration of secondary compounds, this is not explored further. However, this does mean that some of the emission correction could be a result of meteorological bias (likely boundary layer height) rather than purely emission bias. Further research would be required to separate these biases. As a result of this approach, the tuning of the emissions will combine errors in the emissions, sink processes and meteorology.

Anthropogenic emissions

In this class were the emissions of alkanes (C_2H_6, C_3H_8, ALK4 (alkanes with 4 or greater carbons)), alkenes (C_2H_4, PRPE (alkenes with 3 or greater carbons), aromatics (benzene, toluene, mp-xylene, o-xylene, trimethylbenzene, other aromatics), alcohols (methanol, ethanol and larger alcohols), CO, NO, SO_2, NH_3, black carbon, organic carbon.For each of these species the same scaling was applied over the whole model domain.

This domain level scaling was intended to reflect national level changes and errors in the emissions. Some of the changes in industrial activity in Beijing, will correlate with those in Shanghai. Similarly, policies to reduce emissions in Beijing should likely also apply in Shanghai. The approach is not perfect. The industrial development of western China has been faster than in the East (Crane et al., 2018), and there have been different regulatory frameworks applied in different regions. However, given the lifetime of some of the species (CO, O_3, aerosol) scaling the emissions at a national level seemed appropriate.

The model was tuned to simulate observed NO_x (NO + NO_2) concentration rather than the individual NO and NO_2 species. It was not possible to tune NO and NO_2 and achieve the correct O_3 concentration.

SO_2 and NH_3 were tuned to give the best agreement with the observed aerosol SO_4^{2-} and NH_4^+ mass concentrations respectively as the aerosol mass and surface area were considered more important than replicating their gas phase components.

Isoprene

Isoprene emissions are highly variable (Dylan B. Millet et al., 2008), thus tuning the whole of domain based on observations from Beijing would seem inappropriate. Forested regions are not likely to respond in the same way as urban regions. Instead, the emissions for the Beijing urban region were tuned, leaving the rest of the domain with the default MEGAN emissions. Figure 2.5 shows the area of the model that is scaled. The urban area was restricted to just the Beijing City region as this area would be expected to be reasonably homogeneous from a biogenic perspective. Only the base emission was scaled, MEGAN then calculates the diurnal cycle and day to day variability based on meteorological conditions.

Figure 2.5: Map of Beijing area. Black box shows the area over which isoprene emissions were scaled. Red cross indicates the location of the campaign site. Grey squares shows the 0.25° model grid boxes.

2.6.1 Tuning methodology

The chemistry of the atmosphere is non-linear. Changing the emissions of one compound does not necessarily lead to a linear change in the concentration of that compound, and it can also lead to changes in the concentration of other compounds. Thus it is non-trivial to find a set of emissions that fit all of the observations.

To simplify the analysis, the tuning focused initially on the mean concentration calculated by the model over the period of the campaign. Once the mean values were approximately correct attention turned to the campaign mean diurnal cycle.

For the initial comparison between the modelled and measured campaign means, the diurnal cycle applied to the model emissions was removed. The model was then run multiple times, updating the scaling of NO_x, isoprene and aromatic compounds, until the campaign mean value of those concentrations were comparable to the observations. This iterative approach was considered complete when the simulated mean matched within +/- 5 % of the observed mean. Once the threshold agreement had been reached, the mean emission of the remaining VOCs and CO were tuned. Finally, the emission of primary aerosol species (black carbon and organic carbon) and aerosol precursor species (NH_3 and SO_2) were tuned.

Once the campaign mean model concentrations were comparable to those observed, the diurnal emissions profile were tuned to fit the observations. Isoprene diurnal emissions were not tuned, as their diurnal emissions profile is set by the MEGAN Inventory. The tuning of the diurnal profiles was performed in a similar species order to the means, with initial emphasis focusing on NO_x and the aromatics. The same diurnal profile was used for all of the species within the alcohols, aromatics and alkenes class.

Once tuned the model showed much better agreement between the model and the measurements.

2.7 Tuning Results

Figure 2.6 shows the mean calculated diurnal cycle for the concentration for the primary gas species before and after tuning. After tuning the diurnal mean was close to the observation mean for most species, diurnal mean values can be seen in Table 2.7.

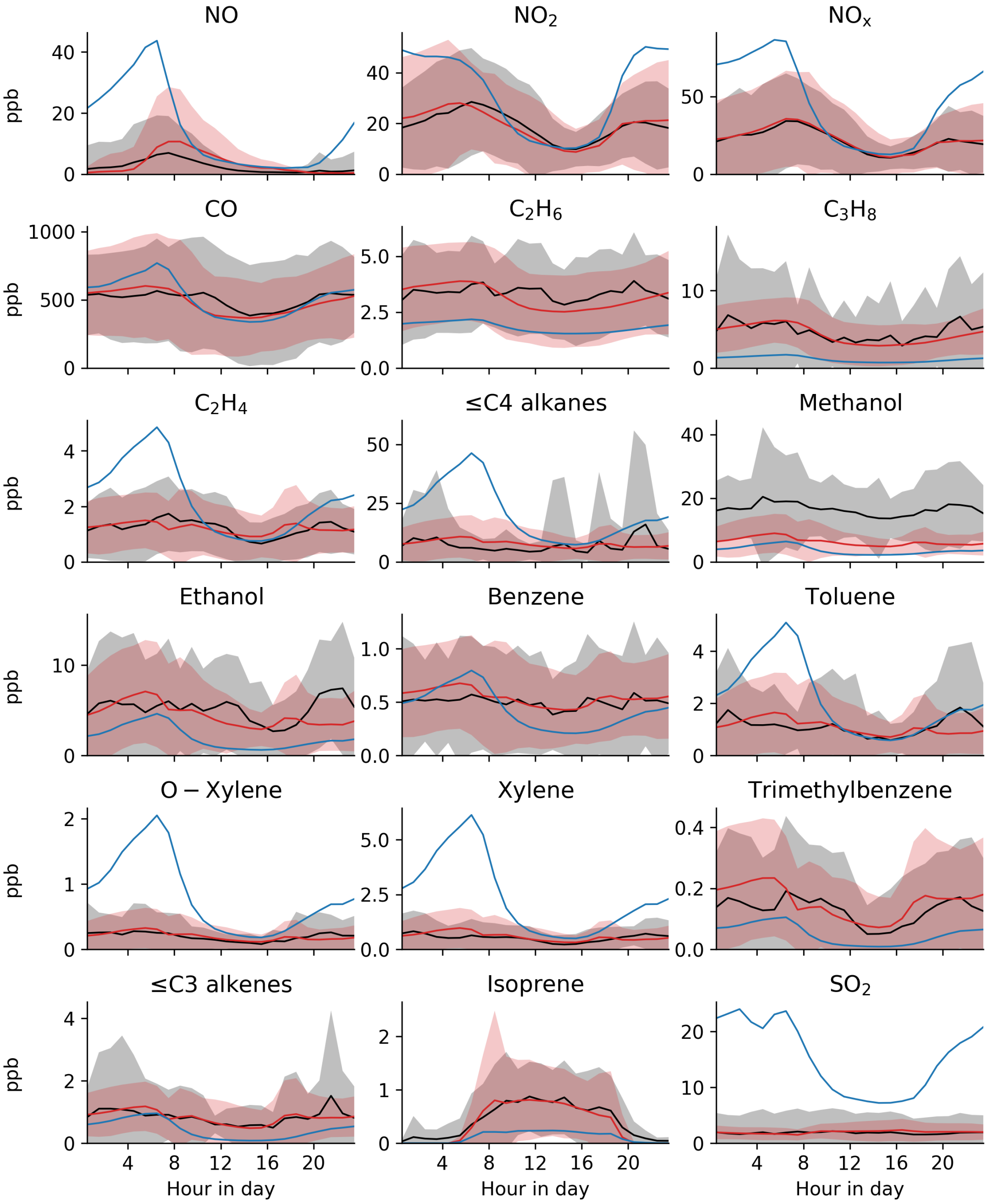

Figure 2.6: Mean hourly diurnals of the untuned model (blue), tuned model (red) and observed (black) for the gas phase primary species. Shaded area indicates one standard deviation in the hourly mean.

Compound	Obs mean	Model mean	Obs noon	Model noon	Obs midnight	Model midnight
NO_x / ppb	22.27	22.43	16.47	17.18	21.04	22.53
CO / ppb	504.84	482.89	473.16	376.57	536.30	550.08
Ethane / ppb	3.37	3.16	3.41	2.58	3.14	3.52
Propane / ppb	4.70	4.31	3.35	3.04	5.21	5.01
Ethene / ppb	1.21	1.21	1.19	1.03	1.14	1.24
Alkanes C4 + / ppb	6.99	7.68	5.40	6.62	7.29	7.60
Methanol / ppb	16.50	6.31	15.38	5.20	15.89	6.37
Ethanol / ppb	5.24	4.47	5.76	3.53	4.65	4.50
Benzene/ ppb	0.50	0.54	0.48	0.45	0.52	0.58
Toluene / ppb	1.11	1.09	0.91	0.90	1.28	1.08
O − Xylene / ppb	0.20	0.20	0.12	0.15	0.26	0.21
Xylene / ppb	0.53	0.58	0.33	0.40	0.77	0.62
Trimethylbenzene / ppb	0.14	0.15	0.09	0.09	0.15	0.20
Alkenes C3 + / ppb	0.84	0.83	0.58	0.57	0.89	0.92
Isoprene / ppb	0.42	0.37	0.81	0.79	0.06	0.00

Table 2.4: Mean diurnal values for the modelled and observed primary gas phase species.

The scale factor applied to the inventories after tuning can be seen in Table 2.5.

Table 2.5: Summary of primary compound scale factors on the daily mean emission rate.

Species	Inventory	Mean emission factor
NO	2010 MIX	0.87
CO	2010 MIX	1.08
Ethane	2008 Tzompa-Sosa	3.14
Propane	2014 XIAO	5.39
Alkanes C$\geq$4	2014 CEDS	0.67
Ethene	2014 CEDS	1.07
Alkenes C$\geq$3	2014 CEDS	7.13
Alcohols	2014 CEDS	5.03
Benzene	2014 CEDS	2.09
Toluene	2014 CEDS	1.00
Xylene	2014 CEDS	0.50
TMB	2014 CEDS	6.70
Isoprene	2014 CEDS	3.50

Table 2.6: Inventory and emission scale factors for the primary gas phase species

NO required a 0.87 mean scale factor, resulting in a mean emission rate of 4.55 mg m^{-2} h^{-1}. The emission flux is higher than the 3.55 mg m^{-2} h^{-1} flux calculated by Squires et al., 2020 but falls well within the one standard deviation of that calculation (3.69 mg m^{-2} h^{-1}). A 13% reduction in NO emissions between the MIX inventory year (2010) and the campaign year (2017) is close to the -17% predicted trend in Beijing (B. Zheng et al., 2018).

Ethane and propane emissions required large scale factors of 3.14 and 5.39, respectively. Alkane emission inventories have been observed to severely underestimate (by up to a factor of eight) emissions in Chinese urban centres (Z. W. Mo et al., 2018). These missing emissions are believed to be due to difficulties in quantifying fugitive fuel emissions (Sillman, J. A. Logan, and S. C. Wofsy, 1990). Butane and larger alkanes (ALK4) required a 33% reduction. The reason the CEDS over estimates ALK4 is unclear.

CEDS performs well on new ethene species requiring only a 7% increase in emissions. The combined larger alkenes (PRPE), however, required the largest scale factor (7.13) of all the primary species investigated. This represents a significant underestimation by the CEDS inventory and may represent errors in the assumptions made about the composition of Chinese fuel in the inventory.

The scale factor required by the aromatics varies between species, indicating a poor understanding of aromatic emissions activity (Watson, Chow, and Fujita, 2001). The emission flux of benzene (0.64 nmol m^{-2} s^{-1}), toluene (2.64 m^{-2} s^{-1}) and xylene (1.65 m^{-2} s^{-1}) were all within one standard deviation of the mean fluxes (2.00, 1.94 and 2.07 nmol m^{-2} s^{-1} respectively) observed during the campaign (Acton et al., 2020). Similar to PRPE, TMB required a large 6.7 scale factor. This again may represent errors in the assumptions about the composition of Chinese fuel in CEDS inventory.

Alcohols required a 5.03 scale factor on emissions. Methanol was the only primary species that could not be readily fitted to the observation. The measured methanol flux during the campaign was 8.0 nmol m^{-1} s^{-1} (Acton et al., 2020), while the model emission flux was 11.0 nmol m^{-1} s^{-1} and still underestimated concentrations (Figure 2.6). To get agreement between the measured and modelled methanol by increasing the emissions would have require a substantial increase in the emissions beyond the observational constraint. It seems likely that much of unaccounted for methanol is secondary in origin and the model underestimates could represent a failure in the chemistry, rather than in the emissions.

Isoprene required a significant 3.5 scale factor on the base MEGAN emissions to match observation. However, this scaling resulted in a mean emission rate of 6.9 nmol m^{-2} s^{-1} close to the measured value of 7.6 nmol m^{-2} s^{-1} (Acton et al., 2020). The substantial scale factor implies there is a fundamental error associated with the parameterisation of isoprene emissions in Beijing. A previous study in southern China (Situ et al., 2014) also found a large (110%) bias (albeit still smaller than the bias found in this study) however, the 2014 study did not focus specifically on the urban areas. The underestimation could be a result of incorrect leaf area index, speciation of flora or incorrect emission rates for the given flora. The exact cause falls outside the scope of this project but considering the high fraction of the OH reactivity due to isoprene in Beijing further work on Chinese isoprene emissions appears essential.

Figure 2.7 shows the mean calculated diurnal cycle for the concentration of the individual aerosol species and the total aerosol surface area before and after tuning. After the model tuning, performance for all compounds and the aerosol surface area has significantly improved. The mean daily, noon and midnight values are shown in Table 2.7, and the necessary scalings given in Table 2.8.

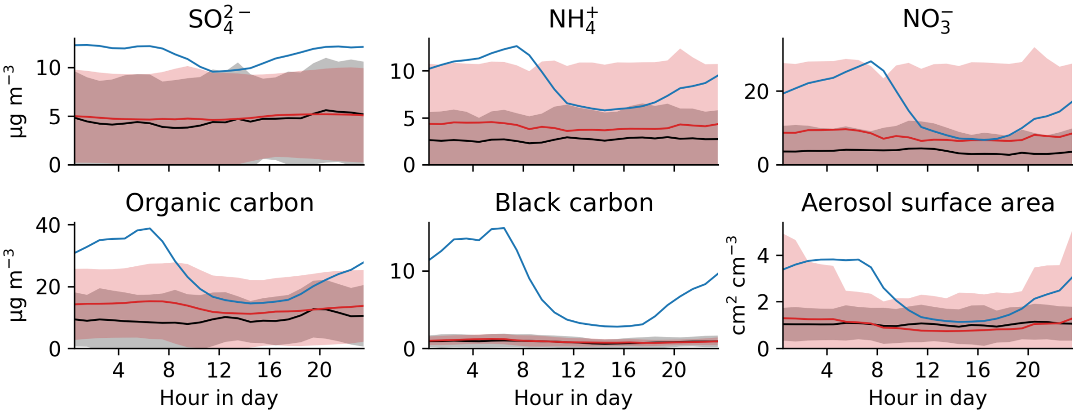

Figure 2.7: Mean diurnals of the tuned model (red) and observed (black) for the aerosol species and aerosol surface area. Shaded area indicates one standard deviation.

Compound	Obs mean	Model mean	Obs noon	Model noon	Obs midnight	Model midnight
SO_2 / ppb	1.93	1.98	1.82	2.15	1.95	1.94
SO_4 / ug m^{-3}	4.63	4.87	4.50	4.64	4.80	4.98
NH_4 / ug m^{-3}	2.69	4.07	2.81	3.70	2.60	4.34
Inorganic nitrate / ug m^{-3}	3.64	7.78	4.34	6.79	3.56	8.69
Organic aerosol / ug m^{-3}	9.67	13.25	10.26	11.44	9.77	14.22
Black carbon / ug m^{-3}	0.85	0.90	0.75	0.77	0.94	0.98
Aerosol surface area / cm^2 cm−3	1.04	0.92	1.04	0.69	1.04	1.22

Table 2.7: Mean diurnal values for the modelled and observed aerosol species and aerosol surface area.

Table 2.8: Summary of primary compound scale factors applied to the daily mean emission rate.

Species	Inventory	Mean emission factor
SO_2	2010 MIX	0.22
NH_3	2010 MIX	0.76
Organic carbon	2014 CEDS	0.38
Black carbon	2014 CEDS	0.14

Table 2.9: Inventory and emission scale factors for the aerosol and aerosol precursor species.

SO_2 required considerable 78% reduction in order to replicate both the SO_2 and the sulphate concentrations. This large reduction is similar to the 62% published (B. Zheng et al., 2018) reductions in SO_2 emissions in China since the inventory was produced (2010).

Ammonia emissions were reduced by 24% in order to replicate the observed Ammonium diurnal. The reduction is close to the 31% NH_3 bias previously calculated in the MIX emission inventory (Q. Q. Zhang et al., 2019).

Organic carbon required a 62% reduction in direct emissions. Distinguishing between primary and secondary organic aerosol (SOA) is problematic, and as such, it is hard to identify how much of the reduction is a result of excessive SOA production within the model, and how much is due to errors in the emissions inventory. Nonetheless, primary organic carbon emissions have been reducing in China in recent years (Lang et al., 2017).

Black carbon had the most considerable reduction of all species observed at 86%. Black carbon emissions have been reducing in Beijing (Y. Chen et al., 2016). However, the magnitude of the reductions found here appear in excess of the published values (38%) and would suggest a possible issue within the CEDS emission inventory.

2.7.1 Diurnal profile

As well as changes to the mean emissions during the campaign, adjustments were made to the diurnal emissions profiles, so that the modelled mean diurnal cycle fitted the observed diurnal cycle.

Figure 2.8 shows the emissions profiles used for NO emissions compared to the default provided by the Global Emissions InitiAtive (GEIA) inventory (Stavrakou et al., 2009). The diurnal profile has a very large peak in the morning and then a smaller secondary peak in the afternoon. The shape and magnitude of the day/night difference are similar to that calculated by Squires et al., 2020, albeit with a smaller afternoon peak. The calculated diurnal is substantially more drastic than the default Beijing NO emission profile. This likely reflects the overwhelming role of transport on NO emissions in Beijing. This may have changed over the years as industrial and emissions sources are moved out of the city to reduce human exposure. Future work would be required to separate the diurnal scale factors into sector-specific factors. However, this could result in an even greater transportation diurnal being required once a relatively flat diurnal from power generation is applied as such this may elude to meteorology exaggerating this diurnal somewhat.

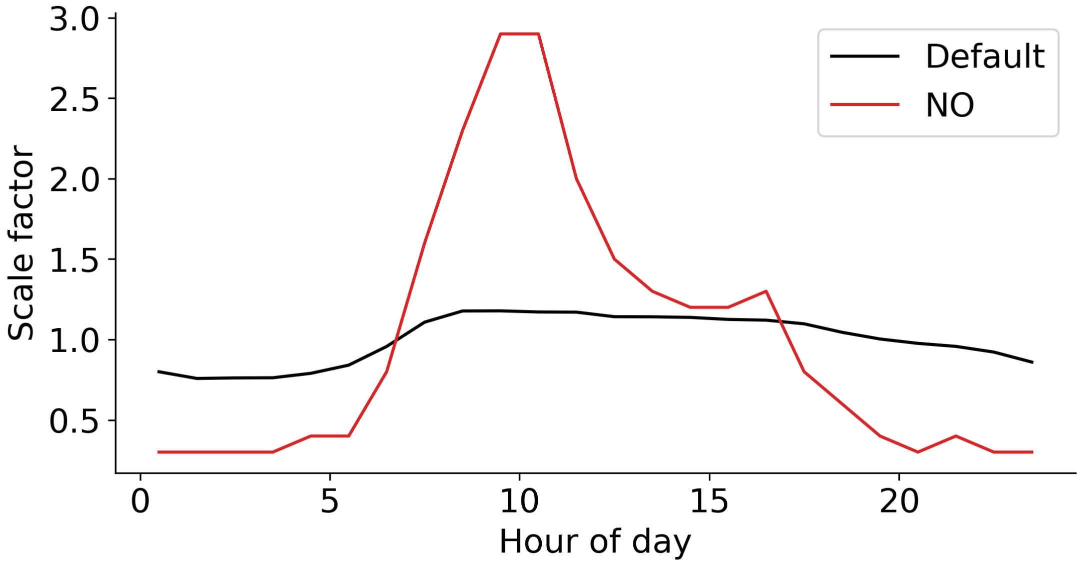

Figure 2.8: Default and tuned diurnal scale factor for NO emissions.

By default GEOS-Chem uses the same diurnal profile for all anthropogenic VOCs and for CO. It was found that the fossil fuel species required splitting into three categories: long lived species (CO, ethane and propane), short lived species (aromatic compounds, alcohols and large alkanes (ALK4)), and alkenes (Figure 2.9).

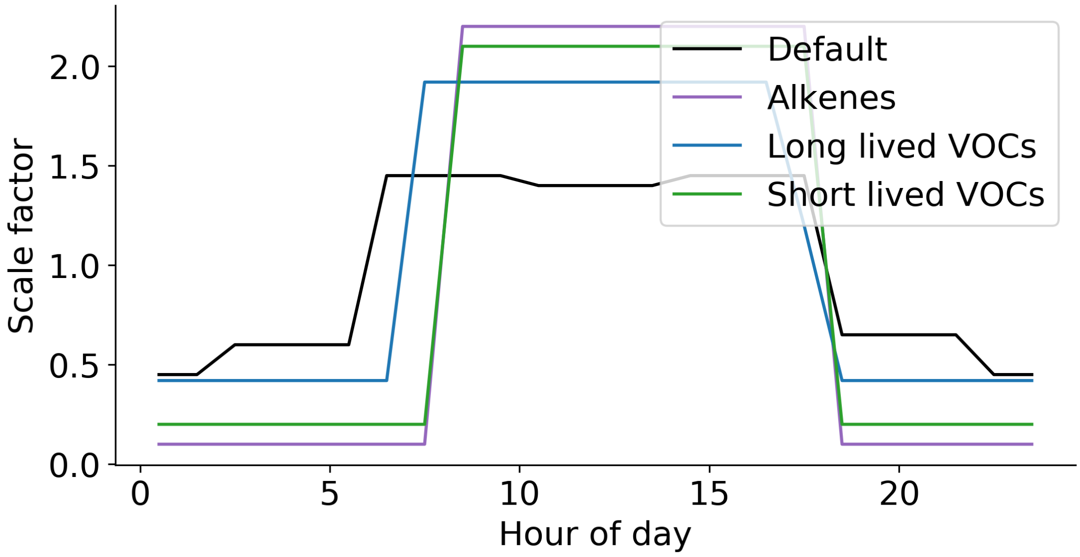

Figure 2.9: Default and tuned diurnal scale factor for VOC emissions.

All three profiles look relatively similar, with a "top hat" style profile, roughly following day and night. The shorter lived the compound, the bigger the jump between day and night time values. This may reflect some of the other model failures. OH concentrations at night are essentially zero in the model, but are measurable in the observations. Thus increased night time concentration in the model may allow a single diurnal profile to be applied to the VOCs and to maintain model fidelity.

Tuning the model allowed there to be a "reasonable" comparison between the observed concentrations of predominantly primary compounds and those simulated in the model. The exact reasons for these differences have not been explored fully, although some indications suggest changes in the emissions between the year they were generated for (2010) and the year of the campaign (2017), and issues with the concentrations OH calculated by the model. Although not perfect the model agreement for the primary species is subjectively now good enough for the model's ability to simulate secondary species to be evaluated.

2.8 Model performance

The focus of this chapter is on the sensitivity of secondary compounds (ultimately O_3) to potential policy driven changes in the emissions of primary compounds such as NO_x and VOCs. Now that the modelled concentration of these species match (to some extent) the observations, the performance of the tuned model to simulate secondary compound concentration is evaluated. As the model was tuned to the NO_x rather than the individual NO_x species, the modelled NO and NO_2 concentrations are assessed followed by HONO. Then the model's ability to simulate formaldehyde is assessed. The model's ability to simulate the RO_x family members (OH, HO_2, RO_2) and OH reactivity is then evaluated. Finally, the model performance for O_3 is analysed.

2.8.1 NO_x species

The model NO emissions were tuned to fit the observed NO_x concentration, but the concentrations of the individual components (NO and NO_2) were not constrained. Figures 2.10 and 2.11 shows the comparisons between the model and measurement for NO, NO_2.

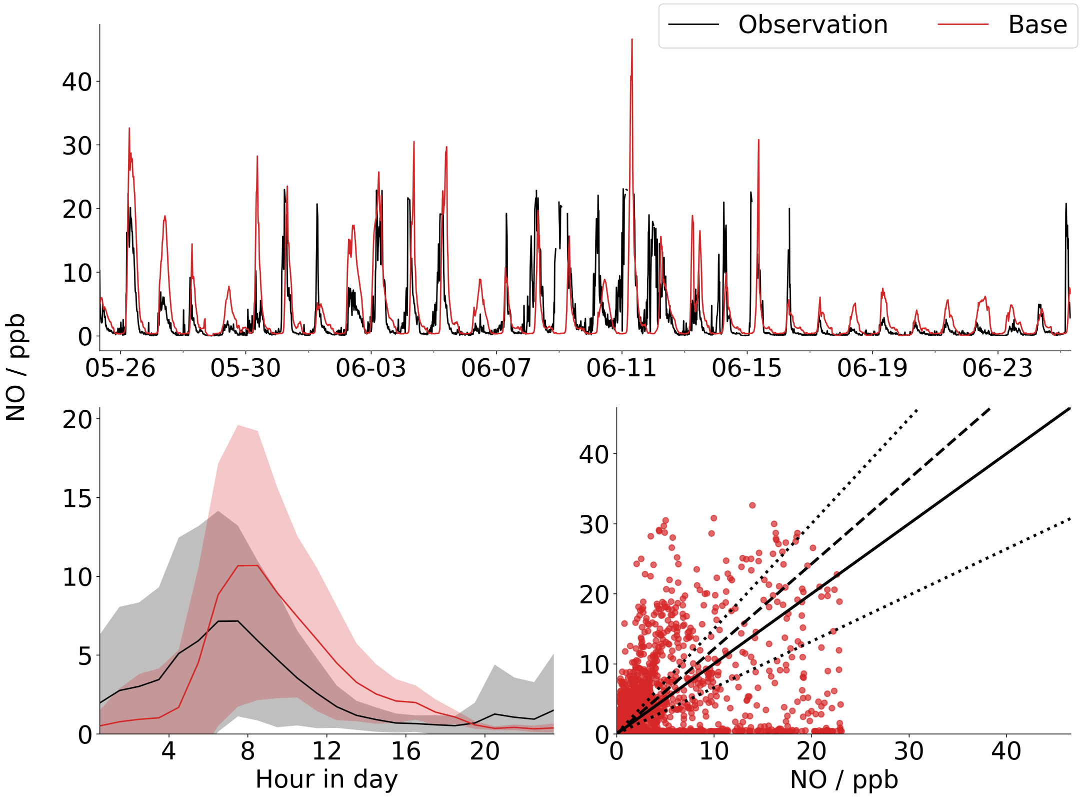

Figure 2.10: Model (red) versus observed (black) NO concentration concentration time series (top), mean diurnal with one standard deviation shaded (bottom left) and a scatter plot (bottom right) of observations (x-axis) against model (y-axis), line of best fit (dashed), 1:1 line (solid) and +/- 50% line (dotted).

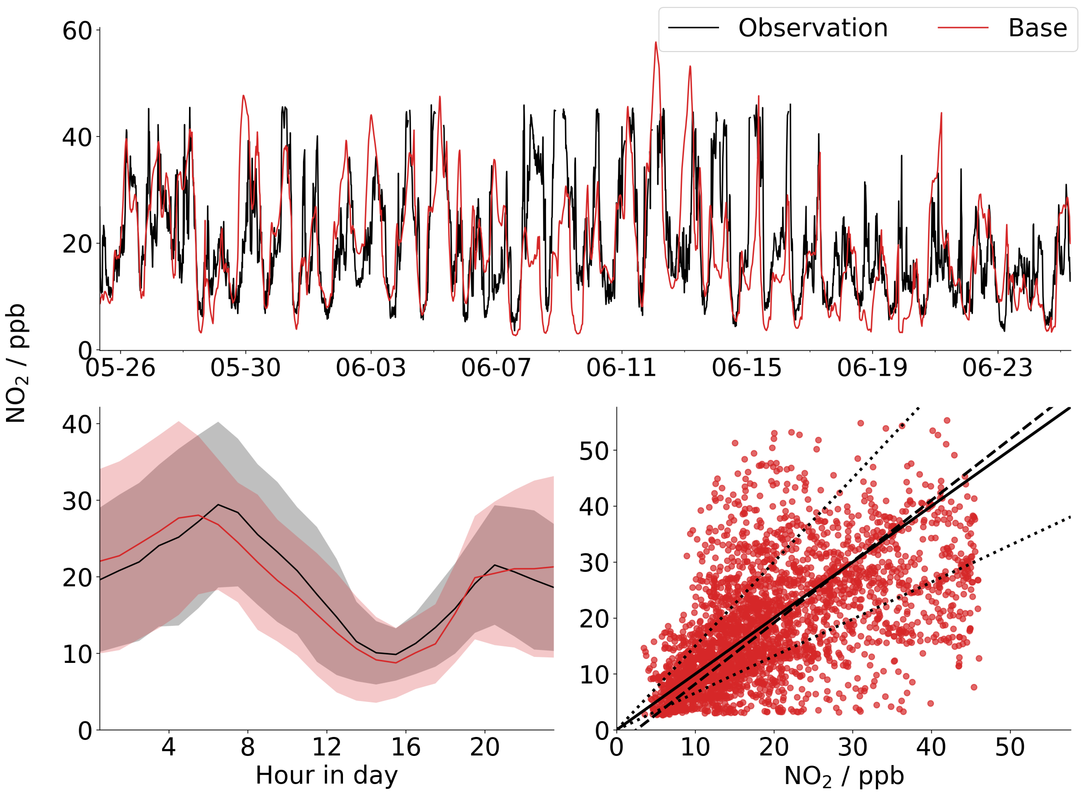

Figure 2.11: Model (red) versus observed (black) NO_2 concentration concentration time series (top), mean diurnal with one standard deviation shaded (bottom left) and a scatter plot (bottom right) of observations (x-axis) against model (y-axis), line of best fit (dashed), 1:1 line (solid) and +/- 50% line (dotted).

Despite the model being able to simulate to NO_x concentrations well (Figure 2.6) with a campaign mean of 22.27 ppbv in the observation and 22.43 ppbv in the model, it simulates the NO and NO_2 concentrations individually less well. The mean modelled NO concentration during the campaign was 2.6 ppbv, compared to the observed mean value of 3.2 ppbv, a high bias of 18%. The NO_2 mean had a smaller bias (5%) but in the opposite direction with a modelled value of 18.6 ppbv compared to an observed value of 19.5 ppbv. Figure 2.12 shows the comparisons between the modelled and measured ratio.

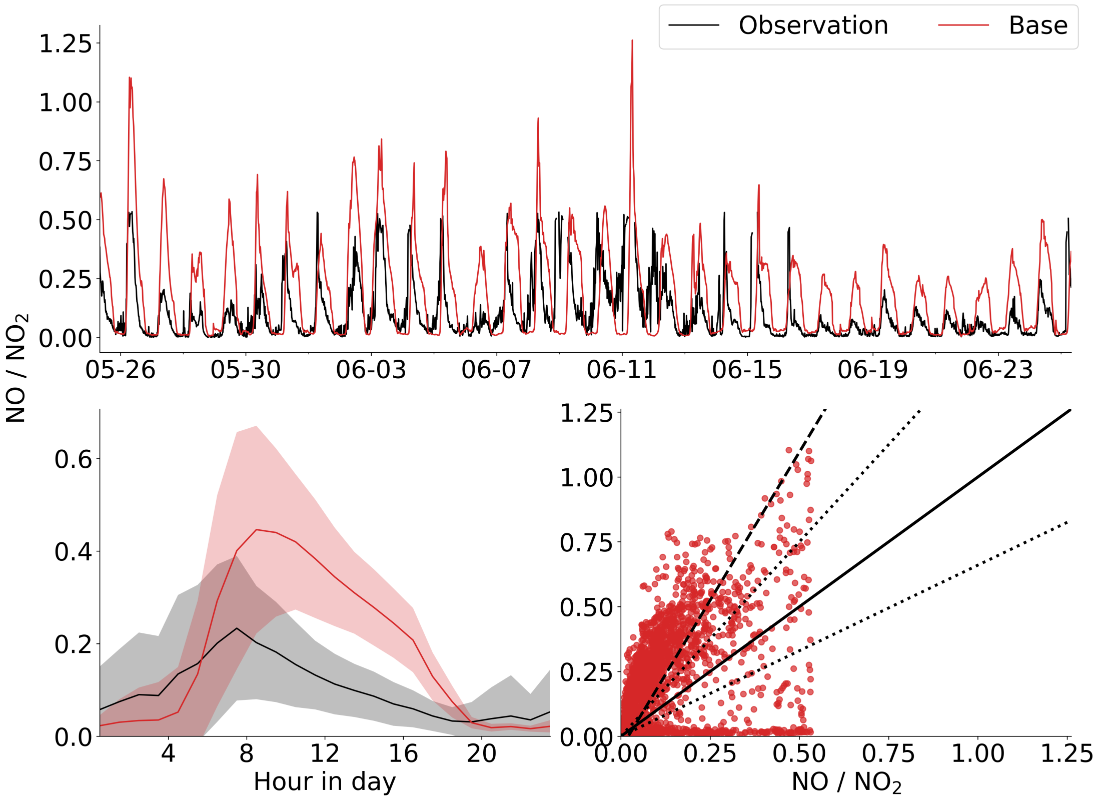

Figure 2.12: Model (red) versus observed (black) NO:NO$_2$ ratio concentration concentration time series (top), mean diurnal with one standard deviation shaded (bottom left) and a scatter plot (bottom right) of observations (x-axis) against model (y-axis), line of best fit (dashed), 1:1 line (solid) and +/- 50% line (dotted).

During the daytime, the ratio should be described well by the Leighton relationship (Chin et al., 1994). Assuming a steady-state for the concentration of NO (or NO$_2$) the rate of chemical production (dominated by NO$_2$ + hv) is balanced by a chemical loss (dominated by NO + O$_3$). Loss rate analysis of NO at noontime found that only a few per cent of NO reacted with RO$_2$ and HO$_2$ rather than O$_3$. The relationship can be thus expressed by Equation 2.1. Where jNO$_2$ is the rate of NO$_2$ photolysis and k is the rate of the O$_3$ + NO reaction.

$$\frac{[NO]}{[NO_2]} \approx \frac{j_{NO_2}}{k[O_3]} \qquad \text{(Eq. 2.1)}$$

In theory, the modelled NO to NO$_2$ ratio should be the same as the observed if the observed O$_3$ concentration and jNO$_2$ photolysis rates are well simulated. In general, the model does replicate the observed O$_3$ concentration well, and while jNO$_2$ does have a mean bias (Section 2.5.1), for the majority of days, the jNO_2 performs well. Thus the measurements appear inconsistent with the Leighton relationship.

Processes occurring below the grid scale of the model are always a potential explanation for NO_x model failure. Local emissions of NO_x may not have reached steady-state, but this would likely lead to higher NO observed concentrations than modelled as the NO enriched emissions decay to the steady-state NO to NO_2 ratio. However, observations show the opposite of this, with the observed NO:NO_2 ratio lower than would be expected from steady-state.

Additional fast cycling of NO to NO_2 would reconcile the observations although it is unclear what could be providing this. The model underestimates RO_2 but to convert NO into NO_2 at the same rate as O_3 (at 70 ppbv and 303 K), there would have to be around 4.5 x 10^9 cm^{-3} of RO_2, whereas the mean observed daytime value is closer to 1x10^9 cm^{-3}. Halogen oxides (BrO, IO, ClO) can also convert NO into NO_2 but were not measured during the campaign. Again however to convert NO into NO_2 at the same rate as O_3, there would need to be 63, 67 or 79 pptv of BrO, IO or ClO respectively. The concentration of these radicals in the atmosphere is poorly understood (Tomas Sherwen et al., 2016), but given the concentrations measured in oceanic sites (Mahajan et al., 2010) of these compounds rarely exceeds a couple of pptv it would seem surprising if there could be tens of pptv in an urban setting.

A problem with any explanation that converts more NO to NO_2, is that it would also increase the rate of O_3 production. Added production would lead to a greater O_3 concentration than is currently simulated, and the model simulation of O_3 is already close to observation.

Several interferences are possible on the measurement of NO and NO_2. Decomposition of NO_y species can lead to NO_2 concentrations appearing artificially high (Reed et al., 2016). It is also possible for species containing alkenes (Alam et al., 2020) to create interference on both NO and NO_2 observations. It is currently not possible to explain why the model is unable to reproduce this ratio. Given the importance of this ratio to our understanding of atmospheric chemistry, this is not reassuring but offers a great place to start important future work.

During the night, the biases are reversed, with modelled NO being too low (50%) and NO_2 being too high (~10%). During the night, there is no photolytic source of NO, and most NO observed is thus from direct emissions. Give the site is located in central Beijing the model likely underestimates the ratio during the nighttime due to local NO emissions not yet having reached steady-state.

Much of the day to day variability is however simulated in the model, with the transition between more and less polluted periods being simulated by the model. This gives the correlation between the model and measured NO and NO_2 Pearson's coefficient (R) of 0.47 and 0.56 respectively.

2.8.2 HONO

Figure 2.13 shows the model with the GEOS-Chem V12.8 default nitrous acid (HONO or HNO_2) sources (green), with an update described below (red) and the observations (black). The default model hugely underestimates HONO during both the day and night.

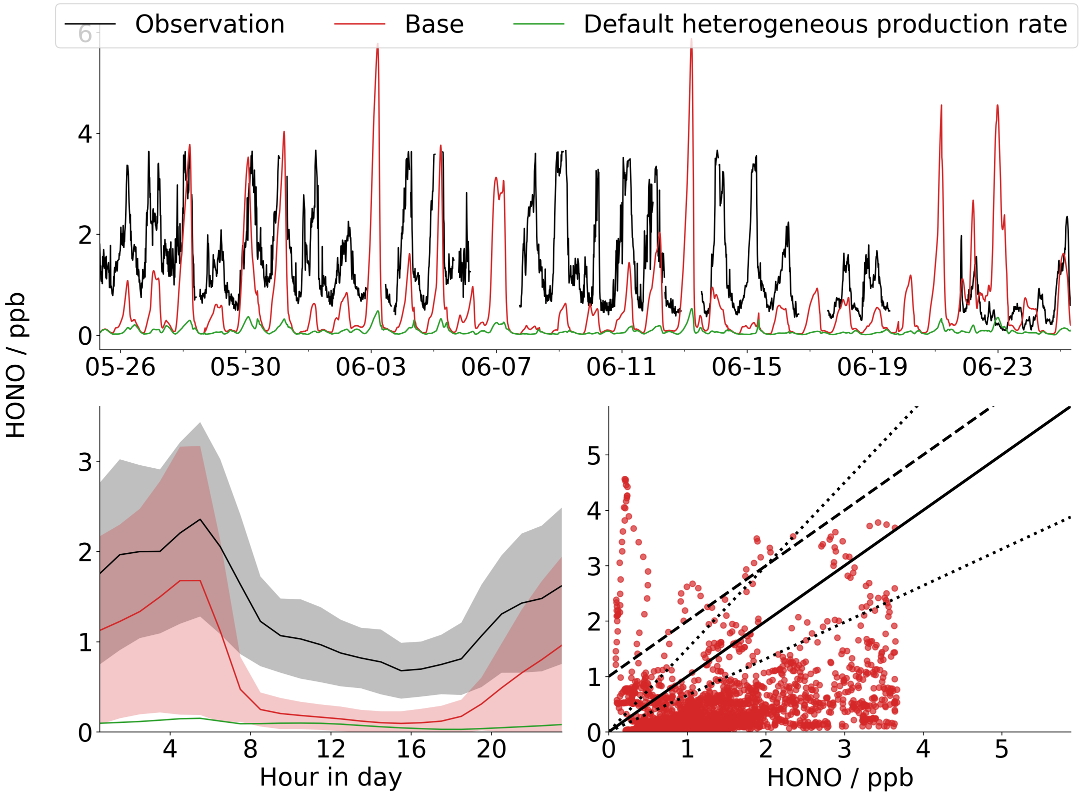

Figure 2.13: Model (red) versus observed (black) HONO concentration concentration time series (top), mean diurnal with one standard deviation shaded (bottom left) and a scatter plot (bottom right) of observations (x-axis) against model (y-axis), line of best fit (dashed), 1:1 line (solid) and +/- 50% line (dotted).

HONO is produced via three production pathways; homogeneously via the gas-phase reaction between OH + NO, heterogeneously via the reaction of NO_2 on particulates or surfaces, and heterogeneously via a light-dependent reaction (L. Zhang et al., 2016). The gas phase source is relatively well constrained and is represented in the model. The model also contains a non-light dependent reaction of NO_2 on aerosols to produce HONO. The default for this is the IUPAC and JPL recommended values for the reactive uptake coefficient (Holmes et al., 2019). The model does not contain a light-dependent heterogeneous source.

With this default configuration, there was a substantial underestimate in the HONO (green line in Figure 2.13). There is however, some uncertainty about the heterogeneous uptake coefficient, with evidence for it being much faster in urban environments (X. C. Lu et al., 2018). The mechanism for this appears to be the complex layer of "organic urban grime" that builds upon aerosol (and potentially other surfaces) enhancing hydrolytic disproportionation of NO_2 (K. F. Ho et al., 2010).

Consequently, the NO_2 uptake coefficient was changed to give a value consistent with that observed in Beijing(1 x 10^{-4}) compared to the default values of 1 x 10^{-6} on organic aerosol and 5 x 10^{-6} on sulfate aerosol (X. C. Lu et al., 2018). The values are also the same as the value used in GEOS-Chem before the Holmes et al., 2019 update. This substantially improved the performance of the model in replicating the nighttime HONO concentration (red line in Figure 2.13), although still resulted in the model being approximately 30% biased low at night.

During the day time the model still substantially underestimates the HONO concentration. There are a number of potential missing sources in the model including direct tailpipe emissions from transport (Liang et al., 2017), soil emissions (Meusel et al., 2018), photolysis of nitrate particulate (Ye, Zhou, et al., 2016) and photolysis of deposited HNO_3 and nitrate on the surface (Ye, Gao, et al., 2016). Additionally, as j value bias was found for other species (NO_2 and $O(^1D)$) it is possible that there was bias in the jHONO.

Overall, even with the nighttime improvement, the model still had an Root Mean Square Error of 1.3 ppb, with a mean mixing ratio of 0.47 ppbv and thus a substantial 260% error. The implications of this underestimate in the HONO for the photochemistry of Beijing is discussed in Section 2.8.9.

2.8.3 Formaldehyde

Formaldehyde (CH_2O) is a prominent product of VOC photochemical oxidation, particularly during the atmospheric degradation of isoprene (Wittrock et al., 2006). The ability of a model to reproduce the observed CH_2O concentration is a useful marker for how well the model replicates the rate of VOC oxidation. The model mean concentration during the campaign was 3.5 ppbv, while that observed was slightly higher at 4.0 ppbv (Figure 2.14).

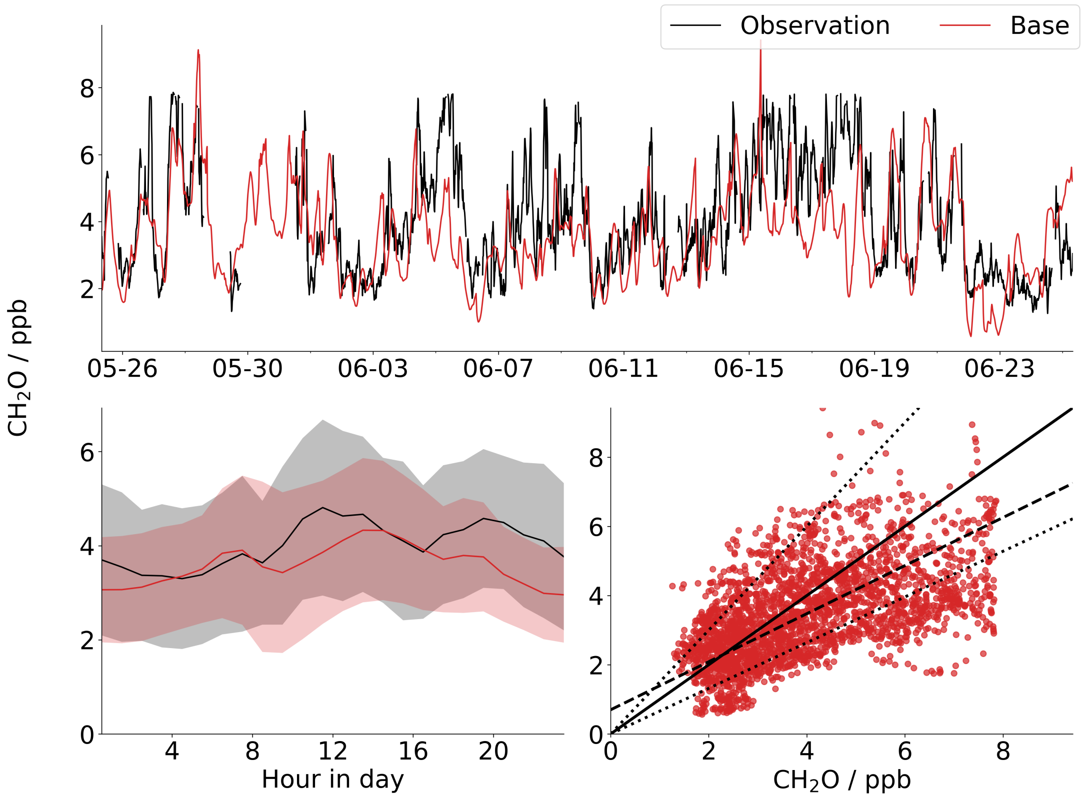

Figure 2.14: Model (red) versus observed (black) CH_2O concentration concentration time series (top), mean diurnal with one standard deviation shaded (bottom left) and a scatter plot (bottom right) of observations (x-axis) against model (y-axis), line of best fit (dashed), 1:1 line (solid) and +/- 50% line (dotted).

Qualitatively the model has some success in simulating the overall trend during the campaign, with the differences between high and low days being reproduced. This gives the model an R value of 0.54. Similar results are found from model evaluations of CH_2O in other locations such as the south east united states (Marvin et al., 2017).

The model's ability to simulate CH_2O would provide some evidence that its rate of hydrocarbon oxidation is, roughly comparable to that observed. However, this will be in contrast to the evidence from the observation of OH Reactivity and the RO_2.

2.8.4 OH

The local hydroxyl radical (OH) concentration is rarely evaluated within chemical transport models. Indirect global or hemispheric evaluations are more common (Krol, Leeuwen, and Lelieveld, 1998; Bey, Jacob, et al., 2001; W. J. Bloss et al., 2005), but it is usually thought that given the short lifetime of OH (0.01 to 1s), evaluation using a transport model, even with a 25km grid resolution would not provide useful information. The difficulty in making OH observations also limits their availability (Heard and Pilling, 2003). Nevertheless, as the dominant oxidant in the atmosphere, the OH concentration governs the lifetime of most gas phase compounds. Our ability to accurately reproduce the OH concentration has a profound effect on our ability to simulate atmospheric chemistry and thus it is useful to evaluate model performance, even with the caveat of the model's grid resolution.

Figure 2.15 shows a comparison between the OH observations made using the FAGE technique (Section 2.3) and those simulated. Although the model captures the rough shape of the observations with a daytime peak at noon and lower concentrations during the morning and evening, it significantly underestimates the concentration at all times.

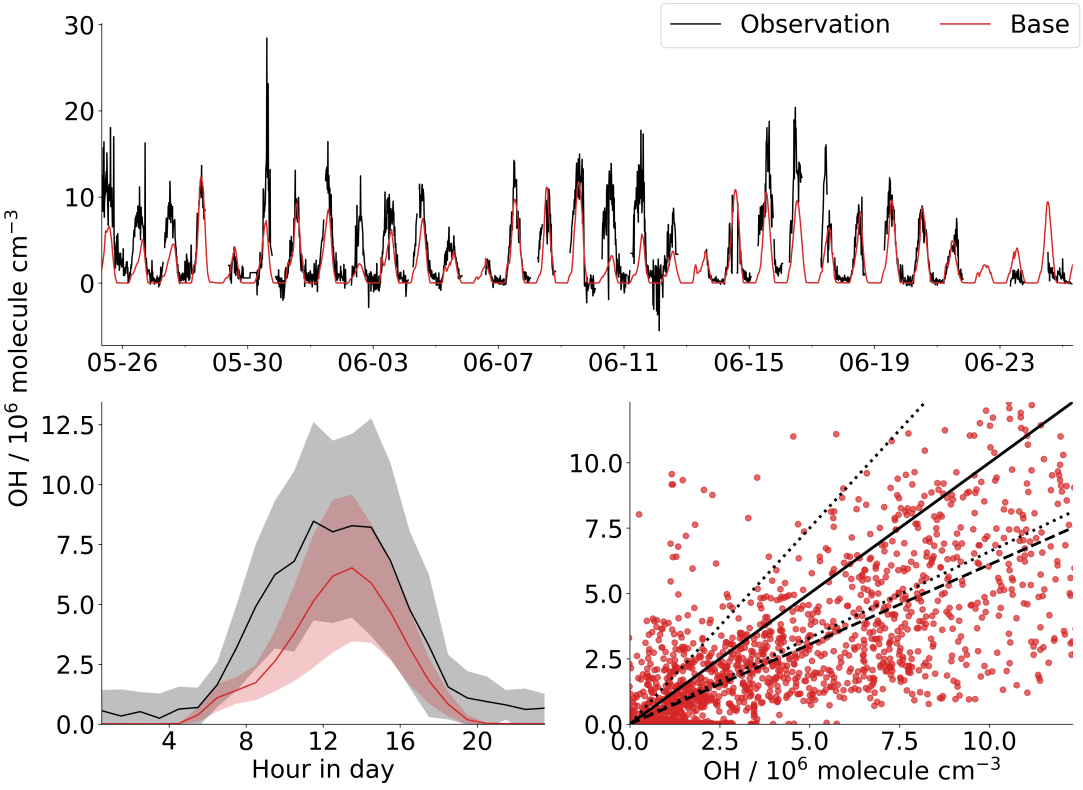

Figure 2.15: Model (red) versus observed (black) OH radical concentration concentration time series (top), mean diurnal with one standard deviation shaded (bottom left) and a scatter plot (bottom right) of observations (x-axis) against model (y-axis), line of best fit (dashed), 1:1 line (solid) and +/- 50% line (dotted).

The mean OH concentration during the campaign was 2.6 x 10^6 molecules cm^{-3} in the model and 3.7 x 10^6 molecules cm^{-3} in the observations. The model thus underestimates the mean OH concentration by 30%. The bias was comparable to the 30% found by Z. F. Tan, Rohrer, et al., 2018 in a box model study run over the same region but for a different year. This is larger than found in a study of this campaign (Whalley et al., 2020).

At noontime the model underestimation was smaller at 1x10^6 molecules cm^{-3}, 12.5% lower than the 8x10^6 molecules cm^{-3} observed. Given the model's substantial overestimation of $J(O^1D)$ (Section 2.5.1) it is surprising that the model underestimates OH. However, photolysis of HONO plays a dominant role in the sources of OH in Beijing (Whalley et al., 2020). Thus model underestimation in morning HONO (Section 2.8.2) may be responsible for the modelled OH concentration rising slower than the observation during the morning (Couzo et al., 2015), and the general underestimation of OH be due to the underestimation of HONO during the day. The model over estimation of $J(O^1D)$ may, to some extent, be compensating for the model underestimation of HONO.

During the night, the model simulates essentially zero OH whereas the observations appear to detect some OH, although these concentrations are close to, or at the limit of detection of the instrument. At night there are no photolytic sources of radicals, however, there are "dark sources" in the atmosphere from the reaction of alkenes with O_3 or NO_3 (Geyer et al., 2003). However, as the model underestimates night time NO concentrations (Section 2.8.1), even if it gets the dark source of peroxy-radicals right there would be a suppression of nighttime OH as there would be an underestimation of the rate of nighttime conversion of peroxy radical into OH.

The observed OH concentration shows a high degree of day to day variability with the OH concentration varying by an order of magnitude between the lowest (28^{th} May) and highest days (29^{th} May). The model captured much of this variability resulting in an R-value of 0.75. Much of this variability in OH is driven by variability in the photolysis rates (Section 2.5.1). Days with low photolysis rates correlate with days with low OH, reflecting the change between cloudy and clear days.

2.8.5 HO_2

As with OH, the hydroperoxy (HO_2) radical concentration calculated by chemistry transport models is seldom evaluated. However, calculating a correct hydroperoxy radical concentration is vital for reproducing O_3 production in the models as the reaction between NO and peroxy radicals is the source of O_3 in the troposphere.

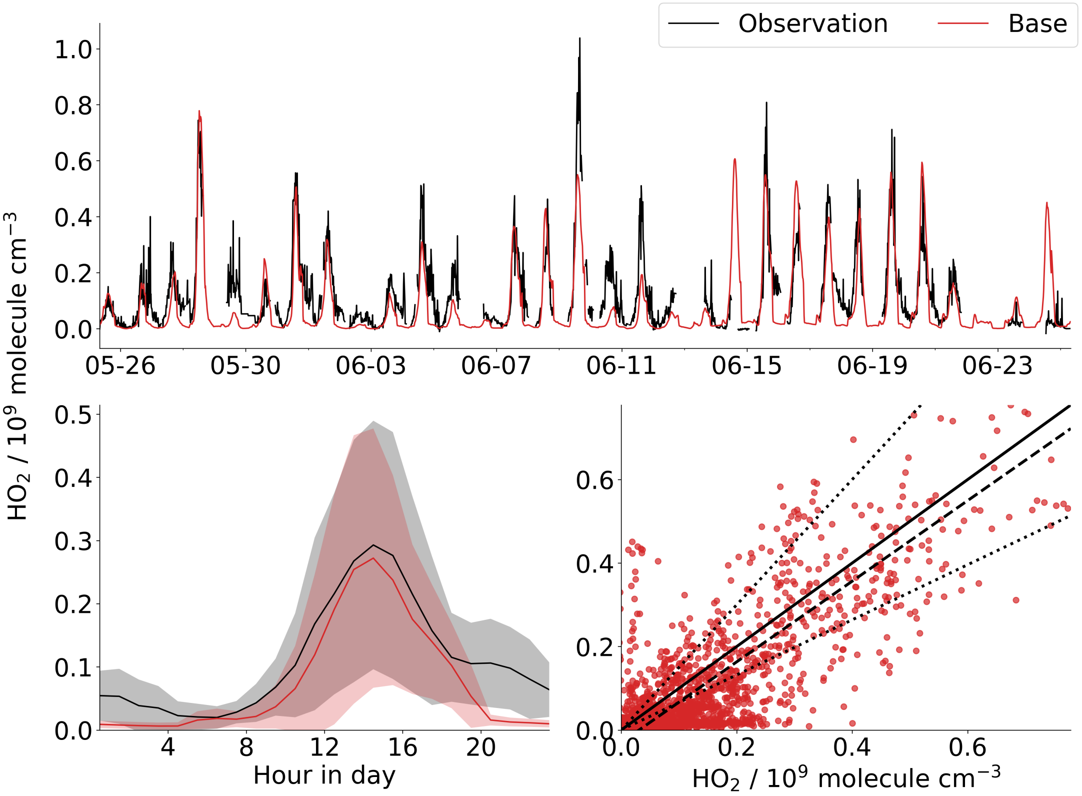

Figure 2.16: Model (red) versus observed (black) HO_2 radical concentration concentration time series (top), mean diurnal with one standard deviation shaded (bottom left) and a scatter plot (bottom right) of observations (x-axis) against model (y-axis), line of best fit (dashed), 1:1 line (solid) and +/- 50% line (dotted).

Throughout the campaign (Figure 2.16), the mean observed HO_2 concentration was 1.2 x 10^8 molecules cm^{-3}, whereas the model simulated 0.8 x 10^8 molecules cm^{-3}.

Thus on average, the model underestimated the HO_2 by 33%. The RMSE was 9.3 x 10^6 molecules cm^{-3} which at approximately 50% of the mean value which compares favourably with a previous box model study by Z. F. Tan, Rohrer, et al., 2018 for the same region.

During the day, the model agreement with the observations is generally good. There is a slight underestimation during the morning, which (as seen with OH) is likely due to the shortage of nighttime HONO resulting in not enough morning radical production. This is in contrast to the results of Whalley et al., (2020) who find their model substantially overestimates the HO_2 concentration during the day time. This probably reflects the difference in HO_x concentrations calculated by their model (higher as they use the observed HONO concentrations), and then differences in HO_x speciation as their model uses the measured NO and the GEOS-Chem model overestimates NO concentrations (see Section 2.8.1) which would reduce HO_2 concentrations.

Similar to OH, the model predicted a significantly lower concentration of HO_2 at night than are observed. There is a noticeable evening shoulder in HO_2 concentrations which is not simulated in the model. "Dark" production routes of RO_x do exist (Geyer et al., 2003). The reaction of the double bonds in alkenes with either O_3 or NO_3 can lead to the production of radicals (Kroll et al., 2002). The model simulation of both ethene and PRPE (>= C3 alkenes) is generally good after tuning (Figure 2.6). Nighttime isoprene showed concentrations of around 0.1 ppbv, whereas the model simulates close to zero, due to the emissions scheme having no dark emissions source. Additionally, there was a significant concentration of mono-terpenes at night (15.0 pptv) measured by the PTR-MS (H. Zhang et al., 2020), which were not considered at all by the model chemistry. Some additional work understanding night time chemistry in GEOS-Chem is needed.

2.8.6 RO_2

Around 80% of the peroxy-radical pool (HO_2+RO_2) is in the form of organic peroxy-radicals (RO_2) (Figure 2.17), reflecting the sizeable organic fraction of the reactivity, compared to that due to CO (Section 2.8.7).

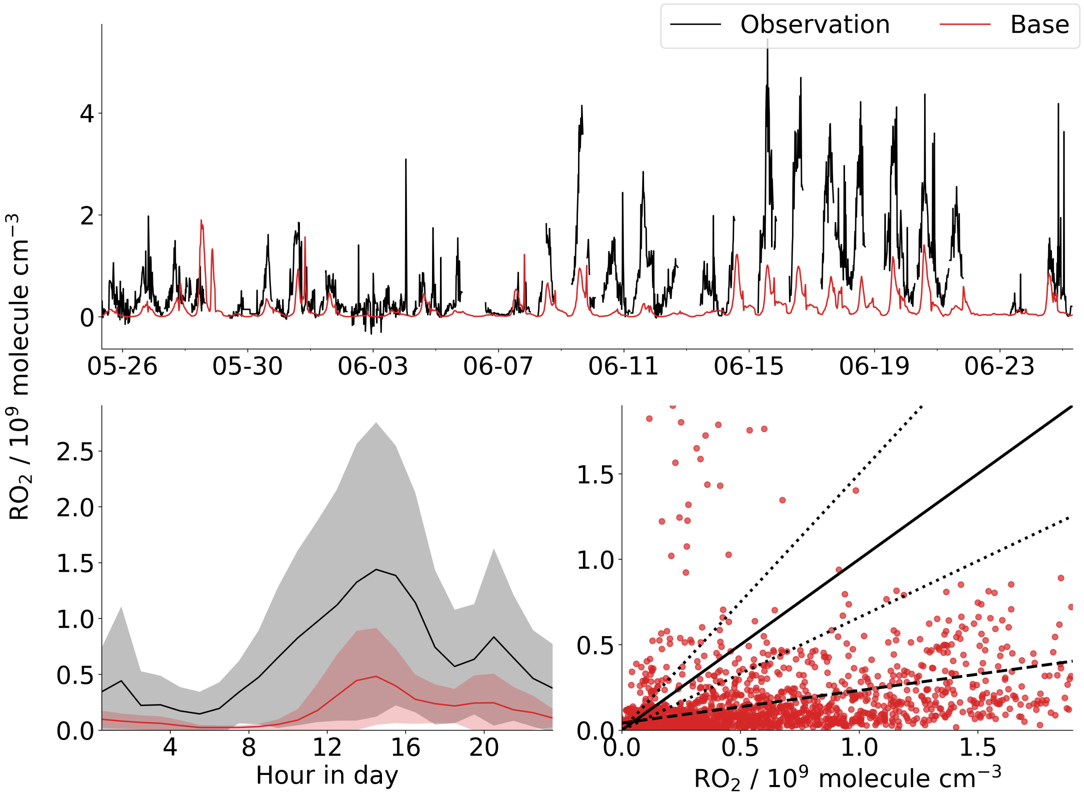

Figure 2.17: Model (red) versus observed (black) RO$_2$ radical concentration concentration time series (top), mean diurnal with one standard deviation shaded (bottom left) and a scatter plot (bottom right) of observations (x-axis) against model (y-axis), line of best fit (dashed), 1:1 line (solid) and +/- 50% line (dotted).

The model substantially underestimates RO$_2$ concentration. Whereas the mean observed concentration was 5.9 x 10^8 molecules cm^{-3}, the model only simulated 1.6 x 10^8 (Figure 2.17).

The most significant underprediction of RO$_2$ was during the morning. As with OH and HO$_2$, this was likely due to a lack of radicals due to an underestimation of nighttime HONO. However, unlike the other radicals (OH and HO$_2$), the bias was consistent through the day. This is consistent with the model's sizeable underestimate in the OH reactivity.

The observations show a noticeable shoulder in RO$_2$ concentrations in the evening. This is seen in other studies (Z. F. Tan, Hofzumahaus, et al., 2020) which is attributed to the ozonolysis source of radicals extending further into the afternoon than the photolytic source. The model shows some evidence of this but at a lower level than that observed.

2.8.7 OH Reactivity

OH reactivity is a measurement of the OH lifetime and represents the first-order loss rate constant for OH. The OH reactivity can be measured directly in the atmosphere (Yang et al., 2017). It can also be extracted from the models via the summation of the first-order rate of every reaction that removes an OH radical (Equation 2.2).

$$k' = \sum_X [X] k_{OH+X} \qquad \text{(Eq. 2.2)}$$

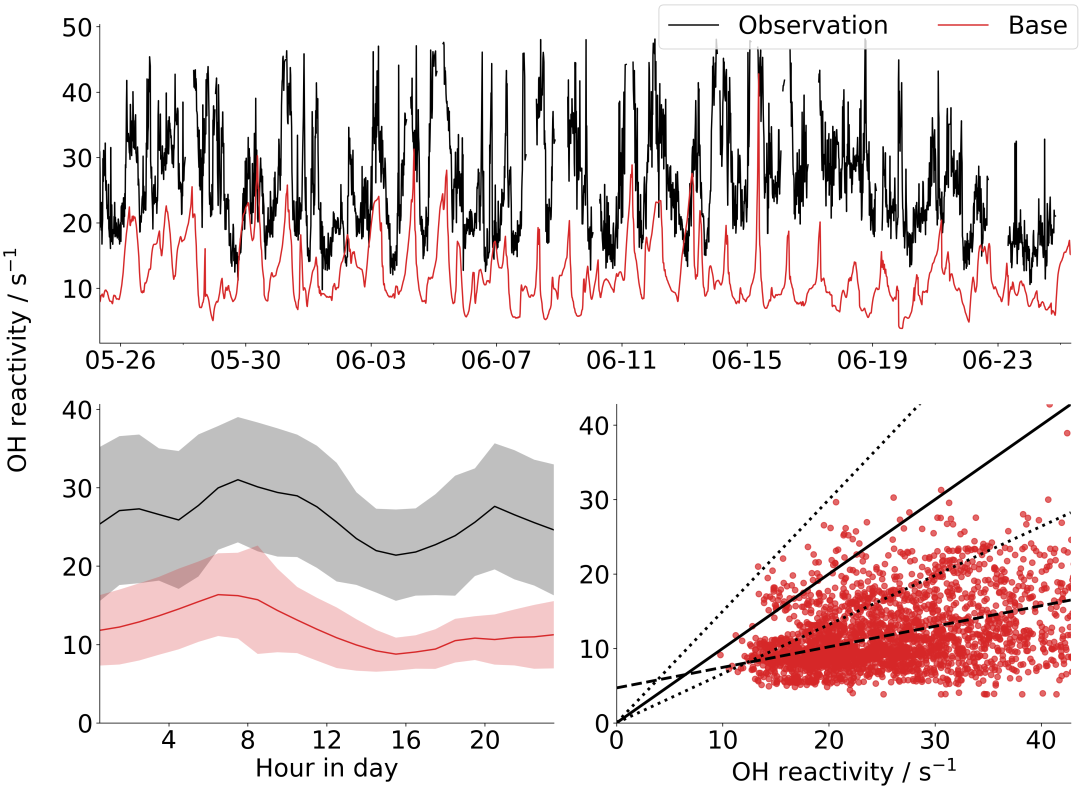

Figure 2.18: Model (red) versus observed (black) OH reactivity concentration time series (top), mean diurnal with one standard deviation shaded (bottom left) and a scatter plot (bottom right) of observations (x-axis) against model (y-axis), line of best fit (dashed), 1:1 line (solid) and +/- 50% line (dotted).

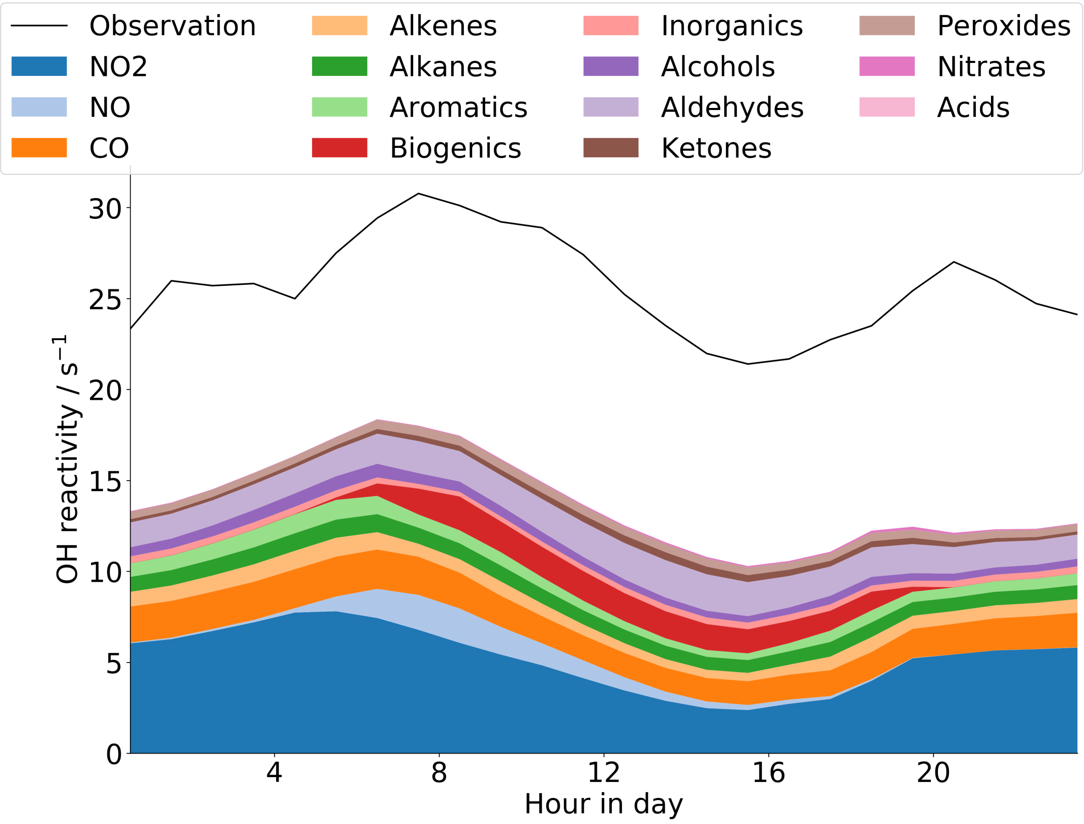

Figure 2.19: Stacked plot of modelled diurnally averaged (mean) OH reactivity by compound class, with total observed OH reactivity in black.

It can be seen in Figure 2.18 that the model underpredicts OH reactivity by about 55%. This under prediction is a well-known issue with models (Porter, Safieddine, and Heald, 2017; Z. Tan et al., 2019; Lisa K. Whalley et al., 2016). It is higher than that found from the study of (Whalley et al., 2020), who explored the same dataset but from a box model perspective. They also found their model underestimated the reactivity (by around 10 s^{-1}), a lesser extent than that seen here (around 13 s^{-1}). Yang et al., 2017, again in a box model in the same region of China, found a 21% model underestimate.

Previous studies have concluded that low bias is possibly explained by unmeasured VOC and oxidation products. Some of this may be species missing from the GEOS-Chem chemistry scheme, but Whalley et al., (2020) used the MCM chemistry for their modelling and still found a significant underestimation. If there is a missing VOC, it would need to provide an almost continuous source of reactivity, as the difference between the measured and modelled reactivity is similar (13 s^{-1}) during the day and the night.

One difference between the underestimate in the box model (Whalley et al., 2020) and GEOS-Chem is likely due to the box model considering the primary and degradation reactivity from monoterpenes whereas GEOS-Chem does not, and the GEOS-Chem underestimate of the NO_2 concentrations (Section 2.8.1).

Unmeasured compounds that might provide some additional reactivity could be sesquiterpenes. These are highly reactive (lifetime of 7 minutes at an OH concentration of 8x10^6 cm^{-3}) biogenic compounds. For those species to make up the difference in reactivity (12 s^{-1}), there would need to be (using an MCM recommend OH rate constant of 1.97 x 10^{-10} cm^3 molecules^{-1} s^{-1}) 1.5 ppbv of those compounds, which given their short lifetime, would require a large flux. There might be a whole unmeasured class of compounds (Janechek, Hansen, and Stanier, 2017; Kumar, Chandra, and Sinha, 2018) or heterogeneous reactions (H. Xiao and B. Zhu, 2003) but with current understanding, it is hard to explain the missing OH reactivity.

Some support for a missing source of reactivity comes from the RO_2 observations (Section 2.8.6). The modelled RO_2 is significantly underestimated by the model. A substantial increase in the reactivity (if the OH reactions made RO_2) could help to reduce this problem.

2.8.8 O_3

Given all of the problems the model appears to have with the fast photochemistry (NO_x and RO_x), the model does a surprisingly good job of simulating the mean O_3 (Figure 2.20). The mean modelled O_3 value of 51.6 ppb is close (2.9%) to the observed 53.1 ppb. This is better than recent studies using chemical transport models in the same region (+7.5% X. Lu et al., 2019, +5% P. Wang et al., 2019, +16% Hajime Akimoto et al., 2019. However, this performance is not a surprise, as the emissions in the model have been tuned to give the right concentration of the primary species.

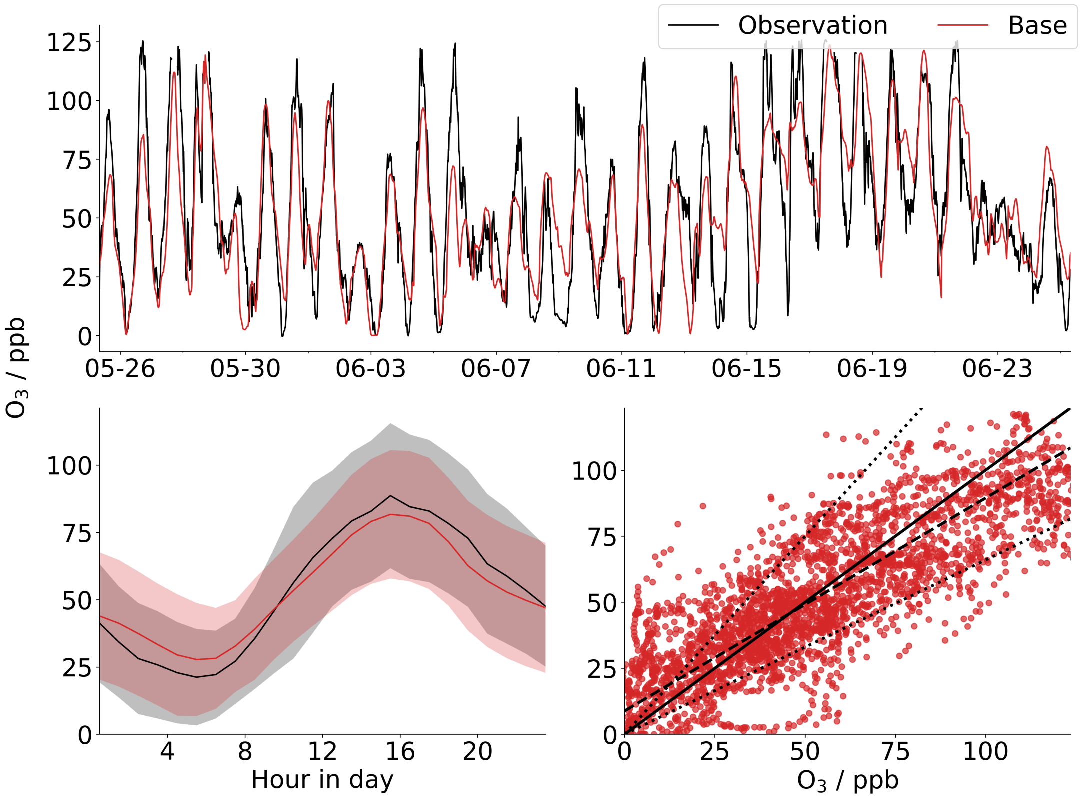

Figure 2.20: Model (red) versus observed (black) O_3 concentration concentration time series (top), mean diurnal with one standard deviation shaded (bottom left) and a scatter plot (bottom right) of observations (x-axis) against model (y-axis), line of best fit (dashed), 1:1 line (solid) and +/- 50% line (dotted).

The model captures much of the day to day variability in the observations, with an R value of 0.82. This is likely primarily due to the model's ability to simulate the relative variability in the daily photolysis rates (Section 2.5.1).

The diurnal mean O_3 over estimated just before dawn, and underestimated at around 4pm, so the average O_3 production during the day is underestimated by around 10 ppbv or around 20%. This would allow for some additional O_3 production due to missing RO_2 but would unlikely allow for a large increase in O_3 production without a substantial increase in O_3 sinks.

2.8.9 Discussion

The overall objective of this chapter is to explore the sensitivity of the model's O_3 to different potential policy levers (NO_x reductions and VOC reductions). Using the model for experiment necessitates some faith that the model's response to those leavers has some relationship to the likely responses in the real atmosphere. From the perspective of the model's ability to simulate O_3, this appears to be good. The model performs surprisingly well at simulating O_3. However, there are many ways that the model could get the right O_3 concentration for the wrong reason. The chemical sources and sink terms balance to the correct answer while those terms are not necessarily correct. There is therefore, the necessity to assess other aspects of the model, and it was in the secondary compound performance that the model was found most wanting.

There still remains some significant problems in the model's ability to simulate quantities such as the NO to NO_2 ratio, the concentration of RO_2 and OH etc. However, the model skill is not that different from that shown by box models (see for example Whalley et al., 2020). A number of future research topics are highlighted by this analysis: sources of daytime HONO, reasons for the failure to simulate the NO to NO_2 ratio, missing reactivity, missing RO_2 sources, nighttime sources of radicals etc. But the tuned model appears to provide a substantially closer representation of the photochemistry of Beijing than the untuned one. Attention can now be turned to understanding whether Beijing is in a NO_x or VOC limited chemical regime (Section 2.9), the role of different VOCs in leading to O_3 production (Section 2.11) and the influence of heterogeneous chemistry (Chapter 3).

2.9 Photochemical regime

Air pollution control strategies for O_3 are typically based on understanding whether a city is in a NO_x or VOC limited regime (Chatani et al., 2014). When the photochemical regime is NO_x limited, an increase in the NO_x concentration leads to an increase in the O_3 concentration (Sillman, 1999). In this regime, the radical loss is predominantly via the self reaction of peroxy radicals, thus increasing NO_x emissions increases the amount of NO available to react with peroxy radicals to produce NO_2, without decreasing the concentration of radicals, leading to more O_3 production. Radical partitioning favours HO_2 and RO_2 over OH. As OH production limits the production of peroxy radicals, increasing VOCs does not lead to an increase in peroxy radicals, and so does not lead to an increase in O_3.

As NO_x emissions rise, more peroxy radicals react with NO rather than other peroxy radicals, resulting in increased OH concentrations. The NO_2 + OH reaction becomes the largest radical loss pathway. At this point, increasing NO_x emissions further leads to a decrease in peroxy radical concentrations, and so in this regime, increasing NO_x emissions decreases the O_3 concentration. Additional VOC would now rapidly react with the higher OH concentration to produce more peroxy radicals which would themselves rapidly react with NO forming NO_2. As such, an increase in VOC would increase O_3. This regime is referred to as VOC limited.

Increasing the NO_x emission even further would eventually result in a high O_3 + NO reaction rate which would cause a substantial amount of the odd oxygen ($O_x = O_3 + O(^1D) + O(^3P) + NO_2$) being held as NO_2 rather than as O_3. At this point, increasing NO_x emissions rapidly reduces O_3 concentrations, although odd oxygen concentrations remain the constant.

Recently it has become apparent that another factor at play in urban ozone photochemistry. Although uptake of HO_2 onto aerosol has been known to be important for over 20 years (Jacob, 2000), it's importance for urban environments has only been recognised more recently (K. Li et al., 2019). This radical loss pathway is not considered in detail in this chapter but forms the basis of Chapter 3.

In order to identify which regime (NO_x limited or VOC limited) Beijing was in during the campaign, the model was run scaling VOC (both biogenic and anthropogenic) and NO (anthropogenic only) emissions in the model domain by factors of 0.8 and 1.25 (0.8 = 1/1.25). Lightning, biomass and soil emissions of NO_x were unchanged as their influence on Beijing were negligible.

Figure 2.21 shows the spread of percentage changes in the daily max O_3 concentration at the observational site for the period of the campaign for these four simulations. On average (solid coloured vertical lines), a decrease in NO emissions or an increase in VOC emissions increased O_3. Conversely, when NO was increased, or VOCs were decreased O_3 concentrations decreased. Therefore on average Beijing would be referred to as being in a VOC limited regime (Xing et al., 2018).

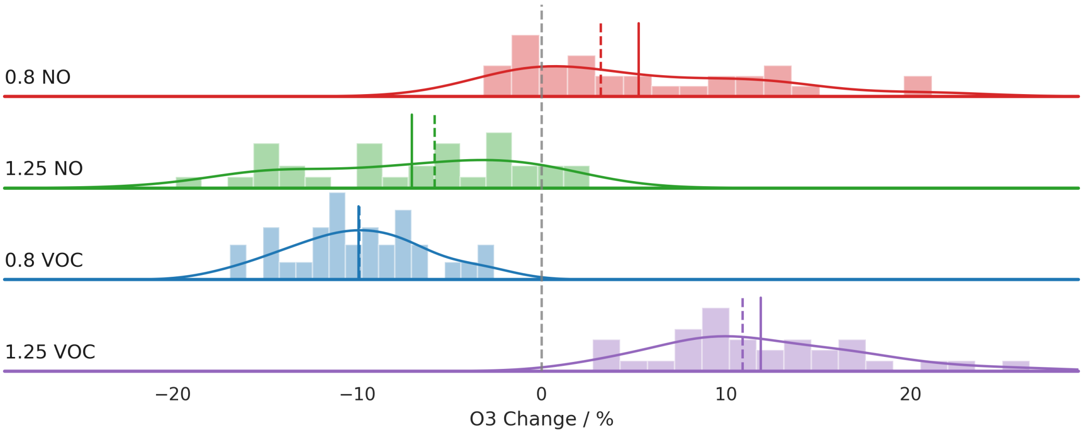

Figure 2.21: Distribution plot of modelled 4 pm O_3 during the campaign with different scale factors on NO and VOC emissions. Mean (solid) and median (dashed) are indicated with the vertical lines.

There is a marked difference in the distribution of daily O_3 concentrations changes (Figure 2.21) from changing NO or VOC emissions. Whereas the changes in VOC emissions leads to a consistent increase or decrease in O_3, changes in the NO emission can lead to both increases or decreases depending on the day. The VOC perturbation distribution appears as a single mode, whereas the NO perturbations appear to be bimodal.

The time series of the change in the daily maximum O_3 with a reduction (0.8 scale factor) in NO can be seen in Figure 2.22. The change to the NO emission elicited a different response depending on the day of the campaign. While much of the campaign was clearly in a VOC limited regime (a reduction in NO emissions to to an increase in O_3), over a third of the campaign appeared to show no O_3 sensitivity or presented a slight (< 2 %) NO limited regime.

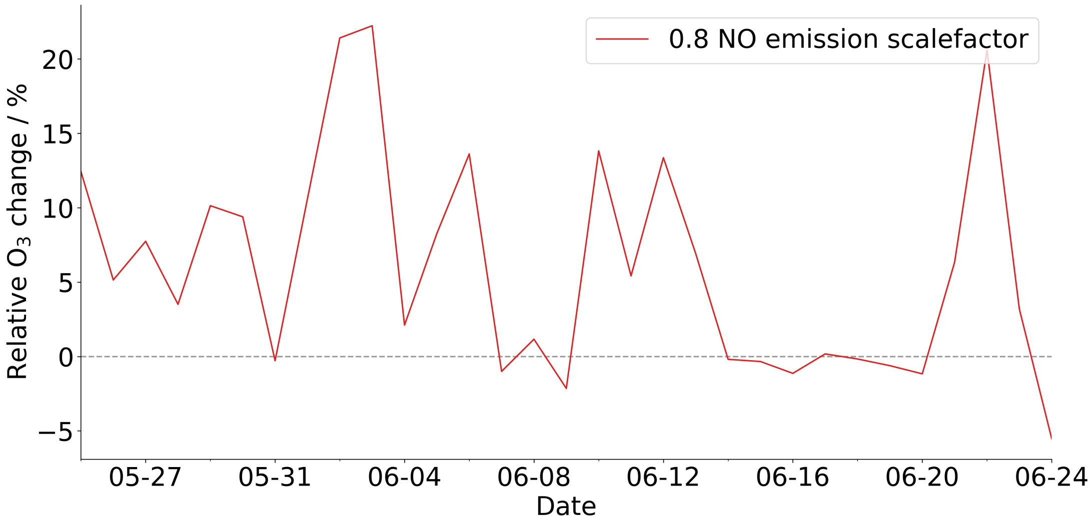

Figure 2.22: Time series of the reduction in daily maximum O_3 with a 0.8 scale factor on NO emissions.

The standard tool for understanding the photochemical regimes is the ozone isopleth (Thielmann et al., 2001). These were historically generated with box models, but given the advances in computational resources, these can now be calculated with chemical transport models.

2.10 Ozone isopleths

Historically, assessing and explaining the different photo-chemical regimes (NO_x vs VOC limited) is achieved by considering isopleths of O_3 concentration under different emissions of NO_x and VOCs (Sillman and He, 2002). These are typically generated using box models run with chemical and meteorological conditions thought "typical" of the city under considerations (Sillman and He, 2002).

In contrast to box models, chemical transport models simulate large areas simultaneously and have sophisticated transportation mechanisms allowing for upwind and downwind mixing, vertical transport, deposition schemes, coupled aerosol chemistry, etc. The computational burden of running these models at sufficiently high resolution prevented their use for resolving isopleths. However, increases in computational resource now allow isopleths to be constructed from chemistry-transport models.

The GEOS-Chem model was used to construct O_3 isopleths for Beijing by varying the NO (anthropogenic only) and VOC (anthropogenic and biogenic) emissions independently. The scaling was symmetric in log space with nine logarithmically distributed steps between 0.1 and 10 (0.10, 0.18, 0.32, 0.56, 1.0, 1.8, 3.2, 5.6, 10.0) for both NO and VOCs. This "coarse grid" covered a broad range of emission scaling with a focus on chemical understanding. A finer grid with nine logarithmically distributed steps between 0.32 and 3.2 (0.32, 0.42, 0.56, 0.75, 1., 1.33, 1.78, 2.37, 3.2) was nested inside the coarse grid to offer greater chemical resolution over the policy relevant range.

2.10.1 Aggregated isopleths

Figure 2.23 shows the modelled mean 4 pm (local) O_3 isopleth for Beijing during the period of the campaign. 4 pm was chosen as it represented the daily O_3 peak (Section 2.8.8). Increases in the VOC emissions are shown on the x axis, and increasing NO emissions are shown on the y. The red 'ridge' line running across the graph shows the point of maximum O_3 concentration at a given VOC emission scale factor, and the red dot represents the "tuned" model simulation described in the previous sections,

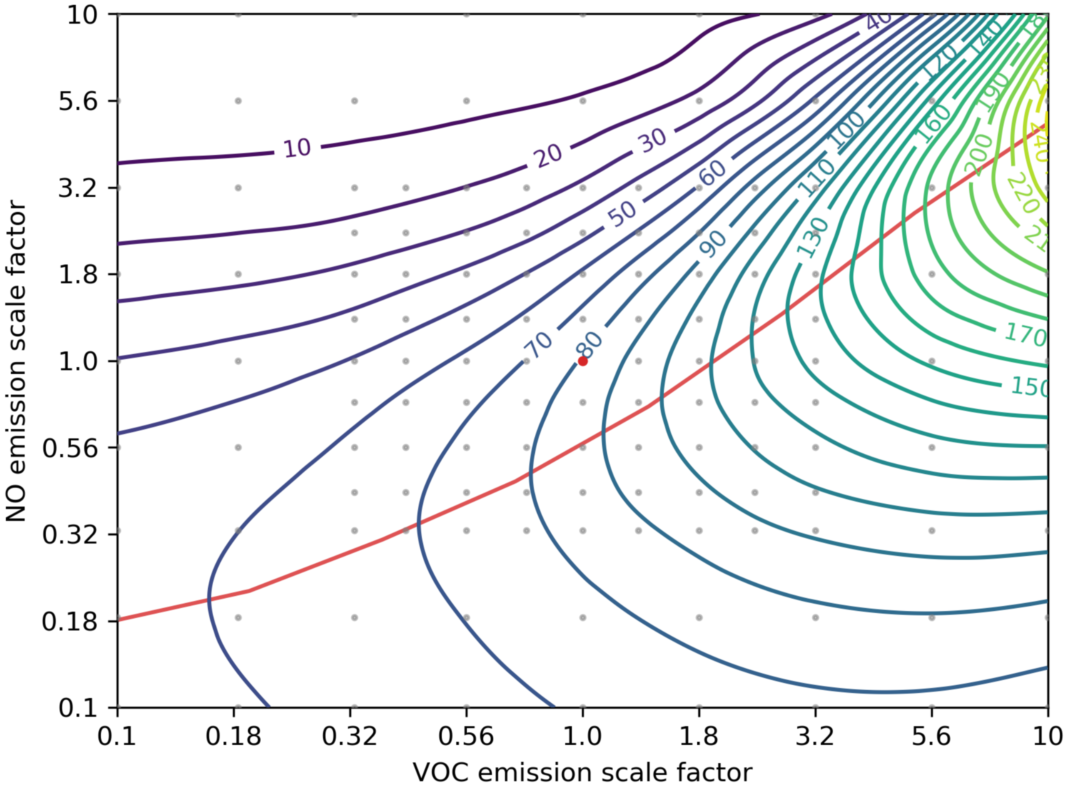

Figure 2.23: Isopleths of the mean 4 pm O_3 concentration (ppb) for the period of the campaign. Grey dots indicate what the VOC (anthropogenic + biogenic) and the NO_x (anthropogenic) were multiplied by in the simulations. Red dot indicates the model after tuning.

The standard features expected from an O_3 isopleth (Finlaysonpitts and Pitts, 1993) are evident. The upper-left half of the plot represents the VOC limited regime, where increases in VOC emissions sees an increase in O_3 concentrations. In this area, increased NO emissions result in lower O_3, and ultimately O_3 is almost entirely titrated away. Conversely, the lower-right half of the plot shows the NO_x limited region, where sensitivity to VOC is very small (horizontal lines), but increases in NO_x emissions lead to an increase in O_3 concentrations.

The city's emissions are currently located at the red dot in the centre of the plot with a mean O_3 concentration on the 80 ppbv isopleth. These conditions are at a higher NO emissions rate / lower VOC emissions rate than the ridge line of maximum O_3 (red line). The model is thus in a VOC limited regime: a decrease in VOC emissions lead to a decrease in O_3, whereas a decrease in NO emissions will lead to an increase in O_3, in agreement with findings in Section 2.9.

Figure 2.23 also shows that reductions in the NO emissions will lead to an increase in the mean 4pm O_3 concentration. A reduction in NO emissions of 50% would lead to an increase in O_3 to 87 ppbv. Reductions beyond that would the lead to reduced O_3 concentrations. But to return to the present-day concentration of 80 ppbv would require a 80% reduction in NO emissions.

Reductions in VOC emissions, however, provides a greater opportunity for reduction in O_3 concentration. A 50% reduction in VOC emissions would produce a 20% drop in the 4 pm average O_3 (80 to 64 ppbv). This reduction could likely be achieved in less than a 50% total reduction by targeting high O_3 forming potential VOCs, which will be discussed further in Section 2.11.

Figure 2.24 shows the median 4 pm O_3 isopleth. The most obvious feature of the plot is the noisy contour lines. As the median is the value above which 50% of the data lies, any change in the rank ordering of the 4pm O_3 will likely lead to a step change in the value of the median hence the noise. Section 2.9 showed that the O_3 response to a change in the NO_x or VOC emission varied significantly between days. The mean plot reduces the importance of these changes by considering all of the data points, in the median plot, a change in emissions could shift the median day. Further evidence can be seen by the fact that the curves are most erratic close to the ridgeline. As emissions are pushed to the extremes, it is likely all the days converge on one regime (at ten times NO emission all days will be very VOC limited).

Although there are differences in between the mean and median isopleth, these differences are rather small and do not change the perspective that the city in the model is in a VOC limited regime.

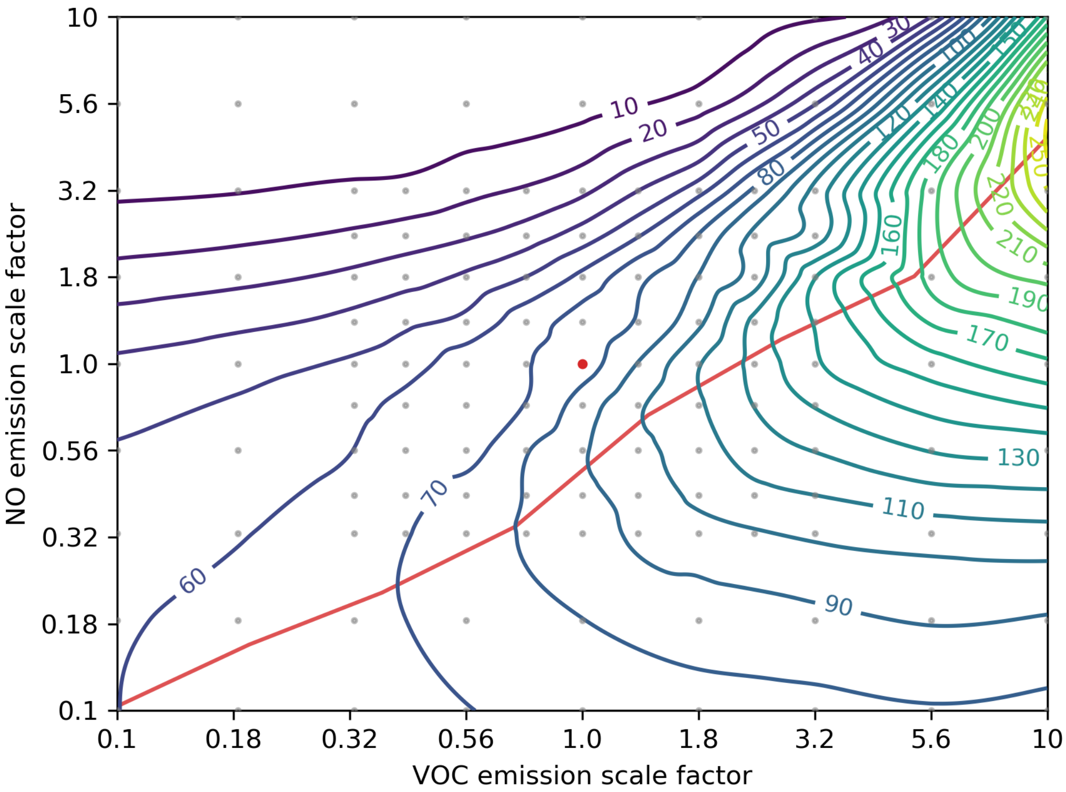

Figure 2.24: Isopleths of the median 4 pm O_3 concentration (ppb) for the period of the campaign. Grey dots indicate what the VOC (anthropogenic + biogenic) and the NO_x (anthropogenic) were multiplied by in the simulations. Red dot indicates the model after tuning.

So far, the focus has been on the average daily 4 pm O_3, but one of the key advantages of using a chemical transport model to produce isopleths is that the full campaign is resolved. O_3 damage to both the environment and human health (Heath, Lefohn, and Musselman, 2009) is non-linear, with the higher O_3 concentration days being substantially more damaging than the lower days. Figures 2.25 and 2.26 shows how the modelled 75th percentile 4 pm O_3 concentration (103 ppbv under current conditions) and the maximum 4pm value (120 ppbv under current condition) would be affected by changing emissions. The overall structure of these isopleths is similar to that of the mean (Figure 2.23) with the present day conditions for the city being at higher NO emissions that the ridge line. However there are some differences.

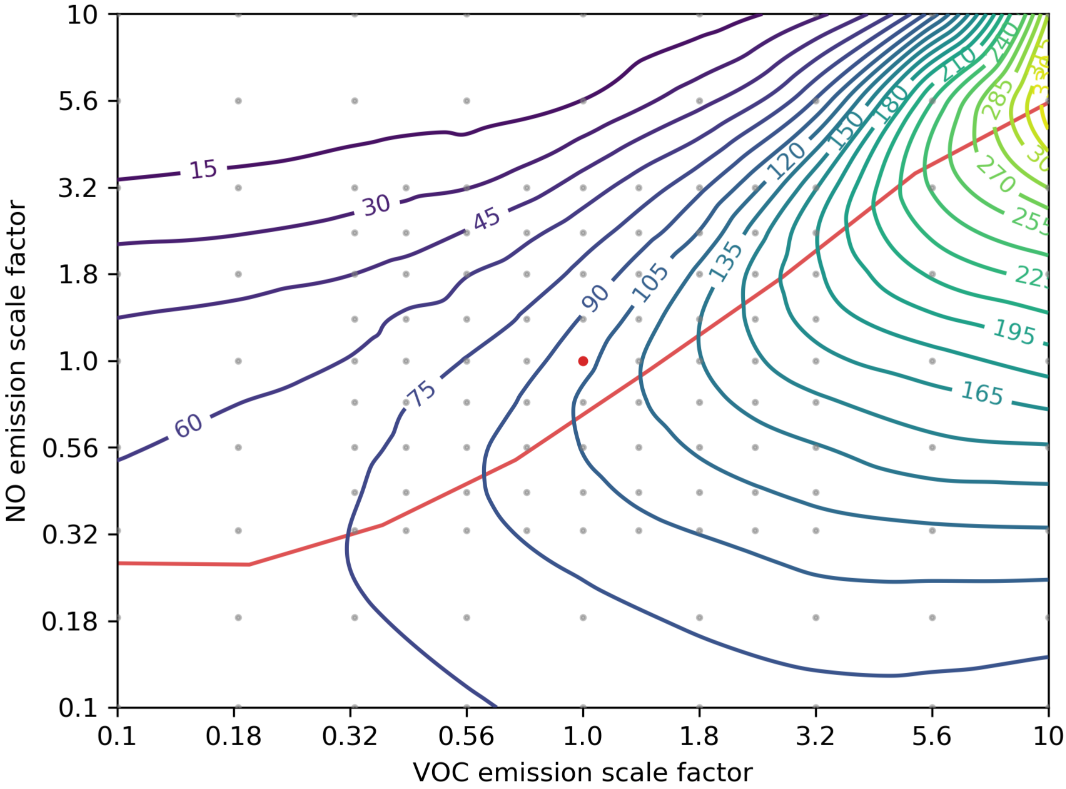

Figure 2.25: Isopleths of the 75th percentile 4 pm O_3 concentration (ppb) for the period of the campaign. Grey dots indicate what the VOC (anthropogenic + biogenic) and the NO_x (anthropogenic) were multiplied by in the simulations. Red dot indicates the model after tuning.

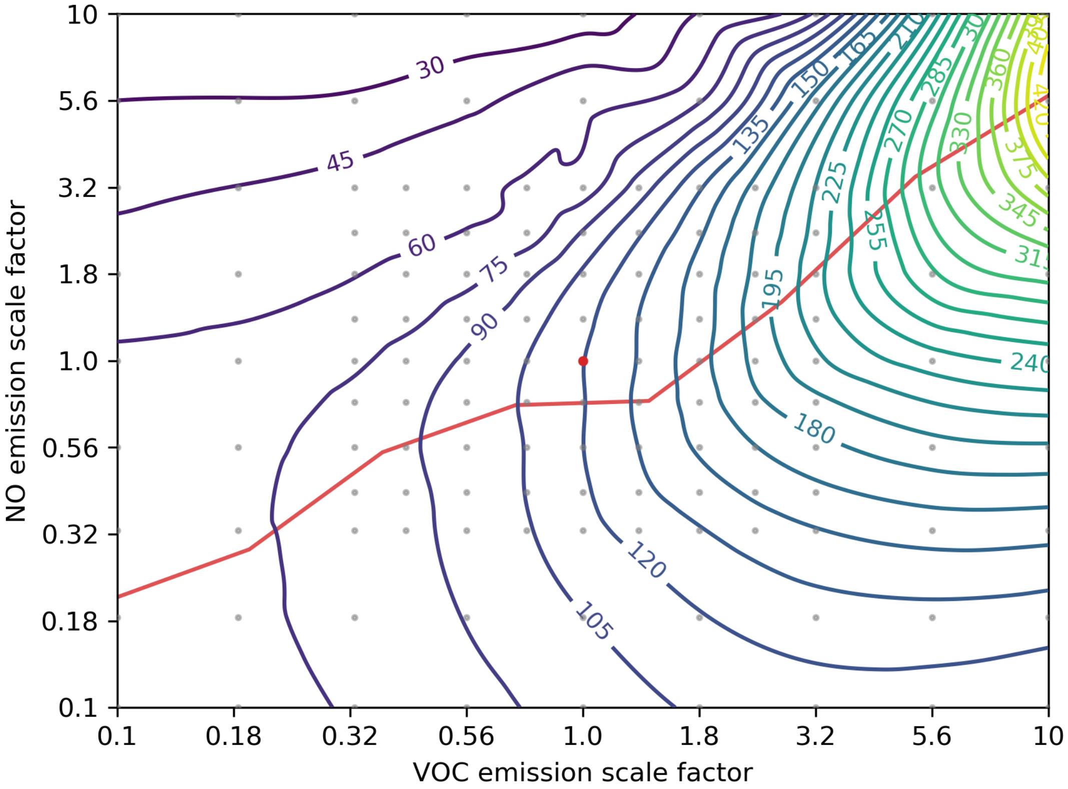

Figure 2.26: Isopleths of the max 4 pm O_3 concentration (ppb) for the period of the campaign. Grey dots indicate what the VOC (anthropogenic + biogenic) and the NO_x (anthropogenic) were multiplied by in the simulations. Red dot indicates the model after tuning.

In both cases, the current conditions (red dots) lies closer to the ridge line than seen in the mean or median isopleths. For the 75th percentile, reducing NO emissions leads to an increase of only a couple of ppbv in the O_3 concentration before the O_3 drops. Reducing the NO would barely increase max O_3 at all. Thus depending upon the metric that policy wishes to improve (mean, 75% percentile, maximum), the disadvantage of NO_x emission reductions changes.

2.10.2 Daily variability

The isopleths on individual days can be examined to provide a more in-depth analysis. Days when O_3 was above the 75^{th} percentile can be split into two periods, the 27th to 28th of May, and the 17th to 21st of June. These two periods with be referred to as the "May" and "June" peaks, respectively and can be characterised by different isopleths.

An example of the May peak period isopleth can be seen in Figure 2.27 (4pm May

28th). This day is characterised by as being VOC limited and the isopleth has a similar form to the average isopleth (Figure 2.23) with NO_x reductions increasing O_3.

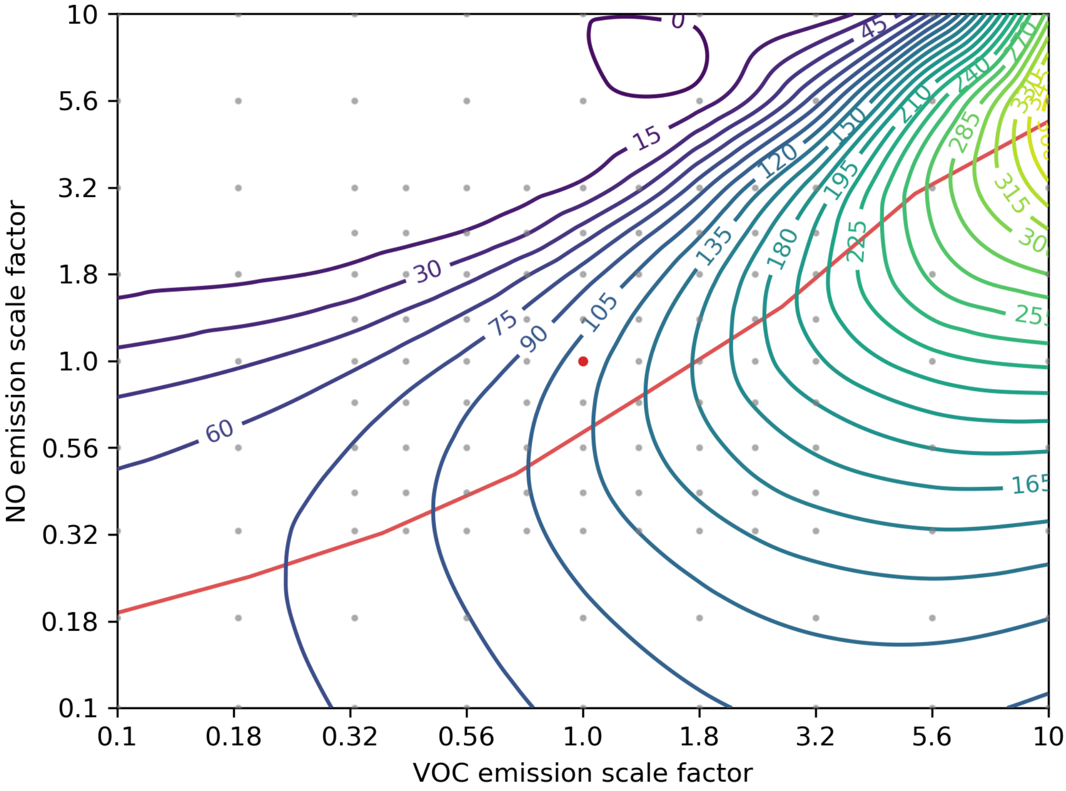

Figure 2.27: Isopleths of the May 28^{th} 4 pm O_3 concentration (ppb) for the period of the campaign. Grey dots indicate what the VOC (anthropogenic + biogenic) and the NO_x (anthropogenic) were multiplied by in the simulations. Red dot indicates the model after tuning.

The June period shows different behaviour. During this (June 20th example shown in Figure 2.28) the O_3 sits on the crest of the ridgeline, lying between the two regimes. While VOC reduction would still reduce O_3 more efficiently than reductions in NO_x, reductions in NO_x would still produce a slight reduction as well. The regime here is neither NO or VOC limited.

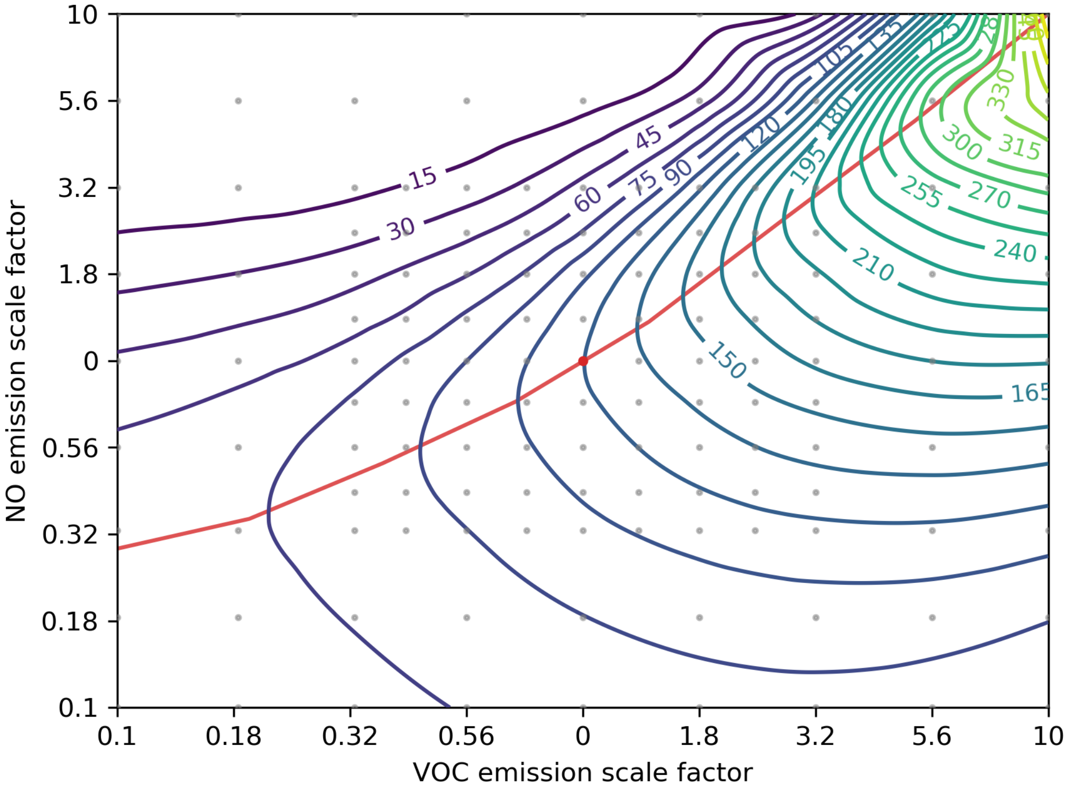

Figure 2.28: Isopleths of the June 20^{th} 4 pm O_3 concentration (ppb) for the period of the campaign. Grey dots indicate what the VOC (anthropogenic + biogenic) and the NO_x (anthropogenic) were multiplied by in the simulations. Red dot indicates the model after tuning.

2.10.3 Implications

What is important to note here is that the June peak period represented the most extended period of very high O_3 during the campaign, with all five days having O_3 concentration greater than 100 ppb. Neither the average nor the 75th percentile isopleths placed the base model on the ridgeline as it was with the June period.

A model study focused solely on the June high period would conclude that Beijing is at the point of maximum O_3 from a NO_x emission perspective, thus focusing emission reductions on both NO and VOC would be the optimal policy strategy. The May peak period, however, indicates NO_x reduction would have minimal effect on the O_3. Thus a greater focus would be on VOC reduction over NO_x. Lastly, a model study that focused solely on the mean would conclude that Beijing is VOC limited, and a NO_x emission reduction would increase O_3. Under these conditions, the emission reduction strategy may be to reduce VOC drastically to offset the increase caused by NO_x reductions.

This shows the importance of examining individual O_3 pollution events when considering the NO_x and VOC mitigation strategies. VOC reduction was shown to be an effective strategy for reducing O_3 concentrations in Beijing under all conditions. The individual contribution of different VOCs, and thus, most efficient VOC to target to reduce O_3 will now be investigated.

2.11 Importance of different VOCs

Section 2.10 showed that the modelled mean, median, 75th percentile and max 4pm O_3 concentration are all sensitive to VOC emissions. However, not all VOCs are the same. Different VOCs display differing capacities to produce O_3 (Duan et al., 2008). In this section, the influence of different VOCs classes on modelled O_3 is explored.

The primary emitted VOCs are split into five classes: CO, alkanes (ethane, propane, and ALK4: alkanes with 4 or more carbons atoms), alkenes (ethene and PRPE : alkenes with 3 or more carbon atoms), isoprene and aromatics (benzene, toluene, xylene, trimethylbezene and combined other monoaromatics). CO is technically not a VOC but we include it here due to its ability to generate peroxy radicals and so O_3. In order to compare each of their contributions to O_3 production, a series of perturbation simulations were performed, where each class of compound was run with one of five multiplication factors; 0.0 (no emissions), 0.2, 0.4, 0.6 or 0.8 applied to all emissions sources (anthropogenic and biogenic) over the whole domain and then compared to the base model.

In Figure 2.29 the percentage changes in the Beijing 4 pm O_3 concentrations can be seen for each of the perturbation simulation. Changing the CO, alkenes and alkanes emissions each has a similar median contribution to O_3 production of around a 6 to 7% reduction. While CO concentrations are much higher than than those of alkanes, which are again much higher than the alkenes (campaign means of 483, 50 and 5 ppbv respectively), their ozone formation potential is reversed (approximately 1, 14, 84 respectively) (H. Xiao and B. Zhu, 2003) resulting in a similar overall influence on O_3. The variability in the sensitivity of the O_3 (as represented by the inter-quartile range) is smaller for the alkanes and CO than the alkenes. As they are longer lived compounds the boundary conditions plays a more important role in defining their concentrations which means that the local influence on the concentration is smaller.

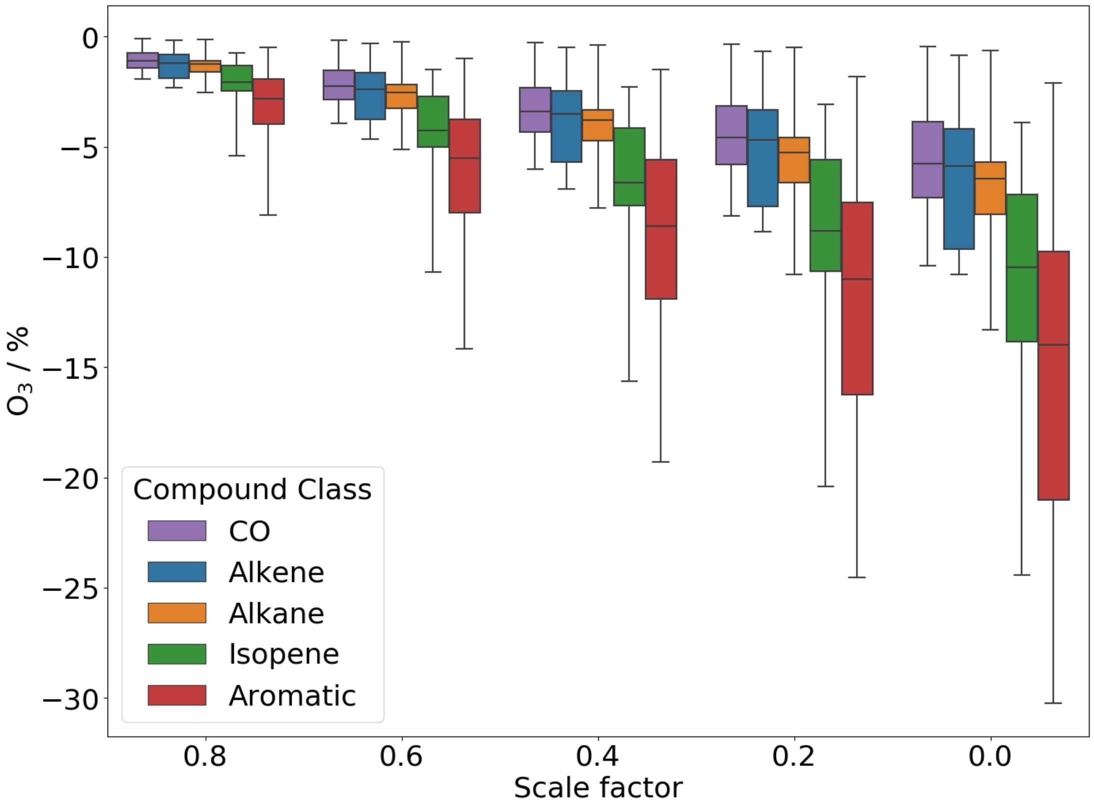

Figure 2.29: Box and whisker plot of change in base O_3 concentration with a scale factor on the emissions of each class of VOC compound.

Removal of isoprene emissions has a more substantial effect on the O_3 concentrations reducing the median 4 pm O_3 by ~11%. The inter-quartile range (8% to 14%) is also broader than for the CO, alkenes and alkanes, as the emissions flux is sensitive to temperature and solar flux which vary day to day unlike the anthropogenic emissions which follow the same diurnal variation in emissions every day. This leads to some days having isoprene as a significant contributor to the OH reactivity, whilst on other days it plays a less significant role.

Aromatic compounds were the most significant contributor to O_3 with their removal inducing a 15% reduction in the median 4 pm O_3 concentration. Aromatics also have the largest inter-quartile range of all the compound classes. A possible reason for this variability is that aromatic compound emissions in the region are predominantly from urban (X. M. Wang et al., 2002), and industrial (Y. Li et al., 2020) centres and have short lifetimes (24 hours or less for methyl-substituted aromatics at 5 x 10^6 OH molecules cm^{-3}) and low background concentration. Therefore air masses coming from polluted regions could have a far more significant aromatic influence on O_3. Similarly, these air masses that do not pass over forested areas could see a reduction in biogenic influence on O_3 and thus a relative increase in aromatic influence. Additionally, relative aromatic influence will increase on days when isoprene emissions are lower due to meteorological conditions.

The influence of removing emissions of these VOCs on the mean diurnal cycle of the short lived species (O_3, OH, HO_2 and RO_2 is shown in Figure 2.30).

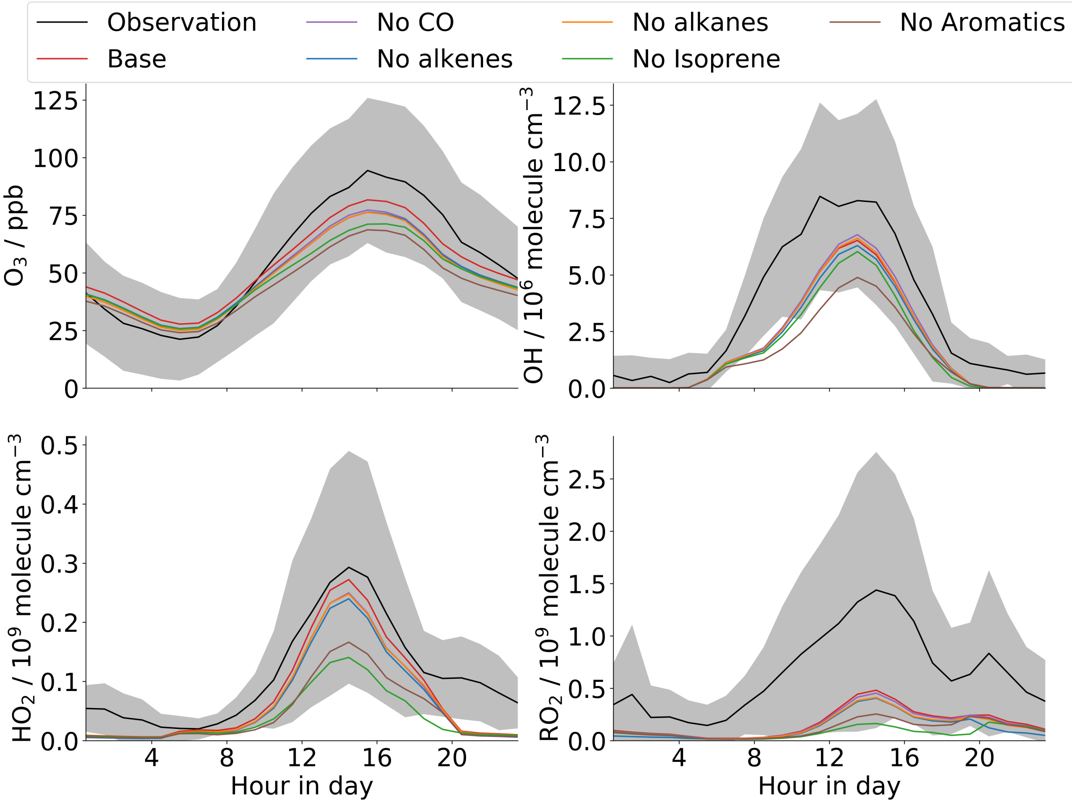

Figure 2.30: Diurnal mean O_3 (top left), OH (top right), HO_2 (bottom left) and RO_2 (bottom right) concentration with the removal of emissions of each class of VOC. Observation in black with shaded one standard deviation.

The change in the mean O_3 diurnal cycle from the different emission classes, unsurprisingly follows that in Figure 2.29. Switching off the aromatics leads to the largest changes, and the removal of the CO the smallest change. However, the influence of the different VOC on the short lived species is more complex.

While aromatics compounds did substantially reduce the HO_2 and RO_2 concentrations (~40% and ~50% respectively), the reduction was not as significant as seen with removing isoprene emissions (~55% and ~62% respectively). This is unintuitive and may represent changes to the O_3 sink term. The concentration of peroxy radical changes more for the isoprene, but the influence is less on the O_3.

VOCs can also influence the concentration of NO which in turn can affect the concentration of O_3. Is the asymmetry between the O_3 concentration and the peroxy-radical concentrations for isoprene and aromatics due to their influence on NO_x concentrations? Figure 2.31 shows the impact of switching off the emission of the different VOCs on the concentration on NO, NO_2, NO_x and NO_y.

Overall the VOCs have little influence on the concentration of NO_y. There are marginally higher NO_x concentrations without the aromatics, which likely reflects the influence of aromatics on increasing the OH concentration. There are also higher NO concentrations without aromatics, which reflect the increase in the NO to NO_2 ratio with the lower O_3 without the aromatics but from this analysis, it doesn't look like the switch in importance between the aromatics and the isoprene for O_3 and peroxy radicals is due to their influence on NO_x concentrations.

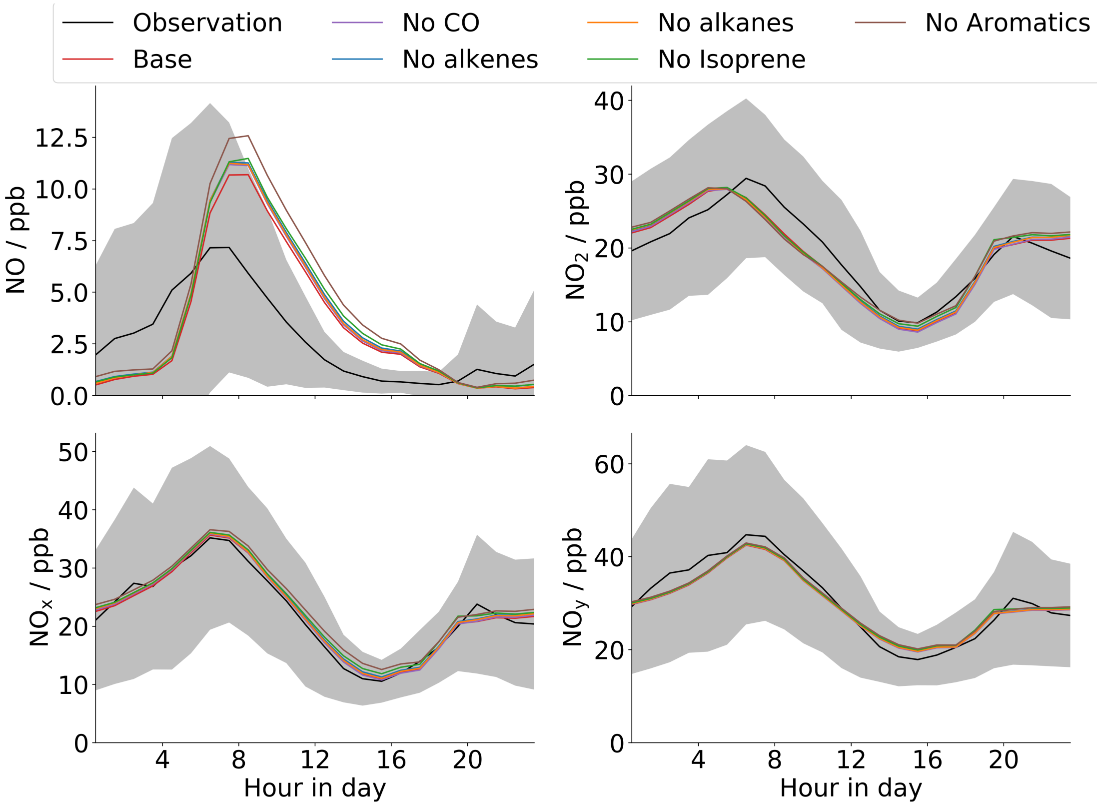

Figure 2.31: Diurnal mean NO (top left), NO_2 (top right), NO_x (bottom left) and NO_y (bottom right) concentration with the removal of emissions of each class of VOC. Observation in black with shaded one standard deviation.

Another explanation might be in the difference in the mechanism choices made for the isoprene (Bates and Jacob, 2019) and aromatic (Goliff, W. R. Stockwell, and Lawson, 2013) oxidation schemes. These are very different schemes. The Bates and Jacob, 2019 scheme for isoprene is designed to work in both high and low NO_x environments. It contains a thorough representation of isoprene peroxy radicals reacting with HO_2, RO_2 and NO. It is based on the 412 reactions and is a state of the art representation of the chemistry of isoprene. The RACM-II aromatic scheme is designed to be much more compact. Radicals generated in this scheme can only react with HO_2, NO and CH_3O_2. The rate constants chosen for these reactions are also simplified and have no temperature dependence. The different response of the O_3 to the changes in the peroxy radicals may be due to the different assumptions made in formulating the chemistry schemes. Further work will be necessary to understand these choices and their implications.

While removing emissions of the VOCs reduced the concentrations of peroxy radicals in all cases the response from OH is different. Removing the emissions of the CO and alkanes increases the concentration of OH. This is due to the reduction in the OH sink and the lack of influence of these compounds on other aspects of the chemistry system. However, removing the emissions of the alkenes, isoprene and aromatics has the opposite effect, decreasing the OH concentration. This is on the face of it surprising given the importance of these compounds as OH sinks (Section 2.8.7). However, all of these species are able to rapidly return radicals on their oxidation by OH. So even through their removal leads to a substantial reduction in the OH sink, it leads to an even larger reduction in the OH sources. Hence overall the OH concentration drops. This is most notable for the aromatics. These are oxidized by OH to form dicarbonyl species that are highly photolabile. They readily photolyse to produce more radicals, resulting in an amplification of radical production. Thus removal of aromatics reduces radical production far more than it decreases the OH sink, and so the OH concentration increases substantially when the aromatics are emitted.

From this analysis both isoprene and aromatics are important for the O_3 concentration in Beijing. Next, the influence of local (city level) versus regional (whole domain) isoprene is explored and then the influence of the different aromatic species (benzene, toluene etc) on O_3 production is assessed.

2.11.1 Local versus Regional Isoprene

Figure 2.29 shows the importance of isoprene emissions for O_3 concentrations in Beijing. This reflects the large emissions rate and the short tropospheric lifetime (1-2 hours) of isoprene. The influence of biogenic species on O_3 production is well established especially in the SE United States (A. M. Fiore, Horowitz, et al., 2005; Yuzhong Zhang and Y. Wang, 2016; Schwantes et al., 2020), but their role in China is less well established (Xie et al., 2008; Z. Mo et al., 2018).

One important question here is location of the isoprene emissions having the influence. Is it the isoprene emitted within the city, or from the isoprene emitted in the regional forests surrounding Beijing? This is difficult to answer definitely as we do not have observations of the atmospheric composition in the surrounding area to assess the model performance against. Local Beijing emissions of isoprene were increased to match the observations dataset (Section 2.6) but those outside of the region were kept at their default MEGAN values.

Figure 2.32 shows the impact of switching off the emission of "local isoprene" (defined by the region shown in Figure 2.5) or all isoprene in the model. Switching off all isoprene emissions leads to a reduction of 12% in the median 4pm O_3 concentration (see also Figure 2.29), whereas switching off only the local isoprene reduces the median concentration by only 3%. Thus much of the influence of isoprene on O_3 on Beijing isn't local but on the regional scale.

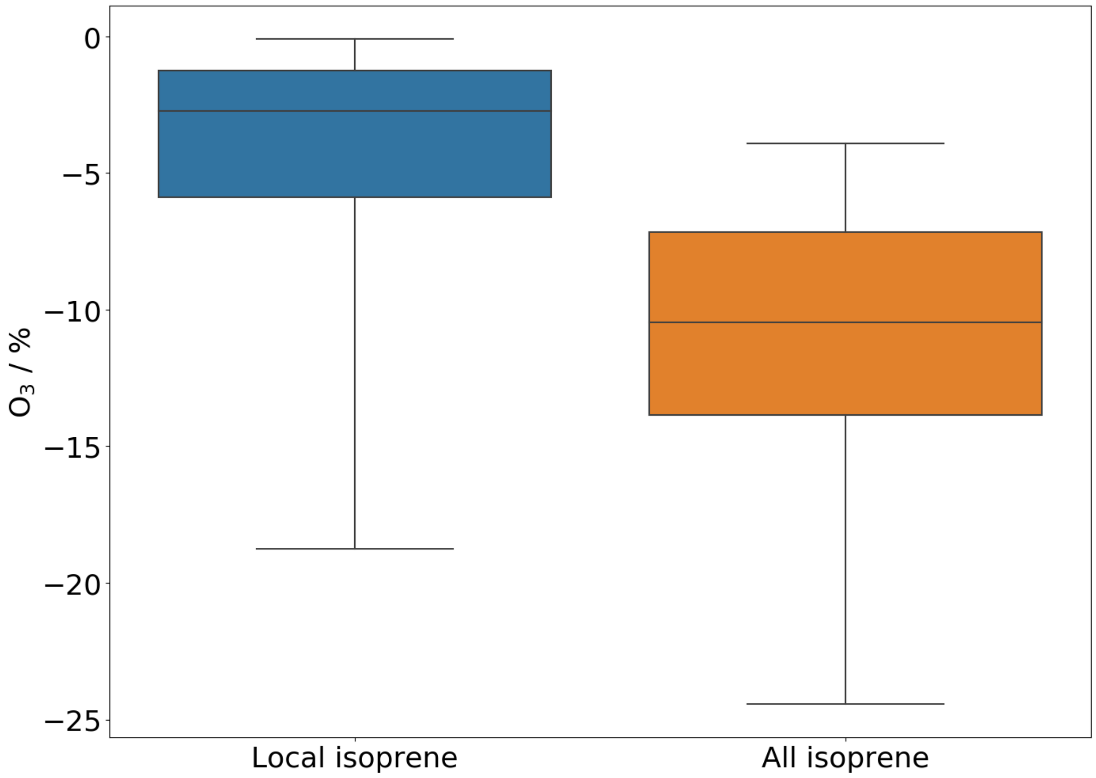

Figure 2.32: Box and whisker plot of change in base O_3 concentration with the removal of all domain isoprene emissions compared to the removal of local (isoprene originating from the Beijing urban area) emissions.

In Figure 2.33, a map of the mean 4 pm percentage change in O_3 from eliminating all isoprene emissions can be seen. The hot spots for O_3 influence line up with urban centres. This reflects that these are the regions where O_3 is being made, as they are the regions with high NO concentrations. Beijing itself displays the highest observed relative reduction in the region (-12%). It is likely Beijing is most sensitive as emissions were scaled by 3.5 times over the city to match observations. As was seen in Figure 2.32 only a quarter of this reduction was a result of local (inside the black box) emissions. As isoprene only has a short lifetime (2 hours at 5 x 10^{-6} OH molecules), the regional isoprene emissions must be having an impact on Beijing O_3 via both the formation of advectable species such as CH_2O, methacrolein and methyl vinyl ketone and in increased transport of O_3 concentration into the city. From a policy perspective it is unlikely that meaningful improvements to O_3 concentration could be achieved via targeting Beijing's foliage (such as growing alternative species of trees). However, this might be desirable for other air quality reasons such as reducing particulate matter and altering street-level wind dynamics.

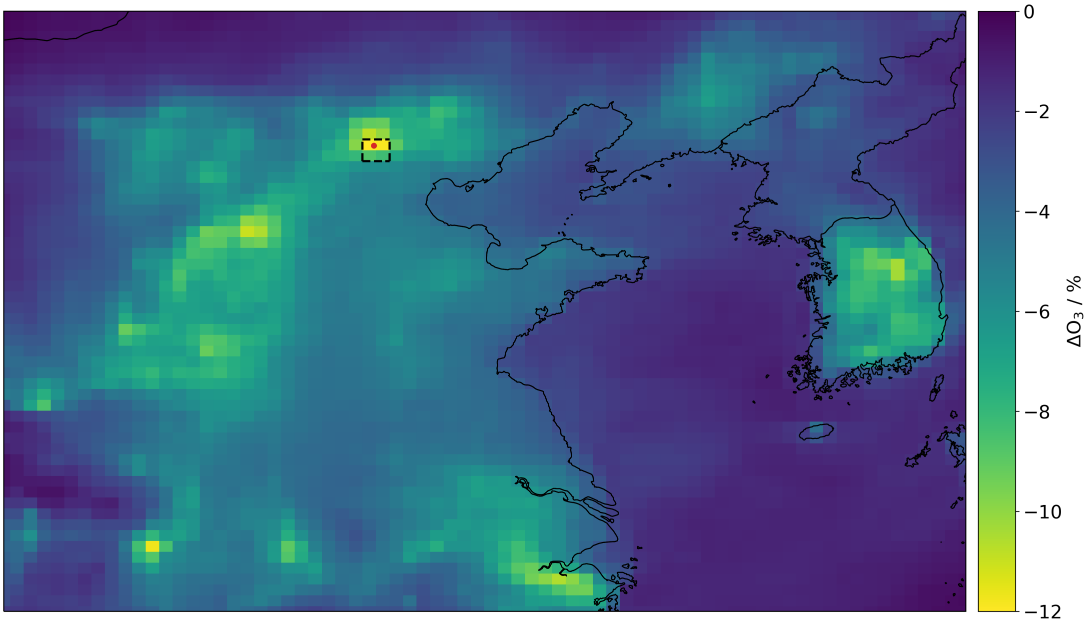

Figure 2.33: Map of the change in mean 4 pm O_3 concentration with the removal of isoprene emissions.

2.11.2 Influence of the individual aromatic species

The aromatic class of compounds provide the largest influence on the O_3 concentrations (Figure 2.29). However this class consists of six different compounds. In order to understand further the impact of the different members of this class of species individual perturbation studies are performed where the emissions of the individual aromatics are switched off (Figure 2.34). Although in the model xylene was split into two species (o-xylene and mp-xylene), all isomers are combined for the analysis.

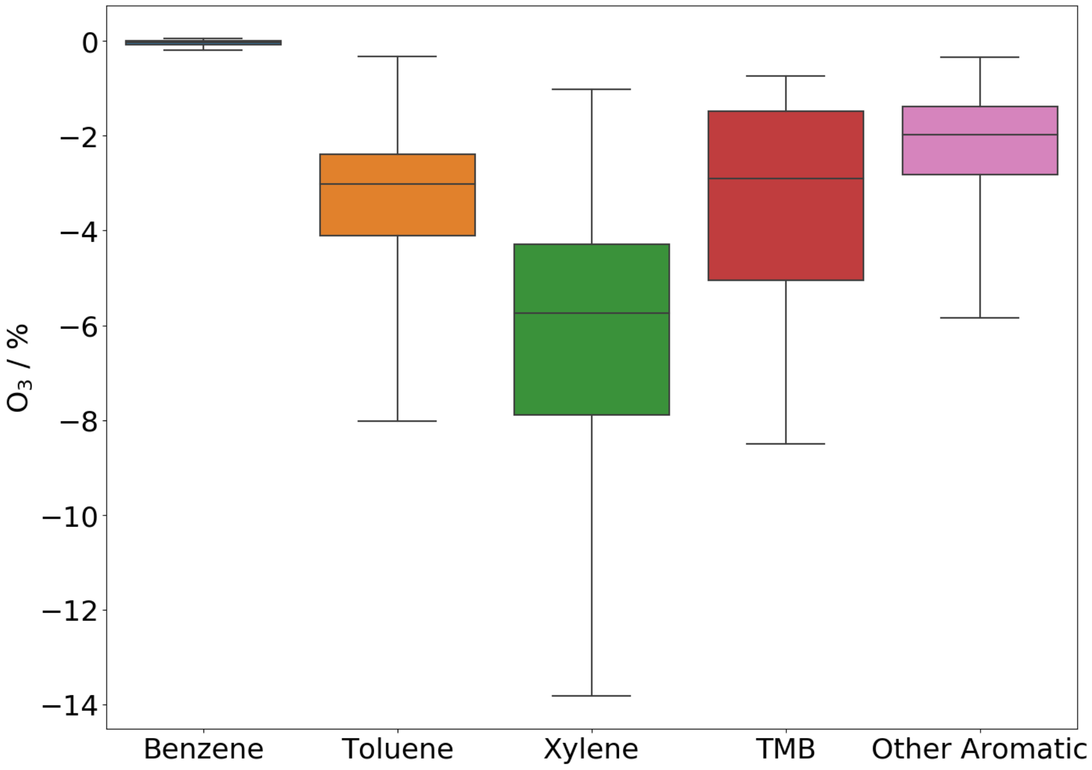

Figure 2.34: Box and whisker plot of change in base O_3 concentration with the removal of emissions from each aromatic species.

Switching off the benzene emissions has minimal effect on the O_3 concentration. Its relatively low concentration (0.5 ppbv see Figure 2.6) and its long atmospheric lifetime (10 days at 1x10^{-6} molecules cm^{-3} OH) results in it not having a particularly large influence on local O_3.

Removing toluene, xylene and TMB emissions has a more profound impact on O_3 with mean concentrations reducing by ~3 and 6, 3%, respectively.

The xylenes have the most profound impact despite their concentrations (0.5 ppbv) being similar to those of benzene. Their rate constant with OH is 20 times faster than the rate constant between benzene and OH, resulting in them having a much more significant influence on local O_3.

Removal of trimethylbenzene (TMB) emissions reduced O_3 by ~3%, comparable to the change seen in toluene. However, as the chemistry implemented to represent TMB is the same as that of mp-xylene there is increased uncertainty about this conclusion. Further work is required to assess whether this methodology is suitable or whether specific reduced TMB scheme are necessary to represent the chemistry of aromatics fully.

The last aromatic species was the "other aromatic" group, which used ethylbenzene as its surrogate. As this class did not have any direct measurements, the inventory used was CEDS (Section 2.4.3) with no scaling. As with TMB, the xylene chemistry was used to represent the oxidation chemistry of the other aromatics. Thus the influence of this class of compounds is uncertain, due to its emissions, speciation and chemistry.

It is informative to look at the concentrations of these compounds calculated without any scaling of the CEDS emissions. Figure 2.6 shows the diurnal mean concentrations of these compounds with their default emissions. The concentration of the xylenes are very much higher than those observed, with day time concentration around twice as high as observed. Thus without the observations, it might be concluded that these species had an even more considerable influence on O_3 production than they have in reality. Further work to resolve these larger aromatics and development of their chemistry is needed to identify whether they are needed or not.

Given these simulations, approaches that target reduce xylene concentrations in Beijing would appear to be the most useful at reducing the O_3 concentration.

2.12 Conclusions

In this chapter, the GEOS-Chem CTM was run nested at 0.25° x 0.3135° resolution over at East China with tuned emissions to explore the sensitivity of the modelled O_3 to NO_x and VOC emissions.

The tuning assumed that all of the model failures were due to shortcomings in emissions. Failures resulting from the meteorology or chemistry were not directly explored. Future work could involve fixing the boundary layer to observed values either from surface lidar or via satellite. Scale factors on emissions for many species were substantially, indicating the CEDS, Tzompa-Sosa, Xiao and MIX inventories do not represent Beijing emissions well. This may have been that substantial changes occurred in the emissions from Beijing between the year for which the inventory was developed (2010 for MIX and 2014 for CEDS) or that the information and assumptions going into the emissions inventory was flawed. Additionally, the default GEOS-Chem emission diurnal profiles, required a greater day to night transition to fit the observations. It is not clear why this would be the case but may reflect Chinese working practices compare to the North American or European ones used to develop the diurnal factors. However, it is likely that much of this profile is related chemistry weakness within the model. After tuning, the primary species were much close to the observations.

The model performance for the secondary species, was in general poor. Some of this was due to the underestimate of HONO concentrations which is the dominant source of OH in Beijing (Whalley et al., 2020). Thus the source of radicals was likely underestimated by the model. There was some compensation from an overestimation of $J(O^1D)$ (which would increase the primary O_3 source of radicals) but in general it appeared that the model's radical sources were underestimated. This is evident in the model's total radical concentration ($OH+HO_2+RO_2$) where the model grossly underestimates, most notably for the RO_2. The model's balance between HO_2 and RO_2 whilst favouring RO_2 was still significantly different from observations. This likely reflects the model's underestimate of OH reactivity suggesting a missing source of reactive carbon in the model. However, the model simulation of CH_2O was reasonably close to that observed, which would suggest that any missing reactive carbon species was not having a significant impact on the smaller carbon species. The model's ability to simulate the balance between NO and NO_2 was also poor with little explanation of why. The observations suggest a lower NO concentration but a higher peroxy radical concentration. Although it seem unlikely that the high peroxy radical concentration is the cause, a mechanism that would reduce the NO concentration would allow for a higher peroxy radical concentration without causing the O_3 concentration to rise too high.

These failings do not appear to be restricted to the GEOS-Chem model. Many similar failings were found by Whalley et al., (2020) using a constrained box model with a very detailed chemistry. Although not subject to some failings due to the constraints (underestimation of HONO, over estimation of NO, overestimation of $J(O^1D)$), the problems associated with underestimating reactivity and RO_2 are still present. It would be surprising if other models, whether box or chemical transport would show a significantly better performance that shown by the model here. Other models may have similar problems but due to a lack of evaluation they may not be so evident.

Given these problems the O_3 simulation appears to be surprisingly good. This likely reflects compensating errors: an underestimate of RO_2 together with an overestimate of NO, an underestimate of the HONO source of radicals with an overestimate of $J(O^1D)$ etc. Additionally it shows O_3 is relatively insensitive to the fine detail once the main drivers are reproduced well. However, the performance is likely as good as could be expected give our understanding of the chemistry.

The modelled O_3's sensitivity to NOx and VOC emissions was explored and the model was found to be in a VOC limited regime. There was some day to day variability, but in general, the optimal way of reducing O_3 in Beijing was to reduce VOC emissions. Reductions in NO_x would likely increase mean O_3 concentrations, but peak O_3 would likely to increase by less.

The most significant classes of species for O_3 production were isoprene and aromatics. They both lead to substantial increases in OH, HO_2+RO_2 and O_3 concentrations. Isoprene emissions from outside of the city was found to be a significant source of O_3 within the city, and xylenes were found to be the most important class of aromatics in generating O_3.

With O_3's relationship between VOC and NO_x emissions explored in this chapter, the impact of HO_2 uptake on O_3 concentration will be investigated in the next.

Chapter 3

The affect of hydroperoxyl radical uptake onto aerosols for ozone concentrations over Beijing

3.1 Introduction

In Chapter 2, the focus was on how NO_x and VOCs influence the production of O_3 in Beijing. This kind of approach dates back to the early analysis of O_3 in cities such as Los Angeles (Haagensmit, 1952; Pratapas and Calcagni, 1983; Seinfeld, 1989). More recently the impact of particulate matter on O_3 production has been discussed. For many years O_3 and particulate matter pollution were considered to be separate problems. However more recently it has become evident that they are to some extent interwoven (Mao, Fan, et al., 2013; S. E. Bauer et al., 2004; Jacob, 2000).

Historically the impact of particulates on air pollution has been thought to mainly occur through the hydrolysis of N_2O_5 on the surface of particulates during the night (Riemer et al., 2003; W. L. Chang et al., 2016). However, more recently it has become evident that uptake of HO_2 onto aerosol can have a significant impact as well (Thornton and Abbatt, 2005). By providing a sink for RO_x radicals this mechanism can reduce the amount of O_3 that an air mass can produce. It thus competes with the other sinks for radicals, the OH+NO_2 reaction and the self reaction of RO_x species (HO_2+HO_2, RO_2+HO_2, RO_2+RO_2 etc.)

Beijing and the surrounding area has been a test bed for this more recent analysis. However, some studies (K. Li et al., 2019; Xu et al., 2012; Lou, H. Liao, and B. Zhu, 2014; J. Li et al., 2018) have claimed that the impact on particulate matter on O_3 is significant, whereas others claim little of minimal impact (Z. F. Tan, Hofzumahaus, et al., 2020).

The uptake rate of a molecule to a surface can be approximated by the first order loss rate coefficient given by Equation 3.1 (Schwartz, 1984).

$$k = A\left(\frac{r}{D_g} + \frac{4}{\gamma\omega}\right)^{-1} \qquad \text{(Eq. 3.1)}$$

Where r (cm) is the aerosol particle effective radius, D_g ($cm^2\ s^{-1}$) is the gas phase diffusion coefficient, γ (unitless) is the reaction probability (also often referred to as the reactive uptake coefficient), ω ($cm\ s^{-1}$) is the mean molecule speed (calculated as $(8kT/\pi m)^{1/2}$), and A is the aerosol surface area concentration ($cm^2\ cm^{-3}$). Under most situations the impact of diffusion is small and so the dominant terms in the equations are the surface area concentration (A) and the reactive uptake coefficient (γ).

There is significant uncertainty on the value of γ. The literature contains values ranging from <0.002 (Taketani, Kanaya, and H. Akimoto, 2008) to 0.4 (Taketani, Kanaya, Pochanart, et al., 2012). Studies have found that particle radius (J. Guo et al., 2019), humidity (Lakey et al., 2015), temperature (Thornton and Abbatt, 2005) and composition (notably the copper content (J. T. Lin and McElroy, 2010)) could all play a role in governing the γ value.

Particulate surface area concentration (A) also plays an important role for determining the impact of heterogeneous processes. Model performance for the simulation of aerosol often focuses on mass below a critical threshold (typically 1, 2.5 or 10 μm) due to the health impacts (Shiraiwa et al., 2017), or on number concentration of particulates greater than critical sizes due their importance for cloud formation (Kulmala et al., 2013). However, the performance of models in the calculation of aerosol surface area is not often reported. In this chapter the impact of HO_2 uptake onto the surface of aerosols, on O_3 concentrations is explored within a model in the context of the APHH campaign described in the Chapter 2.

3.2 The significance of HO_2 uptake on O_3

The model (after tuning) described and evaluated in Chapter 2 is now used to explore the sensitivity of O_3 concentrations to HO_2 uptake The default HO_2 uptake coefficient (γ) used in GEOS-Chem is 0.2 (Mao, Fan, et al., 2013). This falls in the middle of the large range of experimentally derived values (0.002-0.4). Laboratory measurements of the uptake coefficient made with Beijing aerosol (Taketani, Kanaya, Pochanart, et al., 2012) gives a value in the range 0.09 – 0.40. The model value is in that range.

It is plausible that other peroxy radicals such as the methyl peroxy radical could also undergo reactive uptake. However, there are minimal laboratory measurements of this reaction, and thus any analysis would be highly speculative. Further, HO_2 is the simplest peroxy radical with the lowest molecular weight resulting in the highest collision rate. Thus organic peroxy radicals with a reduced collision rate coupled with an expected lower Henry's law coefficient are likely to have a small relative impact on radical loss.

A summary of the model performance for aerosol surface area along with aerosol mass and relative humidity is shown in Figure 3.1

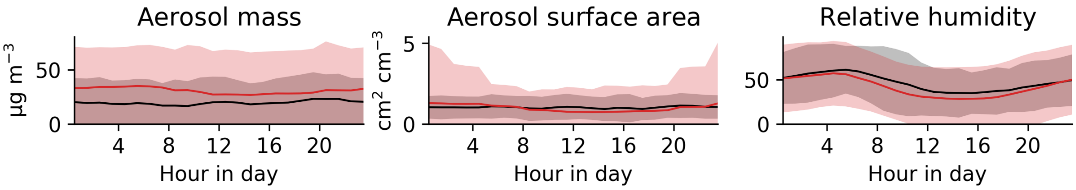

Figure 3.1: Summary of the model (red) aerosol mass and surface area and relative humidity mean diurnal performance compared to observation (black). One standard deviation is displayed by the shaded area.

The model underestimates the surface area by approximately 12% averaged across the day but has a day time (8 am to 8 pm) underestimation of 24%. The cause of this underestimate is unclear, and there is a lack of published aerosol surface area performance over China in chemical transport models. Conversely, the aerosol mass for the measured species was overestimated by 32% which is in agreement with published values for GEOS-Chem for the same region and time period (Dang and H. Liao, 2019). The model may be overestimating the aerosol radius as was previously found in a study on aerosol optical depth bias (Latimer and R. V. Martin, 2019). Relative humidity was not the cause of a radius over prediction as the model slightly (4%) underestimated the relative humidity during the campaign. The model uses a single global value for each aerosol type. If the radius was too small for a specific species this could account for the problem.

As well as the biases in the aerosol sink for HO_2, there are significant model biases in the other radical sinks ($OH+NO_2$, RO_x+RO_x) which need to be considered. A summary of the model radical and NO_2 performance is shown in Figure 3.2

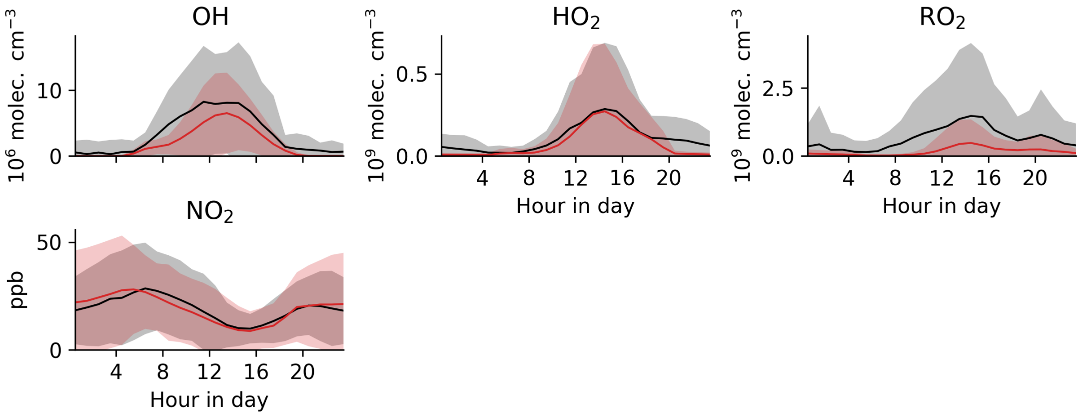

Figure 3.2: Summary of model (red) radical and NO_2 mean diurnal performance compared to observation (black). One standard deviation is displayed by the shaded area.

As discussed in Section 2.8, the model's ability in simulating the other components of the RO_x radicals sinks is problematic. For the $OH+NO_2$ route, the model underestimates both the OH (22% at noon) and NO_2 (14% at noon) concentration during the daytime. For the RO_x+RO_x route, the model slightly underestimates HO_2 (5% at noon) but significantly underestimates the RO_2 (72% at noon). For the radical uptake route the model underestimates surface area (12%) coupled with the slight 5% HO_2 underestimate.

Overall there are significant biases in the performance of the model for the RO_x species, which requires caution when considering the fidelity of the modelled O_3 response to HO_2 uptake. However, there is also utility in exploring the model sensitivity to this parameter and then returning to these uncertainties once an understanding of the model sensitivity has been achieved.

Several methods could provide an assessment of the impact of HO_2 uptake onto aerosols. However, the simplest is to multiply the reactive uptake coefficient used in the model by a constant. Assuming that the diffusive component of Equation 3.1 is small, this is equivalent to increasing the aerosol surface area by the same fraction. Three additional model simulations are therefore performed, which multiply the γ value used by 2, 0.5 and 0.

Figure 3.3 shows the time series of the change in O_3 concentration from multiplying the HO_2 uptake coefficient by 2, 0.5 and 0.0.

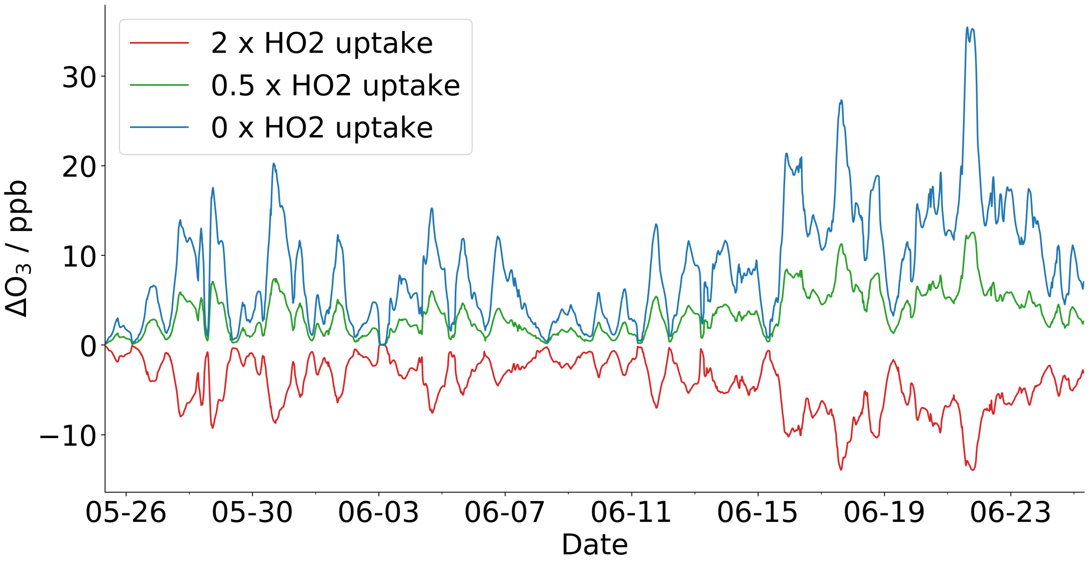

Figure 3.3: Hourly averaged (mean) time series of O_3 change resulting from different HO_2 uptake coefficients relative to the the base model with a γ of 0.2

Switching off the uptake leads to an increase in the O_3, with an increase of 13% (83 ppbv to 94 ppbv) in the mean 4 pm O_3. Doubling the HO_2 uptake reduces 4 pm O_3 concentration by around 8% (83 ppbv to 76 ppbv), whereas halving it increases O_3 by +5.5% (83 ppb to 88). As with reducing NO and VOC emissions (Section 2.9), there is significant day to day variation, with some days showing only a 1% change in O_3 while the highest response days are >10%.

The increase in O_3 for the removal of HO_2 uptake is a significant change to the composition. Switching off isoprene emissions in Section 2.11 led to a 11% reduction in the mean 4pm O_3 compared to the 13% increase found from removing HO_2 uptake.

It is obvious that in the model the uptake of HO_2 onto aerosol is an important processes for controlling the concentration of O_3 in Beijing during the 2017 field campaign. This tends to confirm the work of K. Li et al., 2019, but seems to contrast with the work of Z. F. Tan, Hofzumahaus, et al., 2020. In the next section the balance of radical loss processes is assessed and an effort is made to try and rationalise the different conclusions made by these papers about the significance of this process.

3.3 Radical loss pathways

As discussed in Chapters 1 and 2, the dominant radical removal pathway dictates the photochemical regime of an air mass and so the strategy for controlling its O_3. Conventionally, an air mass is split into either a VOC limited regime (if the NO_2 + OH pathway dominates radical loss), or a NO_x limited regime (if RO_x + RO_x dominates). However, if the uptake of HO_2 onto aerosol plays an important role in determining the radical termination perhaps different classifications should be used.

The size of each radical loss pathway can be calculated in the model and from the observations based on the concentration of OH, HO_2, RO_2, and NO_2 as well as the temperature, relative humidity, aerosol surface area and radius.

The equations to calculate each pathway were taken from GEOS-Chem (JPL recommended). For calculating the HO_2 uptake a γ value of 0.2 was used for both model and observation. Methyl peroxy radical (MO_2) was used as a surrogate for RO_2. The use of MO_2 as a surrogate for all RO_2 could lead to a slight overestimation of RO_2 reaction rates but should be adequate for this analysis. Thus RO_x + RO_x was made up of:

$$\mathrm{RO_2 + HO_2} \tag{R 3.23}$$

$$\mathrm{RO_2 + RO_2} \tag{R 3.24}$$

$$\mathrm{OH + HO_2} \tag{R 3.25}$$

$$\mathrm{HO_2 + HO_2} \tag{R 3.26}$$

$$\mathrm{OH + OH} \tag{R 3.27}$$

Figure 3.4 shows the time series of the rate of radical loss for each hour of the field campaign. Due to gaps in the observation data, not every day of the campaign can be calculated for the observations. There is a strong diurnal cycle reflecting that for short lived species such as these, the sink term needs to balance the source term, which is predominantly photo-chemical. The model is biased low on radical removal throughout the campaign reflective for the underestimation in NO_2, OH and RO_2, and of the low calculated contractions of HONO (Section 2.8)

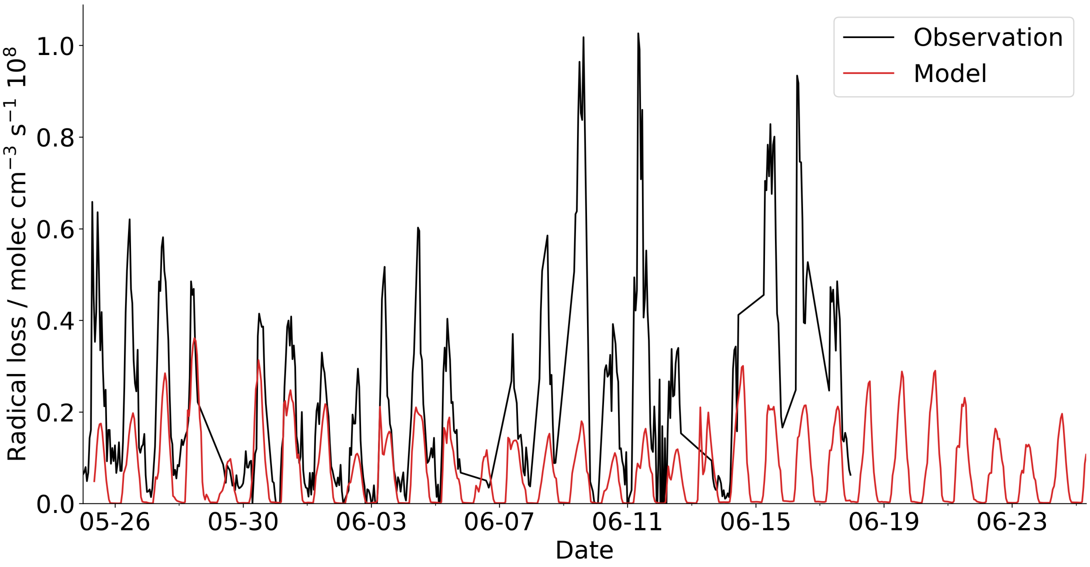

Figure 3.4: Hourly averaged (mean) time series of the modelled (red) and observed (black) radical loss. Where total radical loss is (NO_2+OH) + (HO_2 uptake) + (RO_x+RO_x).

To simplify the analysis, the noon time rates for each day are extracted and expressed as a fraction of the total loss rate For each day, there are three potential components (RO_x+RO_x, HO_2 uptake, and NO_2 + OH) which add together to give the total loss rate. These can be plotted on a ternary plot (Figure 3.5) which represents the fraction of the noon time radical loss which passes through each of the 3 radical loss routes. The days which could not have the observed radical loss calculated due to the missing data have been included in the model plot as blue dots.

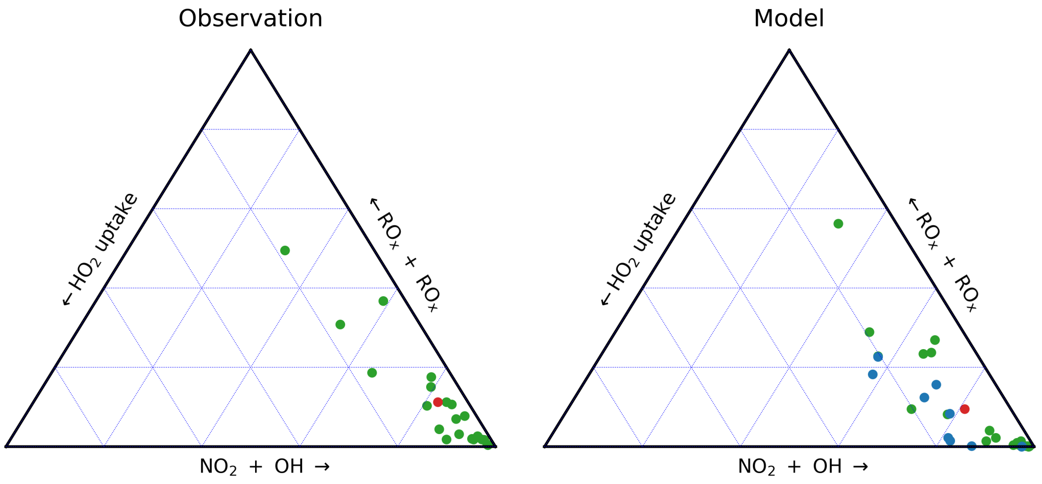

Figure 3.5: Ternary plots representing the relative contribution of each radical loss pathway to the total radical loss for model and observation. An air mass with 100% of loss via RO_x+RO_x would lie in the top corner, lower left corer would be all HO_2 uptake, and lower right all NO_2 + OH. Red dot shows the campaign mean location

Despite the issues with the model's ability to simulate the OH, HO_2 and RO_2 radicals, the mean balance between the sink processes (red dot) is in a very similar location to that found for the observations. On average the NO_2+OH is dominant sink process with a ratio of 0.81:0.095:0.095 (OH+NO_2:RO_x+RO_x:HO_2 uptake) in the observation and 0.83:0.11:0.6 in the model mean. Some day to day variability can be seen with points spreading out quite far from the mean. In general this spread in both the observed and modelled is from the NO_2+OH corner towards the RO_x+RO_x corner.

In order to explore this more fully the model simulation of the campaign is repeated multiplying the anthropogenic NO_x emissions by 1, 0.75 and 0.56 and independently multiplying the HO_2 uptake coefficient by 0.55, 1.00 and 1.80 (the equivalent to changing the aerosol surface area), to give a total of 9 simulations. Figure 3.6 shows impact of these changes on the radical loss pathways in the ternary plots. The middle-left panel, highlighted in red, is the same as the model panel in Figure 3.5, with the campaign mean position indicated by the red dot. For each plot the red dot indicates the mean split in the radical sinks for the campaign with the line leading from that point to the mean point in the base case simulation.

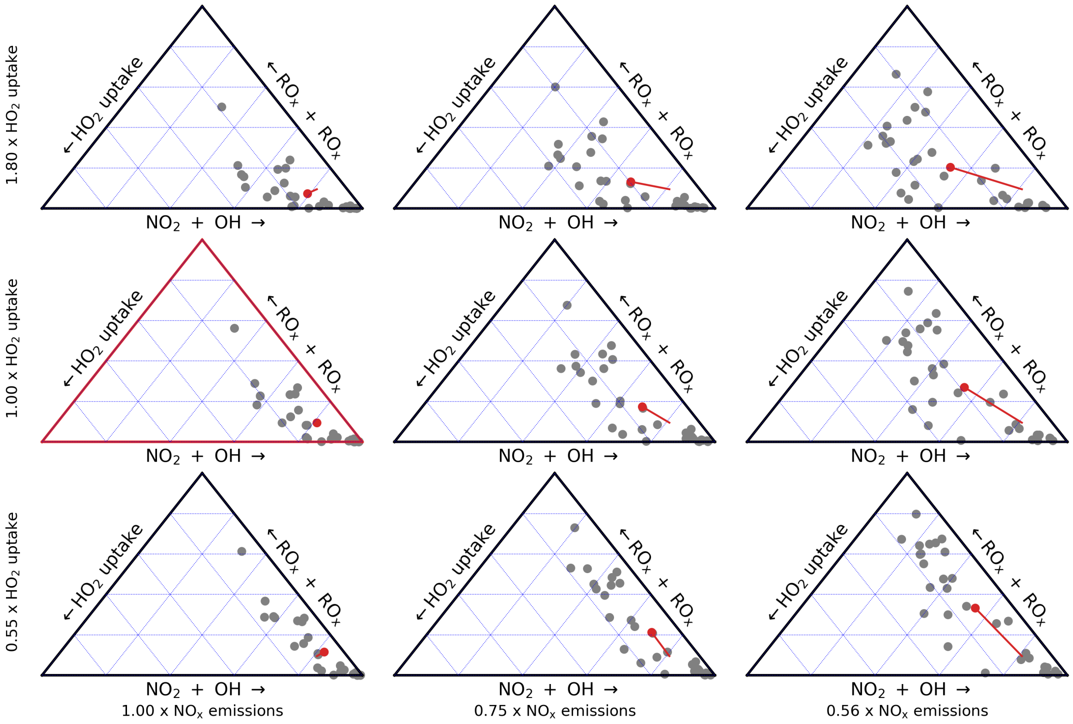

Figure 3.6: Ternary plots of the relative contribution of each radical loss pathway with changing NO emissions and HO_2 uptake. Red dot indicates the campaign mean, and the red line identifies the change in mean location relative to the base model mean.

Decreasing the NO emissions (moving from the base simulation rightwards) leads to the points moving upwards towards the RO_x+RO_x (top) corner of the diagram with a tendency to also increase the importance of the heterogeneous uptake. With reduced NO emissions, HO_2 and RO_2 concentrations increase. The loss is linear with the radical concentration for the heterogeneous uptake, but is the square for the self reaction. Thus as NO emissions drop RO_x + RO_x increases faster than the HO_2 uptake so much of the travel (indicated by the red line) under reduced NO emissions is thus from the NO_2+OH corner to the RO_x+RO_x corner, rather than from the NO_2+OH corner to the HO_2 uptake corner.

Changes to the uptake coefficient (moving to the upper and lower row of the plot) in the model is equivalent to changing the surface area. For the default NO emissions (the left most 3 plots), the changes are relatively small compared to the changes from changing the NO_x emissions. At lower NOx concentrations, the change is larger. Under the standard emissions conditions NO_2+OH dominates the sink. Changes in the uptake of HO_2 will not have a significant overall effect on this. However, as NO emissions reduced the dominance of NO_2+OH will become less and changes in HO_2 uptake can have an impact on the balance of radical loss.

From the results in this section, it can be seen that HO_2 uptake can be an important radical loss pathway, particularly when NO emissions are reduced from their current level. The sensitivity of the model to heterogeneous chemistry thus depends upon the NO_x concentration. To explore this further a large number of model simulations are performed to explore the relationship between changes in NO emissions and changes in the uptake of HO_2.

3.4 NO_x / HO_2 Uptake isopleths

In Chapter 2, O_3 isopleths were used to investigate the effect of changes in VOC and NO emissions on O_3 concentrations. These are a traditional tool used to understand the appropriate photochemical regime for deciding policy for reducing O_3 concentrations. However, it is clear that changes in aerosol surface area can have an influence on O_3 production. A new way of exploring the photochemical environment is needed to be able to explore O_3 production taking HO_2 uptake into account. In this section O_3 isopleths will be produced which vary HO_2 uptake instead of varying the VOC emissions

The method for generating these isopleth plots is the same as described in Section 2.10. The GEOS-Chem model was used to construct O_3 isopleths for Beijing by varying the NO (anthropogenic only) and reactive uptake coefficient (γ) independently. The scaling was symmetric in log space with nine logarithmically distributed steps between 0.1 and 10 (0.10, 0.18, 0.32, 0.56, 1.0, 1.8, 3.2, 5.6, 10.0) for both NO and γ. This "coarse grid" covers a broad range of emission scaling with a focus on chemical understanding. A finer grid with nine logarithmically distributed steps between 0.32 and 3.2 (0.32, 0.42, 0.56, 0.75, 1., 1.33, 1.78, 2.37, 3.2) was nested inside the coarse grid to offer greater chemical resolution over the policy relevant range. The resulting isopleths are constructed from the 137 model runs.

Figure 3.7 shows the mean modelled 4 pm O_3 as a function of the scaled HO_2 uptake coefficient (equivalent to scaling the aerosol surface area concentration) and the NO emission. As with the NO-VOC isopleth (see Figure 2.23) at very high NO emissions (top of the graph) O_3 concentration are low due to NO titrating away O_3. As NO emissions are decreased, this titration effect is reduced, and as the radical concentrations increase, due to lower radical loss from $OH+NO_2$, the O_3 concentration increases. This is a typical response in a VOC limited regime. As NO emissions reduce further they reach a peak and then drop, as the NO needed to make O_3 reduces faster than the radical concentration increases. This is typical of the NO_x limited regime.

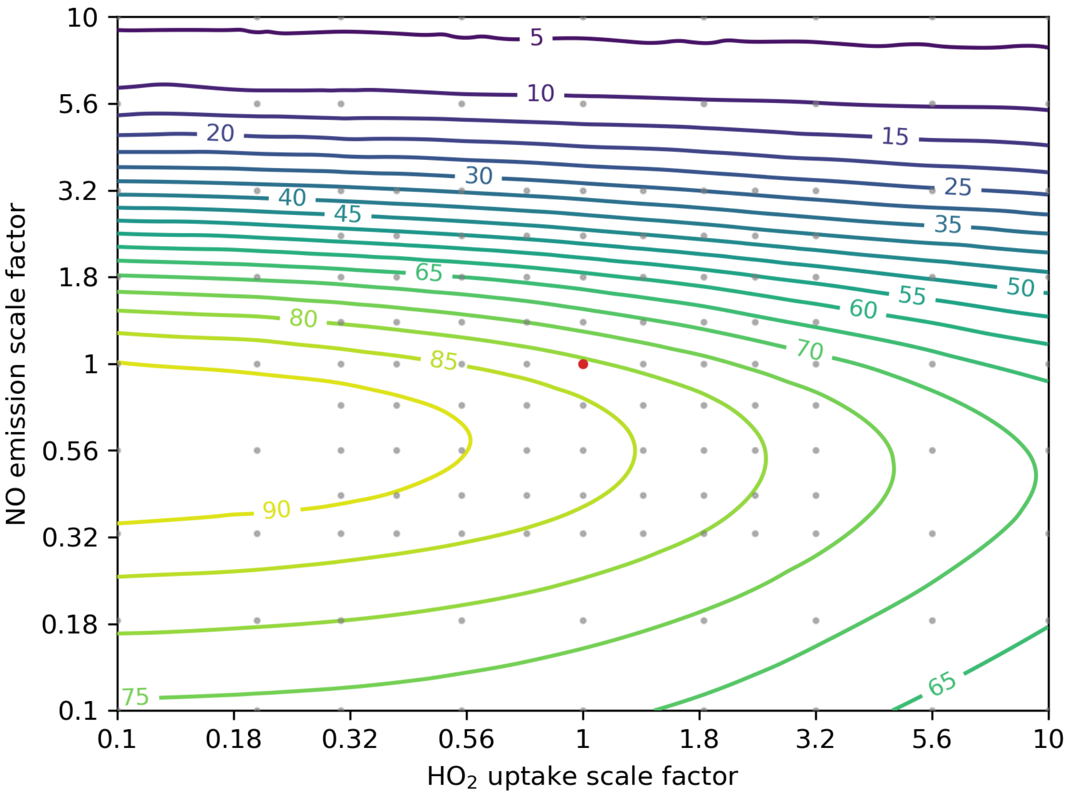

Figure 3.7: Averaged (mean) modelled 4 pm O_3 concentration isopleth for the duration of the campaign. Contours show O_3 concentrations with changing NO emissions (x-axis) and HO_2 uptake (y-axis). Red dot indicates location of the base (unscaled) run. The red lines split the there HO_2 uptake sensitivity regions.

Changing the HO_2 uptake (akin to changing the aerosol surface area), modulates this signal. At all NO emissions, there are higher O_3 concentrations at lower value of uptake, and lower O_3 concentrations at higher uptake. There is a small modulation of the position of peak O_3 at different values of HO2 uptake, with peak O_3 occurring at lower NO emissions, when HO_2 uptake increases.

The sensitivity of the model to changes in the HO_2 uptake (aerosol surface area) can be better seen in Figure 3.8. In the Figure, the axe are the same as Figure 3.7 and shows the change in O_3 for a given 20% change in HO_2 uptake. The results are averaged over bins of a 20% change in NO emissions.

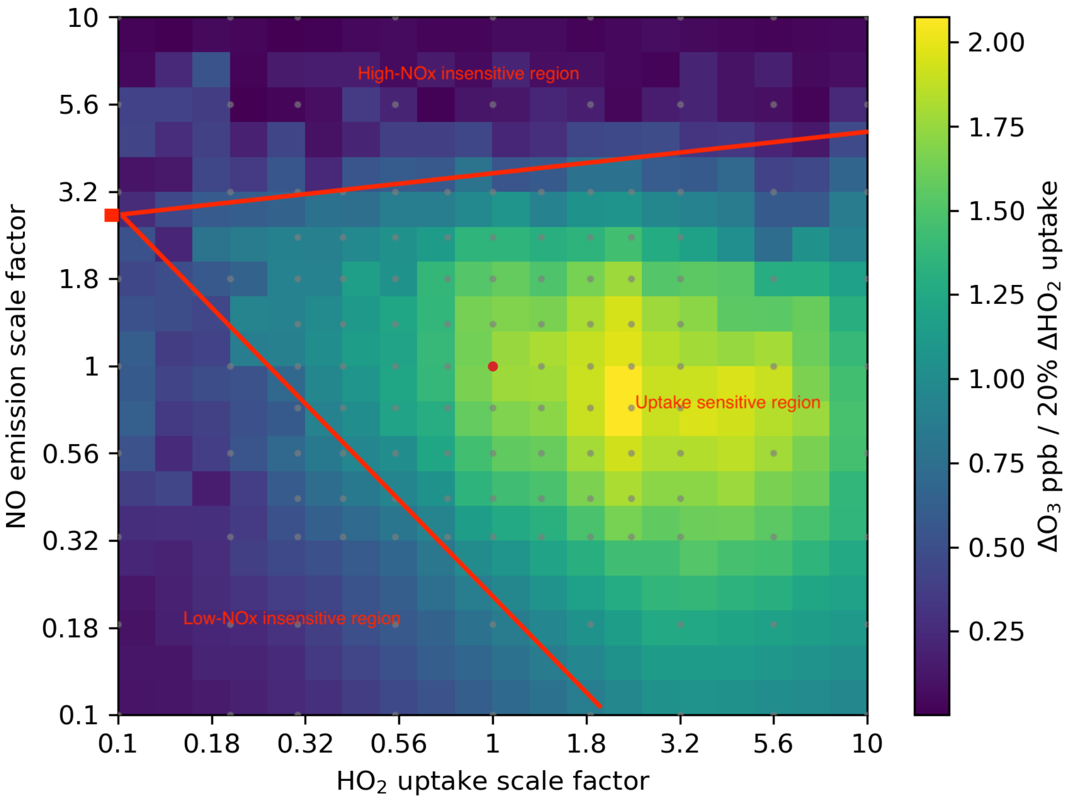

Figure 3.8: Campaign average modelled (mean) 4 pm O_3 sensitivity to a 20% change in HO_2 uptake averaged (mean) over 20% changes in NO emission. Red dot indicates location of the base (unscaled) run.

Looking at the O_3 sensitivity in this way there are three regions of the graph:

- **High NO_x, insensitive** region at high NO emissions, where the model shows little sensitivity and the O_3 isolines in Figure 3.7 are horizontal.
- **Uptake Sensitive** region in the middle, where changing the uptake leads to changes in the O_3, and the isolines run vertically.
- **Low NO_x, insensitive** region towards the bottom left of the graph, where the model become less sensitive to uptake and the isolines in Figure 3.7 again run horizontally.

In order to explain the processes controlling the radicals in these regions, isopleths of radical concentration are drawn up. The 4pm daily O_3 (Figure 3.7) is essentially an integrated quantity reflecting the total daily radical chemistry. However, the radical chemistry is most active around solar noon and so these radical isopleths are drawn up for the average 12 pm to 1 pm period.

Figure 3.9 shows the mean noontime RO_x (=RO_2+HO_2+OH) concentration isopleth. At high NO emissions, RO_x concentrations are low as O_3 the concentration is low, leading to low production of radicals. At moderate NO emissions, RO_x concentrations increase but show little sensitivity to HO_2 uptake (relatively horizontal isolines). In this situation most of the radical (RO_x) loss is due to the NO_2+OH reaction. Changing HO_2 uptake would result in only small changes in the total RO_x loss and so there is little sensitivity to HO_2 uptake. At low NO emissions there is significant sensitivity to HO_2 uptake (vertical isolines). In this regime there is a balance between HO_2 uptake and RO_x+RO_x reactions as a radical sink and so there is sensitivity to HO_2 uptake. At very low values of uptake the isopleths are relatively far apart in the horizontal, but their spacing becomes closer to the middle, with them becoming further spaced out towards the higher uptake rates. This reflects the change in the sink processes. At low values of uptake, RO_2+RO_2 dominates, and an increase in the uptake rate would have only a small impact on the over all rate of loss of radicals. At moderate uptakes, the rate of RO_2+RO_2 and the HO_2 uptake are balanced. Increasing the HO_2 uptake under these conditions leads to a rapid change in RO_x concentrations. As HO_2 uptake increases yet further the lines space out once more. This reflects the balance between HO_2 and RO_2 in making up RO_x. As the HO_2 uptake increases the HO_2 concentration drops and becomes an increasingly small fraction of RO_x. Thus the uptake becomes less and less efficient in moderating RO_x concentration at higher uptake rates.

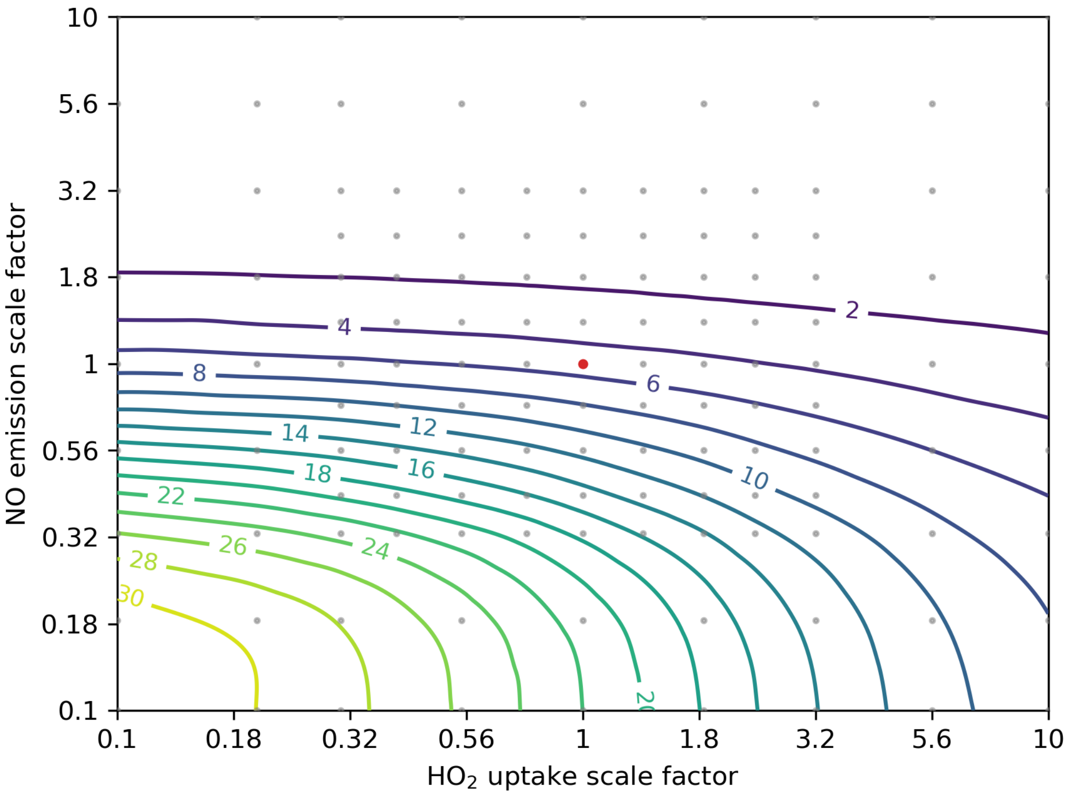

Figure 3.9: Averaged (mean) modelled 12 pm RO_x ([OH] + [HO_2] + [RO_2]) concentration isopleth for the duration of the campaign. Contours show RO_x concentrations with changing NO emissions (x-axis) and HO_2 uptake (y-axis). Red dot indicates location of the base (unscaled) run.

This now explains the 3 regions shown in Figure 3.7. In the *Aerosol Insensitive High-NOx* regime, HO_2 uptake is not an important sink for radicals compared to the NO_2+OH route and so the atmosphere is insensitive to changes in uptake rate. In the *Aerosol Insensitive Low-NOx* regime, HO_2 uptake is again not an important sink for radicals compared to the RO_x+RO_x sink, and so the atmosphere in not sensitive to aerosol uptake. Between these two regions, HO_2 uptake can effectively compete with the other sinks to remove radicals and so the model shows sensitivity to the uptake.

Figure 3.8 shows peak sensitivity occurs at roughly doubling of the base HO_2 uptake coefficient or twice the aerosol surface area and around 80% of the base model NO emissions. At uptake rates faster than that, sensitivity decreases again. This reflects the changing balance of HO_2 and RO_2. As uptake rates increases the HO_2 is removed by aerosol but the RO_2 is not. Ultimately a change in the uptake has little impact on O_3 production as all of the HO_2 has been removed. After a point the O_3 becomes less and less sensitive to HO_2

uptake, as more and more of the total peroxy radical pool is made up of organic radicals.

Given this new understanding, attention is turned to trying to understand the different conclusions that studies have reached about the importance of HO_2 uptake.

3.5 Radical loss isopleths

An alternative way of visualising this changing makeup of the radical sink in the model is to look at the ratio between the different radical sink mechanisms (HO_2 uptake, RO_x self reaction and NO_2+OH) occurring at the different emissions and HO_2 uptake conditions run in the model. Figure 3.10 shows the mean noon time ratio between these rates with blue representing the fraction of HO_2 uptake, green the fraction of RO_x self reaction and red the fraction of the NO_2+OH reaction. Dark red shows regions where NO_2+OH dominates, dark green where RO_x+RO_x dominates, and dark blue where uptake of HO_2 dominates. The yellow regions represents a balance between RO_x self reaction and the NO_2+OH reaction, magenta a balance between NO_2+OH and HO_2 uptake, and cyan represents a balance between ROx self reaction and HO_2 uptake. Also plotted is the campaign 4pm mean O_3 isopleths. The average base case model condition is indicated by the circle on the plot.

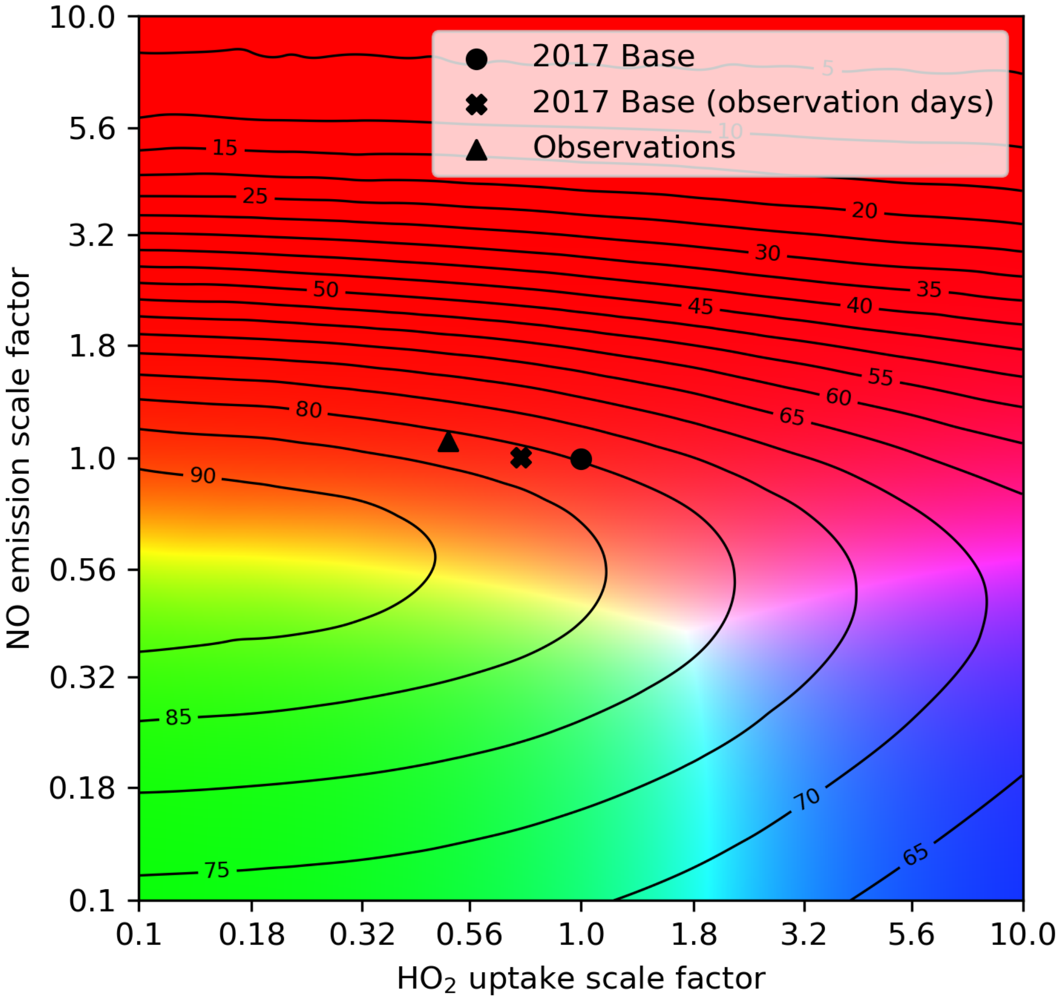

Figure 3.10: The colour coded dominant 12 pm campaign average (mean) radical loss pathway for a given scale factor on NO emission and HO_2 uptake, where red is NO_2+OH, blue is HO_2 uptake and green is RO_x+RO_x. Overlayed is the averaged modelled (mean) 4 pm O_3 concentration isopleth for the duration of the campaign. The round marker indicates the base run, while the triangle and cross indicate the calculated location of the observation and filtered (for days for which there are observations) model respectively.

Given the observations made during the campaign, it is possible to derive a similar observed campaign noon-time average split between the radical loss pathways from NO_2+OH, HO_2 uptake and RO_x self reaction. A degree of approximation is necessary for calculating the RO_x self reaction rate. It consists of the reactions, HO_2+HO_2, HO_2+ RO_2, RO_2+RO_2 and HO_2+OH. There is some degree of variation in the HO_2+RO_2 and RO_2+RO_2 rate constants with RO_2, so the methyl peroxy rate constant is used for simplicity. The location of that ratio between NO_2+OH, HO_2 uptake and RO_x self reaction is then found and marked on Figure 3.10 as the triangle.

The observational position (0.83 : 0.11 : 0.06 for NO_2+OH : HO_2 uptake : RO_x self reaction) is close to the mean model position (0.78 : 0.14 : 0.08) it is closer still when only model days that there are observational data are averaged (0.79 : 0.13 : 0.08).

3.5.1 Understanding other work

Another publication has discussed the sensitivity of O_3 in Beijing to HO_2 uptake. K. Li et al., 2019 described how the O_3 observed in Beijing increased by 8 ppbv between 2013 and 2017. They found that during the five year period in Beijing, the observed PM2.5 levels dropped 41%, while NO_x emissions dropped 23%. Their model was able to reproduce 4.8 ppbv of the observed change in O_3, and they attributed 4 ppbv (83%) of the change to the reduction in particulate matter reducing the HO_2 uptake. The paper gave their radical loss pathways positions (0.44 : 0.24 : 0.33 for NO_2+OH : HO_2 uptake : RO_x self reaction) for 2016, which could be then located on radical loss plot (Figure 3.11) as the star marker. This is located in a different part of the diagram with higher importance for HO_2 uptake (in the magenta zone). It also falls into the most sensitive position in the NO emissions / HO_2 uptake sensitivity plot (Figure 3.12)

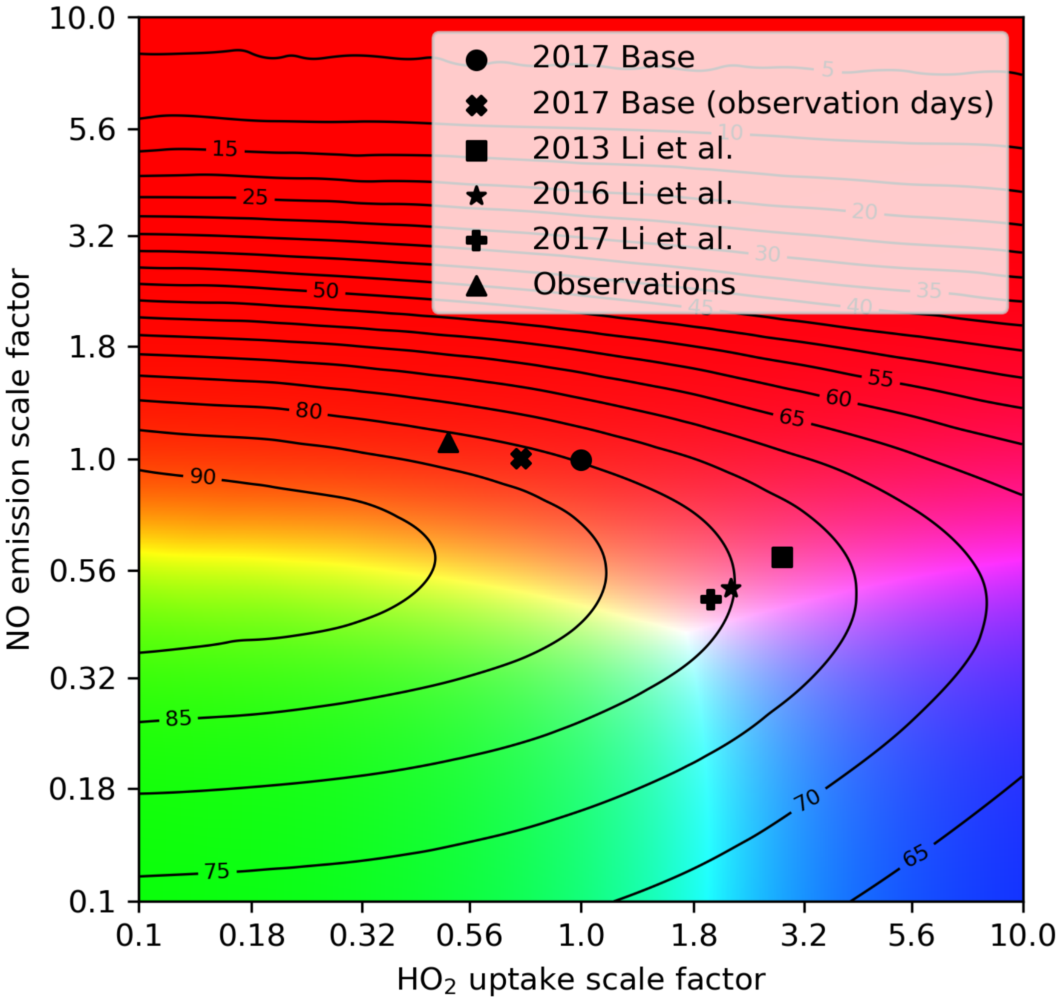

Figure 3.11: The colour coded dominant 12 pm campaign average (mean) radical loss pathway for a given scale factor on NO emission and HO_2 uptake, where red is NO_2+OH, blue is HO_2 uptake and green is RO_x+RO_x. Overlayed is the averaged modelled (mean) 4 pm O_3 concentration isopleth for the duration of the campaign. The round marker indicates the base run, while the square, star, plus indicate the calculated locations for the different years described in K. Li et al., 2019.

As the total change in NO_x emissions and surface area (assuming it surface area and PM2.5 scale proportionally) are known, the assumption was made that the change happened linearly across the five years and the 2013 and 2017 points (square and plus markers respectively) can be placed on the plot via deduced reckoning.

The lack of O_3 sensitivity to a 23% NO reduction compared to the ~5% increase in O_3 observed in Section 2.9 when moving from 1.25 to 1 scale on NO emissions) can be explained by their relative positions on the isopleth. The K. Li et al., 2019 study sat at the top of the NO ridge in the transitional regimes. It can be seen in Figure 3.12 that from the

2013 position (square marker) and moving down to a lower rate of NO emission the route of travel would be along the O_3 curve and result in minimal change. This is in agreement with Section 2.10 that found minimal O_3 sensitivity to NO emissions when sitting close to the maximum O_3 ridgeline.

The 2017 NO emissions should have been similar for both studies, and thus the difference in the photochemical regimes could be based on the resolution the models were run. K. Li et al., 2019 was run at a 0.5°x 0.625°resolution compared to 0.25°x 0.3125°resolution used in this study. As shown in Section 2.6 the actual Beijing urban region only spans a two by two box region at 0.25°thus at 0.5°higher biogenic VOC and lower NO emissions boxes from the surrounding region could be combined with the urban grid boxes. This would alter the NO:VOC ratio and could thus have a profound effect on the photochemical regime.

K. Li et al., 2019 found that a 41% drop in the aerosol surface area in five years resulted in a ~5.0% increase in O_3 in Beijing. Looking at the O_3 sensitivity plot (Fig 3.12) it can be seen that for this study originating from an uptake position 40% higher than the base run would have resulted in a ~3.2 ppb (4.00%) increase in O_3. The difference in sensitivity to changes in uptake appears to be a result of reduced NO_2 + OH influence.

A 41% reduction in uptake applied to the predicted position of K. Li et al., 2019 for 2013 (square marker) would see an O_3 increase of ~4.75%, close to the value calculated in the study. Nonetheless looking at the predicted position of the observations (triangle marker) coming from the position of 41% higher uptake would have resulted in ~2.4 ppb more O_3 (~3.0%). However, when coupled with the ~5% increase in O_3 due to reductions in NO emissions, the total increase is ~8%. An 8% reduction is close to the 7% meteorological normalised observed value (K. Li et al., 2019).

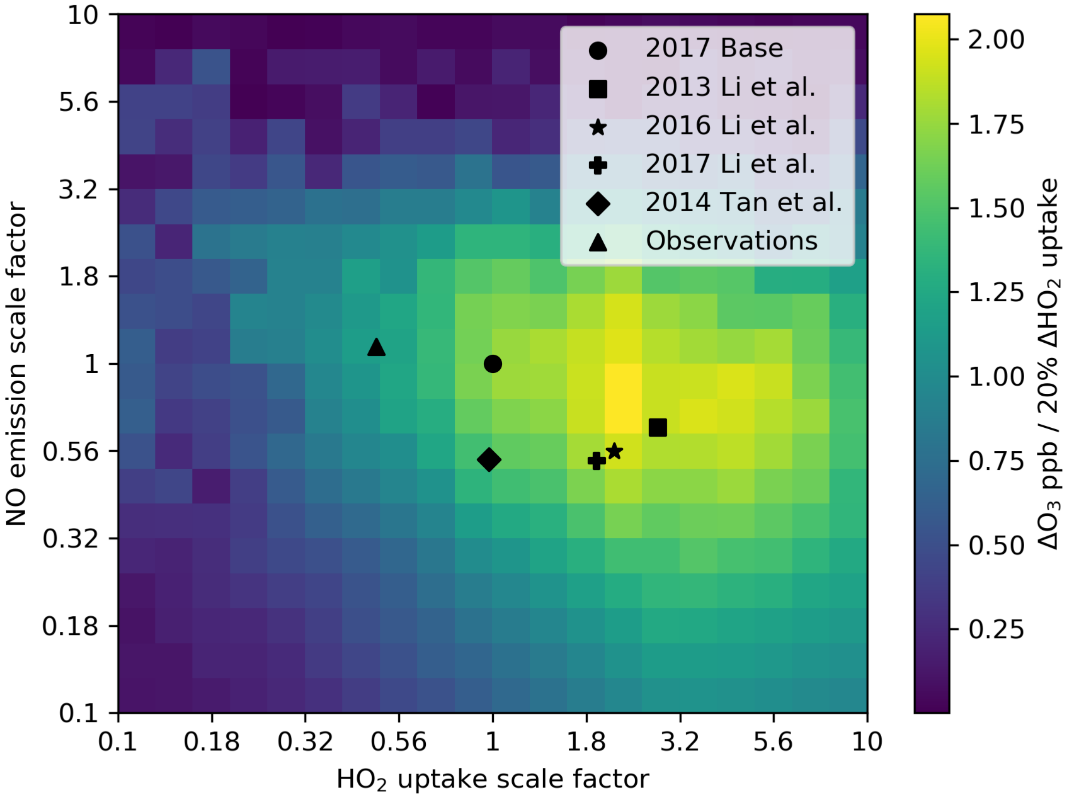

Figure 3.12: Campaign average modelled (mean) 4 pm O_3 sensitivity to a 20% change in HO_2 uptake averaged (mean) over 20% changes in NO emission. The markers identify the positions located in Figures 3.10 and 3.11.

3.6 Spatial variance in radical loss distribution

Figure 3.10 presented the radical loss pathways in term of the fractional loss from NO_2+OH rate (in red), RO_x+RO_x self reactions (green) and HO_2 uptake (blue) and expressed the different fractional combinations by adding these colour representations together.

This analysis was performed for a single site, but it can be extended geographically to explore the balance of sinks through the model domain. Figure 3.13 shows the mean noontime radical loss pathways across the model domain during the campaign period. The Beijing observational site (circle marker) is in a red, NO_2 + OH dominated grid box. Much of the industrialised region to its south, the North China Plain is in a similar regime, as is the area around Shanghai. To its north are regions in green where radical-radical reactions dominate the sink processes. The transition between these regions, indicated by the orange-yellow boxes shows where RO_x +RO_x and NO_2+OH are relatively balanced and so is close to the O_3 ridgeline discussed in Section 2.10.

Over the Bohai Bay to the east of Beijing and into the North of the Yellow Sea, magenta, pink and blue regimes can be seen. In these regions HO_2 uptake is important. As NO_x and aerosol is exported away from the Beijing source region, the NO_x concentration drops faster than the aerosol concentration as it has a shorter lifetime in the atmosphere. This results in the aerosol uptake of HO_2 becoming increasingly important away from the source region, and so the colours move from red to magenta to blue.

The colours on the map of Figure 3.13 are somewhat muted or pastel-shaded, indicating that the regimes may not be clear-cut with multiple sinks playing a role either due to day to day variability or over multiple days. There are several grid cells which appear white which indicates a three-way split between the different uptake routes.

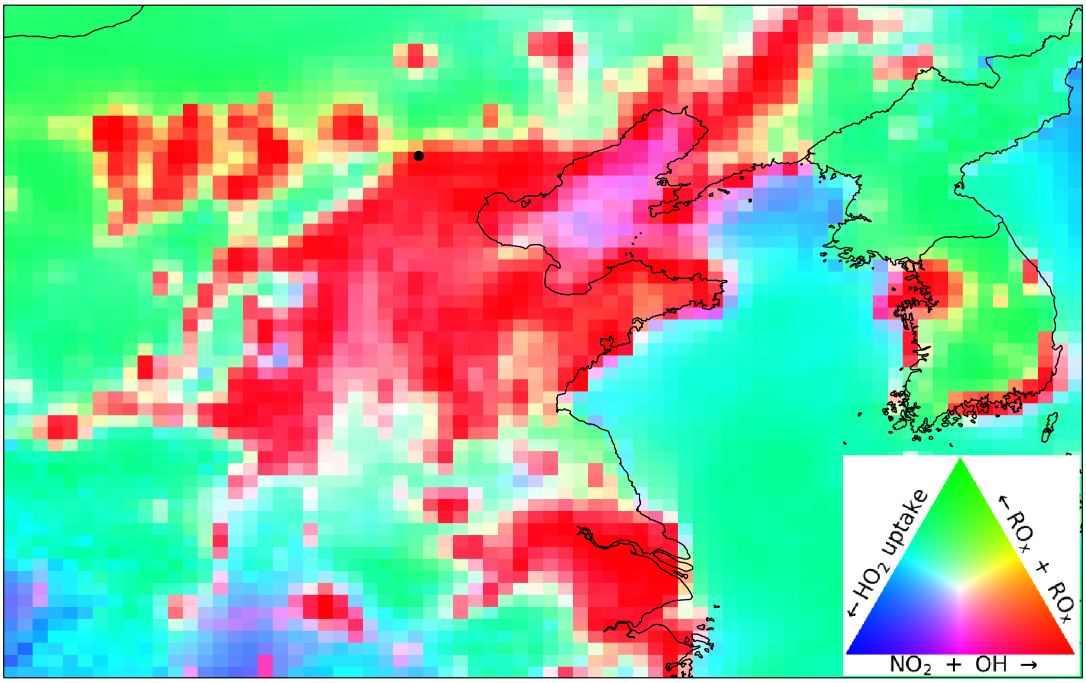

Figure 3.13: Map of East China colour coded with the dominant 12 pm campaign average (mean) radical loss pathway, where red is NO_2+OH, blue is HO_2 uptake and green is RO_x+RO_x. The black dot indicates the location of observation campaign.

From this analysis, it is apparent that there is significant spatial variability in the sink of radicals over East China. Within the domain, there are days and areas in which all three of the sink processes dominate. Remote regions are not necessarily dominated by radical-radical reactions, as the uptake of HO_2 onto aerosols may also play an important role. Two days during the campaign that did not show display a clear NO_2+OH dominance with will now be examined.

In Figure 3.14 the noontime dominant loss pathway for May 28^{th} can be seen with wind quivers overlayed. Here the Beijing site has shifted to a Yellow regime indicating both RO_x + RO_x and NO_2 + OH are dominant. This is in agreement with the findings from Section 2.9 that the regime in Beijing can change between days. The inland green regions on the 28^{th} are more vividly green than the mean plot indicating that they are more definitively in the ROx + RO_x dominated regime. Conversely, over the coastal regions, there is a shift to a HO_2 uptake (blue) regime.

Figure 3.15 shows the combined NH_4^+, SO_4^{2-} and NO_3^- aerosol surface area. A hot spot can be seen over the Yellow sea. The prevailing winds from both polluted industrial region south-east of Beijing, as well as Shanghai, carry polluted wind parcels into the Yellow Sea on the 28th. High humidity associated with coastal regions may also increase the surface area of the aerosol.

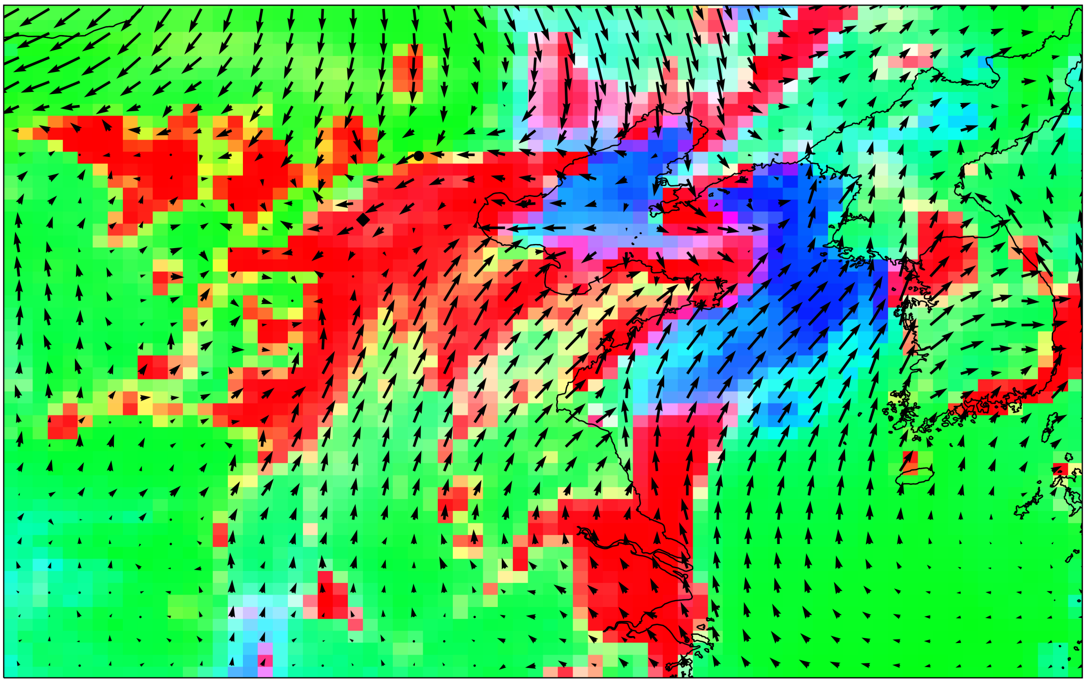

Figure 3.14: Map of East China colour coded with the dominant May 28th 12 pm radical loss pathway, where red is NO_2+OH, blue is HO_2 uptake and green is RO_x+RO_x. The black dot indicates the location of observation campaign. Wind quivers indicate the wind velocity for a given grid box.

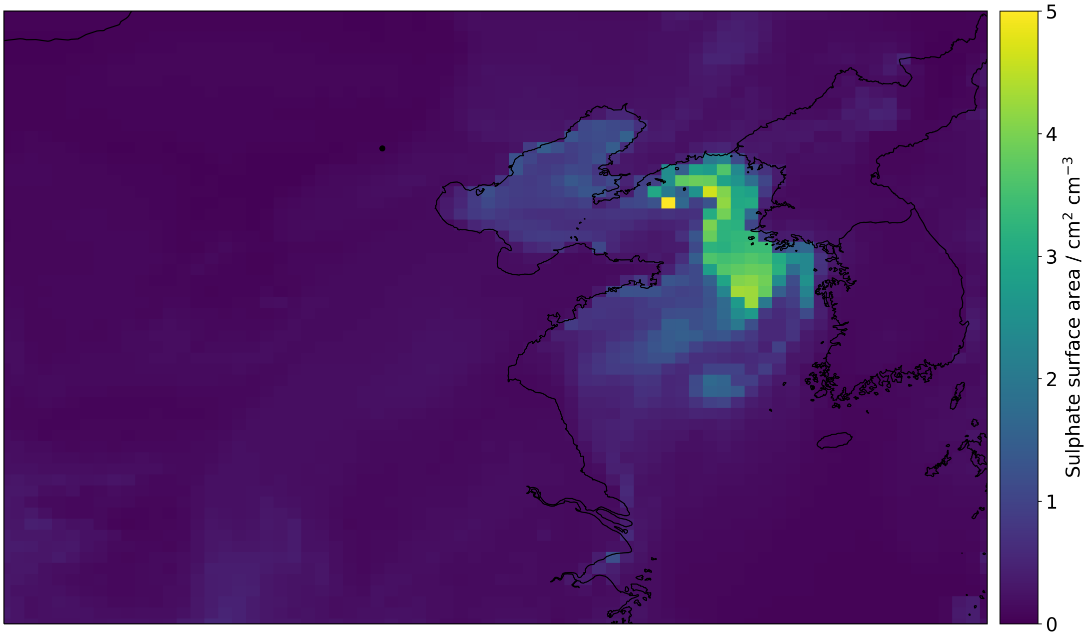

Figure 3.15: Map of modelled NH_4^+, SO_4^{2-} and NO_3^- aerosol surface area at noon on May 28^{th}.

In Figure 3.16 the noontime dominant loss pathway for June 22^{nd} can be seen. Here the inland shift was towards HO_2 uptake regime indicating an increase in particulate. The wind quivers on this day indicate that prevailing winds carried polluted urban wind parcels inland to the north west. A pink transitional regime replaces the boundary regions of the NO_2 + OH dominated areas that were yellow in the mean plot. Figure 3.17 shows the accumulation of NH_4^+, SO_4^{2-} and NO_3^- aerosol north and west of Beijing. Conversely, the coastal regions shift to a green regime as "clean" air from the Pacific Ocean is carried into the region.

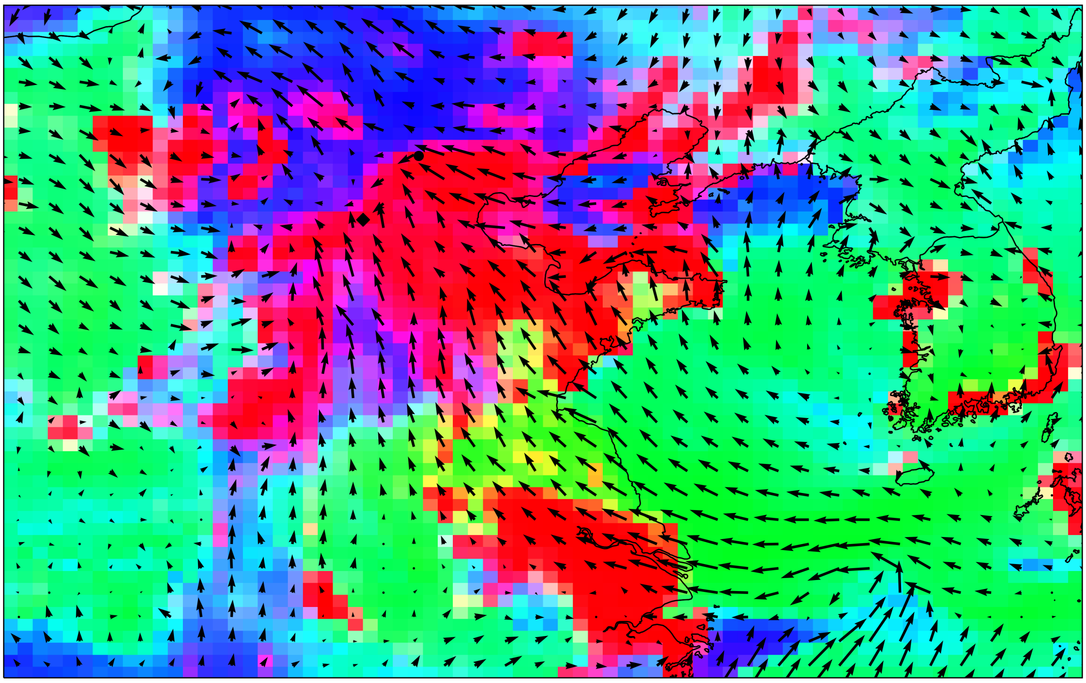

Figure 3.16: Map of East China colour coded with the dominant June 22^{nd} 12 pm radical loss pathway, where red is NO_2+OH, blue is HO_2 uptake and green is RO_x+RO_x. The black dot indicates the location of observation campaign. Wind quivers indicate the wind velocity for a given grid box.

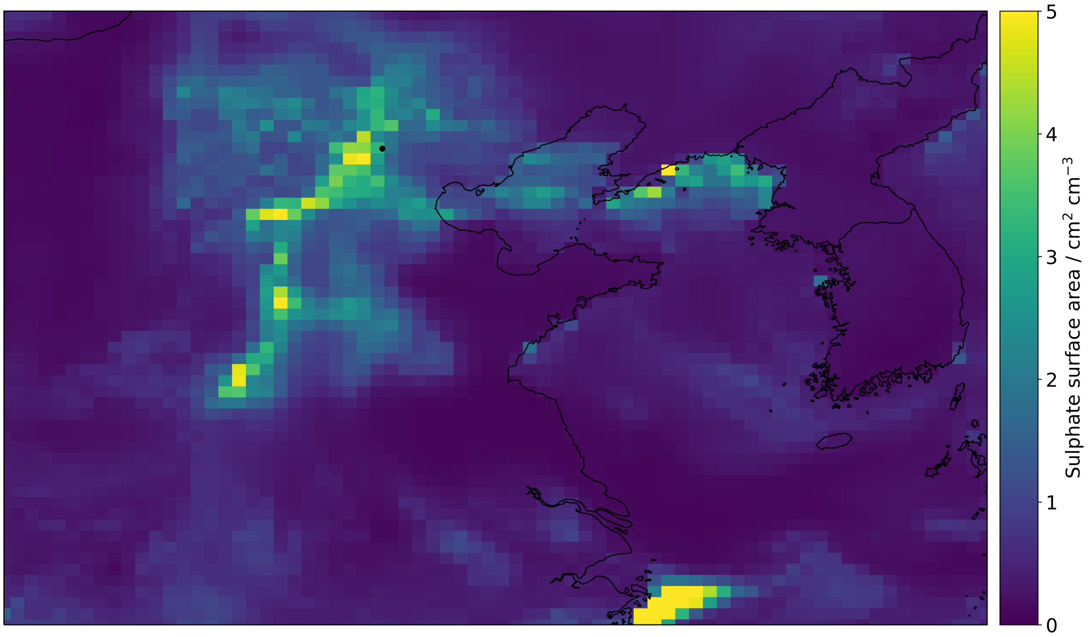

Figure 3.17: Map of modelled NH_4^+, SO_4^{2-} and NO_3^- aerosol surface area at noon on June 22^{nd}.

3.7 Conclusions

HO_2 uptake onto the surface of aerosols has been known since the 1980s but its importance for determining regional photochemical O_3 production has not been recognised until more recently.

This work has attempted to understand the role of HO_2 uptake from the perspective of a chemical transport model tuned to the conditions of a specific field campaign. Under those conditions and with the model chemistry used, HO_2 uptake is as important as isoprene emissions for determining the mean O_3 concentration.

Exploring the model sensitivity to HO_2 uptake, through a series of over 130 model simulations it is possible to define 3 new chemical regimes: Aerosol Insensitive (High NOx), Aerosol Sensitive, Aerosol Insensitive (Low NOx) which give different conditions for a change in aerosol surface area to impact O_3 production.

These regimes are associated with the changing balance of the different radical sinks in the model (HO_2 uptake, NO_2+OH, RO_x+RO_x). Exploring those regimes specially and temporally shows significant variation across the North China region of model with grid cells being in pretty much all combinations of sinks at some point during the campaign.

There are a number of policy conclusions from these simulations:

- Across a region the size of the model domain there is significant spatial variability in radical sinks. This is likely to lead to spatial differences in the response of O_3 to policy decision.
- Continuing to decreasing NO emissions is likely to lead to increases in O_3 concentration in much of the domain, where NO_2+OH dominates radical loss (red areas).
- Beijing itself is on the edge of of the NO_2+OH dominated region and depending on the wind direction can fall into a more RO_x+RO_x, or HO_2 uptake influenced domain.
- Given the model's difficulties in simulating the OH, RO_2, HO_2 and aerosol surface area concentrations there are significant uncertainties about the model's ability to provide useful conclusions about likely future evolution of O_3 in the city.

Chapter 4

Using machine learning for ozone bias correction

4.1 Introduction

The following chapter is based on Ivatt and M. J. Evans, 2020. My contributions included the development of the technique, analysis, all model runs and writing of the paper.

In the previous chapters (VOC vs NOx control, uptake to HO_2 onto aerosols), model failings resulting in O_3 bias have been shown. The failure to reproduce the NO/NO_2 ratio, the RO_2 concentration, the aerosol surface area and the photolysis rates are all examples of areas that the model prediction differs from reality, and so impacts the quality of the O_3 prediction. Process-based models such as GEOS-Chem use a quantitative understanding of physical, chemical and biological processes to make predictions about the environmental state. Given uncertainties in their initial conditions, input variables and parameterisations, these models show various biases. Much of the research into atmospheric chemistry revolves around reducing those biases due to uncertainties in the emissions of compounds into the atmosphere (Rypdal and Winiwarter, 2001), the chemistry of these compounds (Newsome and M. Evans, 2017) and meteorology (Schuh et al., 2019).

Understanding and reducing these biases is a critical scientific activity. However, the ability to improve these predictions without having to improve the model at a process level also has value. For example, air quality forecasting, and the quantification of the impacts air pollutants on human and ecosystem health, would both benefit from improved simulations, even without process-level improvements. Data assimilation techniques (P. Bauer, Thorpe, and Brunet, 2015) are used to incorporate observations into meteorological forecasts and some air quality forecasts (Bocquet et al., 2015). However, other tools and techniques to improve model predictions would be useful.

One of these tools would be a "bias predictor" (where the bias is used here to mean the ratio between the model and the measured quantity) (Harlass, Latif, and W. Park, 2015). This would identify the modelled bias from observations and then define it as a function of some other parameters, whether modelled or observed, allowing the model's bias to be predicted and thus removed. This particular has value when model performance is prioritized over the explicit understanding of how the model works, such as operational air quality models. While the prediction itself may have less value to the scientific community, how the algorithm calculated the value might have worth. By investigating how the algorithm arrived at a prediction could shed light on so-called "known unknowns" where analysis could help identify patterns and correlations in the model failure, which could aid future process-based development.

In this chapter, the development of a bias predictor for O_3 within the GEOS-Chem model, based on a machine learning approach is explored. Once defined, the bias predictor can then be applied more widely (in space or time) to model output to remove the bias, bringing the model closer to the observations. The theory behind machine learning (Sect. 4.2) and the GEOS-Chem configuration used as the base model (Sect. 4.3) are discussed first. The observations of O_3 from four observational networks (Sect. 4.4) and the method (Sect. 4.5) to produce bias predicting algorithm are then discussed. The performance of the bias-corrected model is compared to testing data in Sect. 4.7. A more general evaluation of the performance of bias correction compared to known global biases within GEOS-Chem is described in Sect. 4.8). The predictions resilience to a reduction in the number of training locations and duration of training data is discussed in Sections 4.8.1 and 4.9 respectively. Finally, an alternative method of prediction is considered (Sect. 4.10), and the future potential of such a methodology for air quality forecasting is discussed (Sect. 4.11).

4.2 Machine learning

Machine learning (specifically supervised machine learning) is a general term for any technique in which a computer is used to optimise an algorithm to make predictions based on a training dataset (Ghahramani, 2015). With increases in computational power, several techniques have risen to prominence. Two of particular interest for "large" (>10 million samples) datasets applications are neural nets and decision trees (Maxwell, Warner, and Fang, 2018).

What makes these techniques particularly powerful over traditional regression techniques (such as linear regression) is their ability to process large multivariate datasets (Ballabio, Grisoni, and Todeschini, 2018) while also capturing non-linear relationships (Gardner and Dorling, 2000). Non-linear relationships are particularly important for understanding atmospheric components (such as O_3), where changes in emissions often elicit a non-linear response (Cohan et al., 2005).

A "neural net" is a technique whereby "hidden" layers of nodes connect input variables (features) to an output prediction, mimicking neurons in the brain (Acharya et al., 2018). Understanding and developing these hidden layers is still very much a continuing science. During early experimentation, it was found that substantial and unpredictable changes resulted from alterations to the architecture of the hidden layers (Gardner and Dorling, 1998). Further to this, the hidden layers can be almost impossible to interpret, notably when they are emulating a complex system. Thus the use of neural nets was found to be problematic especially for science (Cynthia, 2019) and alternative methods were explored.

A "decision tree" attempts to make a prediction based on a set of training data, by breaking down the data into smaller subsets based on simple decisions based on the input variables: if the value of an input variable is greater than a specified value. These splits are known as nodes. The ability to readily visualise these nodes provides decision trees with their interpretability (Kingsford and Salzberg, 2008). The data is then further broken down with more and more decision nodes until a terminal node (leaf) is reached. Each leaf contains a prediction for the variable, which is generally the mean of the samples that took that route down the tree. An simple example of a decision tree can be seen in Figure

4.1.

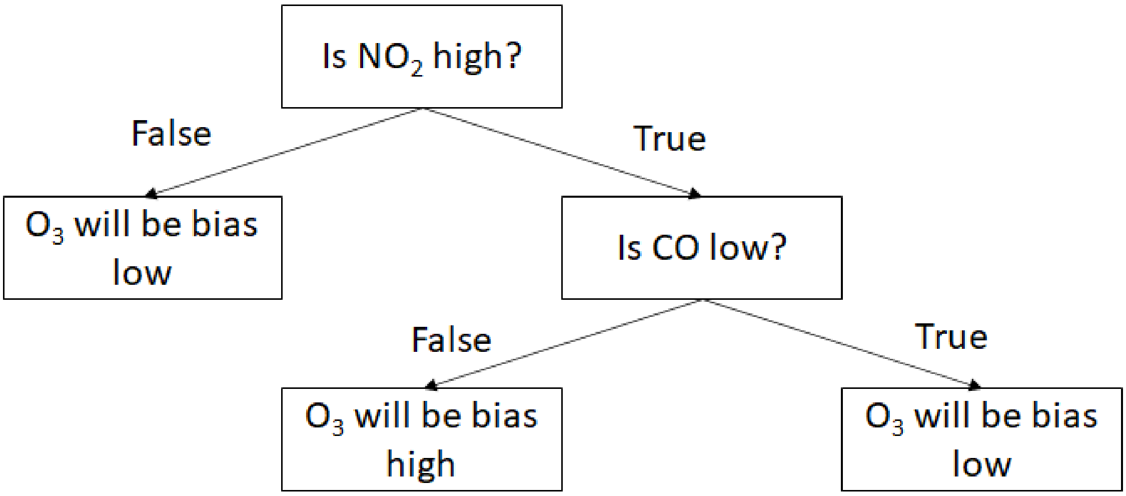

Figure 4.1: An example decision tree

The choice and value of hyperparameters is an important aspect when building decision trees. These hyperparameters define the tree structure. A tree that is too shallow (not enough splits) will miss fundamental relationships in the training data and result in a "weak" or poor predictor. Conversely making the tree too deep will result in "overfitting" and not achieve good generalisation of the training data (Jiang et al., 2019).

Identifying optimum hyperparameters is achieved via cross-validation, whereby the training data is split into subsets (folds). A tree is then trained on all but one fold and tested on the last remaining fold. This process is repeated for each fold, and the performance metric (e.g. mean squared error) of each tree is averaged. Cross-validation is a computationally burdensome task, with the process taking n-folds times longer to train than the final tree, but it helps to define the appropriate values of the hyperparameter. As such, this process cannot be carried out for every combination of hyperparameters, particularly on large datasets. Carefully chosen hyperparameters are vital with consideration of how they will interact with a particular training dataset (Bengio, 2000). Incorrect use of hyperparameters can bring about overfitting and under generalisation of the training data.

Decision trees can be used for "classification" (predicting a variable which falls into two or more categories) or "regression" problems (predicting a variable which can have any value). For developing a bias predictor, it is this regression configuration that will be used in this study.

The main drawback with decision trees is their tendency to overfit (Owaida et al., 2017). It does not take too many splits to reach a point where only a minimal number of training samples contributed to each leaf. The prediction ends up too closely reflecting the quirks of the training dataset, rather than the underlying properties of the training dataset. An improved approach is to use an ensemble of trees known as a decision "forest."

In a decision "forest", a collection of multiple different deep trees can be trained, with each tree being trained on unique subsets of the training data. The predictions of all the trees are then averaged to produce a final prediction. This technique is referred to as a "random forest" (Breiman, 2001). While this is a capable machine learning method, it has now been overtaken by what is known as a "gradient boosted" decision tree ensemble approach.

The specific machine learning technique used in this study was the XGBoost python package implementation of gradient boosted regression (T. Chen and Guestrin, 2016). This ensemble method works by using shallow trees to minimise over-fitting and then relying on a large number of trees to gain the complexity required. The real power of this methodology lies in how the trees are trained. Each tree is trained sequentially on the residuals of the previous tree, meaning that later trees can focus on more extreme and unusual samples (Frery et al., 2017). The algorithm has a relatively quick training time allowing for efficient cross-validation. Lastly, the technique is highly scalable, meaning tests can be run on small subsets of data before increasing to much larger training datasets (Torlay et al., 2017).

The XGBoost algorithm will be used to predict the O_3 bias in the GEOS-Chem chemical transport model, the configuration of which will now be described.

4.3 GEOS-Chem

GEOS-Chem Version V11-01 (Bey, Jacob, et al., 2001) was used for the analysis. In this proof of concept work, the model was run at a coarse resolution of 4^o x 5^o for computational expediency, using MERRA2 meteorology from the NASA Global Modelling and Assimilation Office (https://gmao.gsfc.nasa.gov/reanalysis/MERRA-2/). The model consisted of 47 vertical levels extending from the surface to approximately 80 km in altitude. The model had the "tropchem" chemistry configuration which consists of 68 chemical species and 448 reactions with a linearized version in the stratosphere (Eastham, Weisenstein, and Barrett, 2014). The emission inventories used were EDGAR (Crippa et al., 2018) and RETRO (L. Hu, D. B. Millet, et al., 2015) for global anthropogenic emissions, which were overwritten by regional inventories where available: NEI (USA) (Travis et al., 2016), CAC (Canada) (Donkelaar et al., 2008), BRAVO (Mexico) (Kuhns, Knipping, and J. M. Vukovich, 2005), EMEP (Europe) (Donkelaar et al., 2008) and MIX (East Asia) (M. Li, Q. Zhang, et al., 2017)). GFED4 (Giglio, Randerson, and Werf, 2013) and MEGAN (A. B. Guenther et al., 2012) were used for biomass burning and biogenic emissions respectively. Details of the other emissions used and other details of the model can be found online
(http://wiki.seas.harvard.edu/geos-chem/index.php/HEMCO_data_directories).

The model was run from January 1st 2010 to December 31st 2017 outputting the local model state (the output variables are discussed in Section 4.5) for each grid box for which observational O_3 data was available (the observation dataset is discussed in Section 4.4). Additionally, between January 1st 2016 and December 31st 2016, the model output hourly local state data for all grid boxes (used for global analysis Section 4.8).

4.4 Observations

In order to perform well, an algorithm must be built upon a clean, representative and well-distributed dataset. Cleaning in data science terms means removing missing and corrupt data as well as performing processes to homogenise the data such as time averaging.

For this study, a database of O_3 observations from 1^{st} of January 2010 to the 31^{st} of December 2017 was created. This dataset consisted of ground observations, measurement from balloons (sondes) and measurements from the NASA "ATom" campaign. The database consisted of 65,805,112 data points, the coverage at 4^{o} x 5^{o} resolution is shown in Figure 4.2. Each set of observations will now be explained.

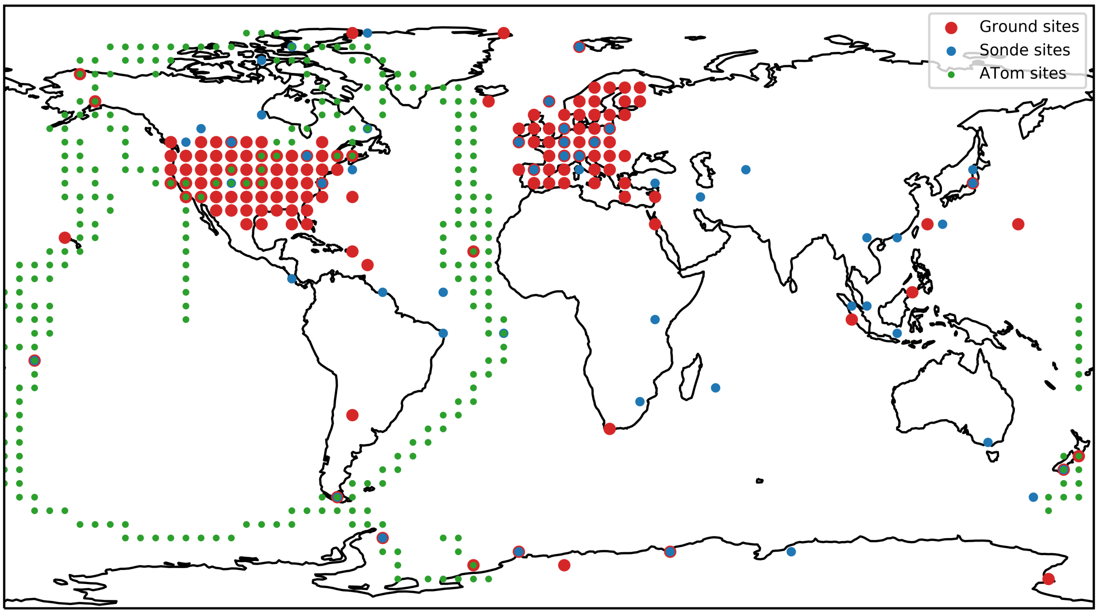

Figure 4.2: Locations of "meta" observations (averaged over model 4°x5° grid boxes) from the surface (EPA, EMEP and GAW in red), the ozone-sonde network (blue) and the ATom flights (Green).

4.4.1 Surface Data

Surface O_3 observations were compiled from an number of air quality and atmospheric composition monitoring networks: the European Monitoring and Evaluation Program (EMEP) (www.emep.int), the United States Environmental Protection Agency (EPA) (https://www.epa.gov/outdoor-air-quality-data) and the Global Atmospheric Watch

(GAW) (https://public.wmo.int/) observation networks. Due to the coarse spatial resolution of the model (4^o x 5^o), sites flagged as "urban" were removed. Urban measurements are often made in regions of steep concentration gradients which can only be captured in higher resolution models. Producing a useful bias correction for a coarse model for these conditions would be difficult as there would be little information to base the bias correction on. Similarly, all mountain sites were removed due to the difficulty in representing the complex topography typical of mountain locations within the large grid boxes. In this study, mountain sites were defined as ground sites at pressure lower than 850 hPa.

Where multiple observations occurred in the same grid box at the same model time step. A mean was taken of all simultaneous observations. This thus led to the generation of a set of "meta-sites" which represent the hourly mean observations on the model grid. Fig. 4.2 shows the locations of the meta sites used in this study.

4.4.2 Sonde Data

Ozone-sonde observations were from the World Ozone and Ultraviolet Radiation Data Centre (https://woudc.org). The stratosphere falls outside the scope of this study, and thus all stratospheric observations were removed. O_3 concentrations greater than 100 ppb was used as the definition of stratosphere (L. L. Pan et al., 2004). As with the ground sites, meta-sites were generated on the three dimentional model grid. All sonde observation below 850 hPa were removed to prevent ground and sondes sites being averaged into the same meta-site.

Sonde observations provide regular long term sampling of the free troposphere. However, there are significant drawbacks relative to ground observations. Firstly, they lack temporal resolution. Most sondes are released at noon once every couple of weeks. Thus they represent at best a "snap shot" of the state of the atmosphere. Secondly, they are distributed around the globe, there is a scarcity of locations. Most continents are only sampled by a handful of sites.

4.4.3 ATom Data

The NASA ATom campaign flew over the Pacific and Atlantic oceans, in a loop from northern poles, through to the southern poles, via the mid-latitudes and tropics (S. Wofsy et al., 2018). The NASA DC8 aircraft sampled from the surface to 15 km measuring the concentration of many compounds including O_3.

Flights took place in four different seasons between July 2016 and May 2018, but only the first three (summer, spring and winter) are used due to availability at the time of writing. As with the sonde data, stratospheric observations were removed (>100 ppb), and observations were averaged into meta-sites based on the model grid. A map of all grid boxes that contain observations is shown in Figure 4.2.

4.5 Training

The training dataset for the algorithm consisted of all ground and sonde observations between January 1st 2010 and December 31st 2015 (now referred to as "training period"). Importantly no ATom data was included in the training dataset, keeping the ATom observations as an independent testing dataset.

The corresponding modelled O_3 predictions for the training period were used to calculate the model bias for each meta observation point. The bias was defined in terms of a scale factor; model prediction divided by observation. The XGBoost algorithm attempts to produce the bias predictor function based on minimising a loss function, in this case mean squared error (MSE). The nature of the MSE would suggest that a bias (model/measured) of 10 was "worse" than one of 0.1 However, the model "failure" is the same in both cases.

To counter that, the algorithm is trained on the $\log_2$ of the bias. Thus a bias of 0.1 ($\log_2(0.1)$=-3.32) was weighted as highly as a a bias of 10 ($\log_2(10)$=+3.32) due to both -3.32 and 3.32 having the same absolute value compared to 0.1 and 10. In total, there were 13,118,334 surface and 250,533 sonde biases in the training dataset.

The local model state at the time and location of each bias was used to produce training features. The model local state consisted of the grid box concentration of the 68 chemicals transported by the model (including O_3) and 15 physical model parameters (see Table 4.1 and 4.2 respectively). While GEOS-Chem required modification to output some of these variables, the decision was made to include as many readily available outputs as possible and to let the machine learning algorithm decide what was important.

Hyperparameter tuning was achieved by 5 fold cross-validation of the training dataset. Root Mean Squared Error (RMSE) was used as the performance metric. The training data was organised by date to ensure each fold contained roughly a year slice of data. The principal parameters tuned were the number and maximum depth of trees.

During cross-validation, similar performance could be achieved with 12 to 18 layers of tree depth, with a reduction in the number of trees needed at higher depth. It was found that the algorithm achieved the majority of its predictive power early in training, with the last 90% of trees producing small gains in RMSE.

As the number of trees was found to have a linear relationship with training time. A compromise was made between training time and predictive strength. 150 trees with a maximum depth of 12, was chosen as the configuration for testing. This selection of hyperparameters with the full set of training data took 1 hour to train on a 40 core CPU node, consisting of two Intel Xeon Gold 6138 CPUs.

The algorithm is now a trained "bias-predictor"; the input variables can now be examined for insight into the predictor's structure.

Table 4.1: Chemical tracers used during training.

Chemical tracers	
NO	Hydrophilic black carbon
O_3	Hydrophobic organic carbon
Peroxyacetylnitrate	Hydrophilic organic carbon
CO	0.7 micron dust
$\geq$C4 alkanes	1.4 micron dust
Isoprene	2.4 micron dust
HNO_3	4.5 micron dust
H_2O_2	Isoprene epoxide
Acetone	Accumulation mode sea salt aerosol
Methyl ethyl ketone	Coarse mode sea salt aerosol
Acetaldehyde	Br_2
$\geq$C4 aldehydes	Br
Methylvinylketone	BrO
Methacrolein	HOBr
Peroxymethacryloyl nitrate	HBr
Peroxypropionylnitrate	$BrNO_2$
$\geq$C4 alkylnitrates	$BrNO_3$
Propene	$CHBr_3$
Propane	CH_2Br_2
Formaldehyde	CH_3Br
Ethane	Methyl peroxy nitrate
N_2O_5	Beta isoprene nitrate
HNO_4	Delta isoprene nitrate
Methylhydroperoxide	5C acid from isoprene
Dimethylsulfide	Propanone nitrate
SO_2	Hydroxyacetone
SO_4^{2-}	Glycoaldehyde
SO_4^{2-} on sea salt	HNO_2
Methanesulfonic acid	Nitrate from methyl ethyl ketone
NH_3	Nitrate from methacrolein
NH_4^+	Peroxide from isoprene
Inorganic nitrates	Peroxyacetic acid
Inorganic nitrates on sea salt	NO_2
Hydrophobic black carbon	NO_3

Table 4.2: Physical parameters used during training.

Physical parameters
Pressure
Temperature
Absolute humidity
Surface pressure
Aerosol surface area
Horizontal wind speed
Vertical wind speed
Isoprene epoxide
Cloud fraction
Optical depth
Solar zenith angle
$Cos(day\,of\,year/360 * 2\pi)$
$Sin(day\,of\,year/360 * 2\pi)$

4.6 Feature importance

There are two metrics for investigating the importance of the input variables in the trained bias-predictor: "gain" is the reduction loss function (mean squared error) gained from splits using that feature, "weight" is the number of times a feature is used to decide a split. The top 10 features by gain and weight for the training dataset is shown in Fig. 4.3 and Fig. 4.4 respectively.

Fig. 4.3 shows those species that had the most profound effect (gain) on the predictive ability of the bias-predictor. The most important feature from the analysis was the concentration of NO_3 (the nitrate radical). This has a high concentration in polluted night-time environments and a low concentration in clean regions or during the daytime (Winer, R. Atkinson, and Pitts, 1984). This feature appeared to be being used to correct the concentration of O_3 in regions such as the US which are polluted and have a significant high bias at night a commonly observed feature (Yerramilli et al., 2012). The next most important feature was the O_3 concentration itself. This may reflect biases in regions with very low O_3 concentrations such as around Antarctica. The third most important feature was the CH_2O concentration. This may indicate biases over regions of high photochemical activity, as CH_2O is a product of the photochemical oxidation of hydrocarbons (Wittrock et al., 2006).

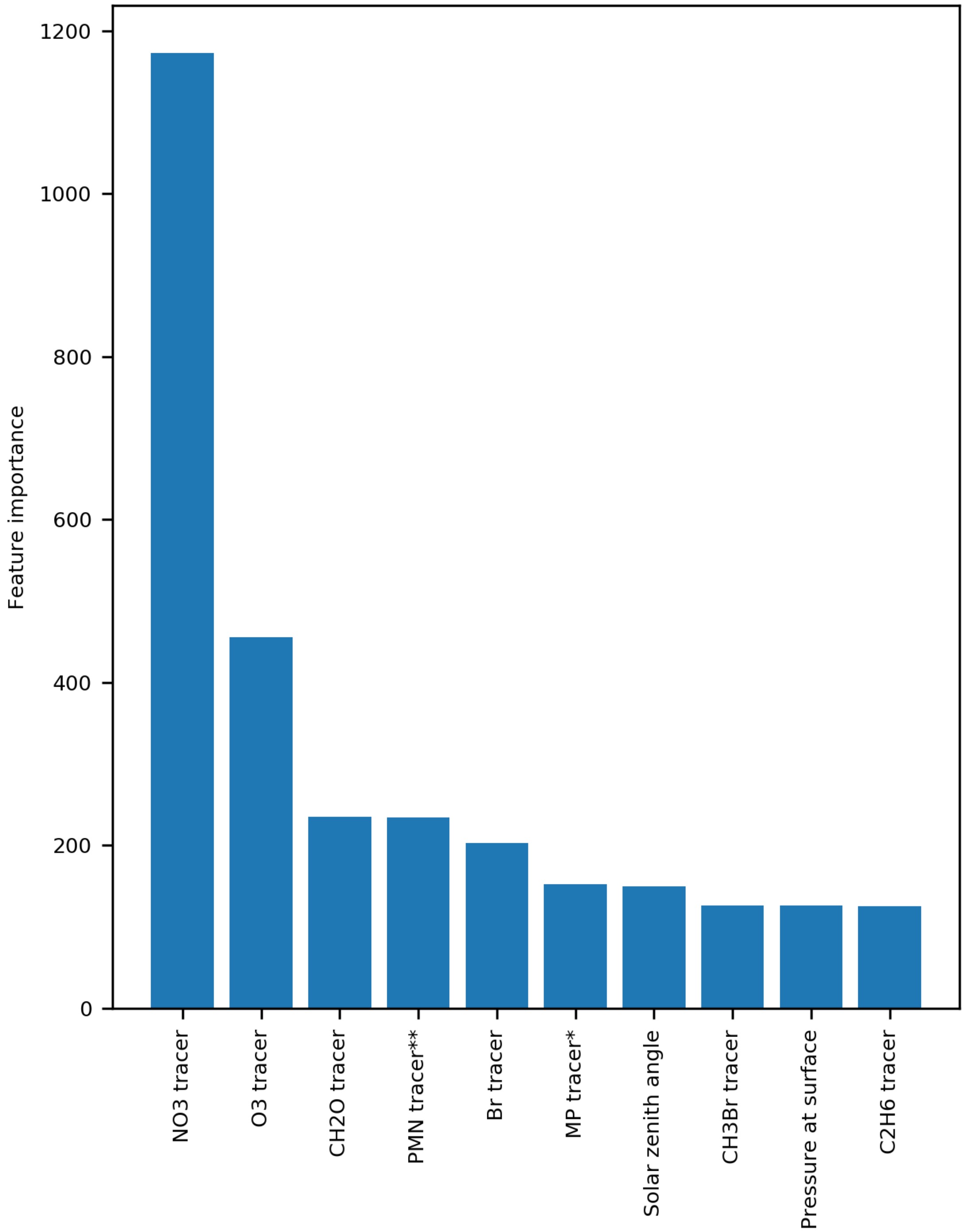

Figure 4.3: Feature importance based on gain(the average gain across all splits the feature is used in). *Methylhydroperoxide, **Peroxymethacryloyl nitrate

Surprisingly except for the O_3 tracer itself, the top 10 features by weight (Fig. 4.4) contained an entirely different set of species to that of gain. Variables that had high weight but low gain would frequently appear further down the decision tree. These are variables used for fine-tuning the correction, splitting small subsets of the data. Pressure had the highest weight, allowing the algorithm to form relationships with altitude. NO was the second most important feature which may be used again as a diurnal marker. NO is also a component of the NO_x and thus has a direct chemical relationship with O_3. The third most prominent feature was the pressure at the surface; this provides topological information. Future work should explore the explanatory capabilities of the algorithm, allowing for the bias-predictor to provide qualitatively as well as quantitative understanding of the model bias.

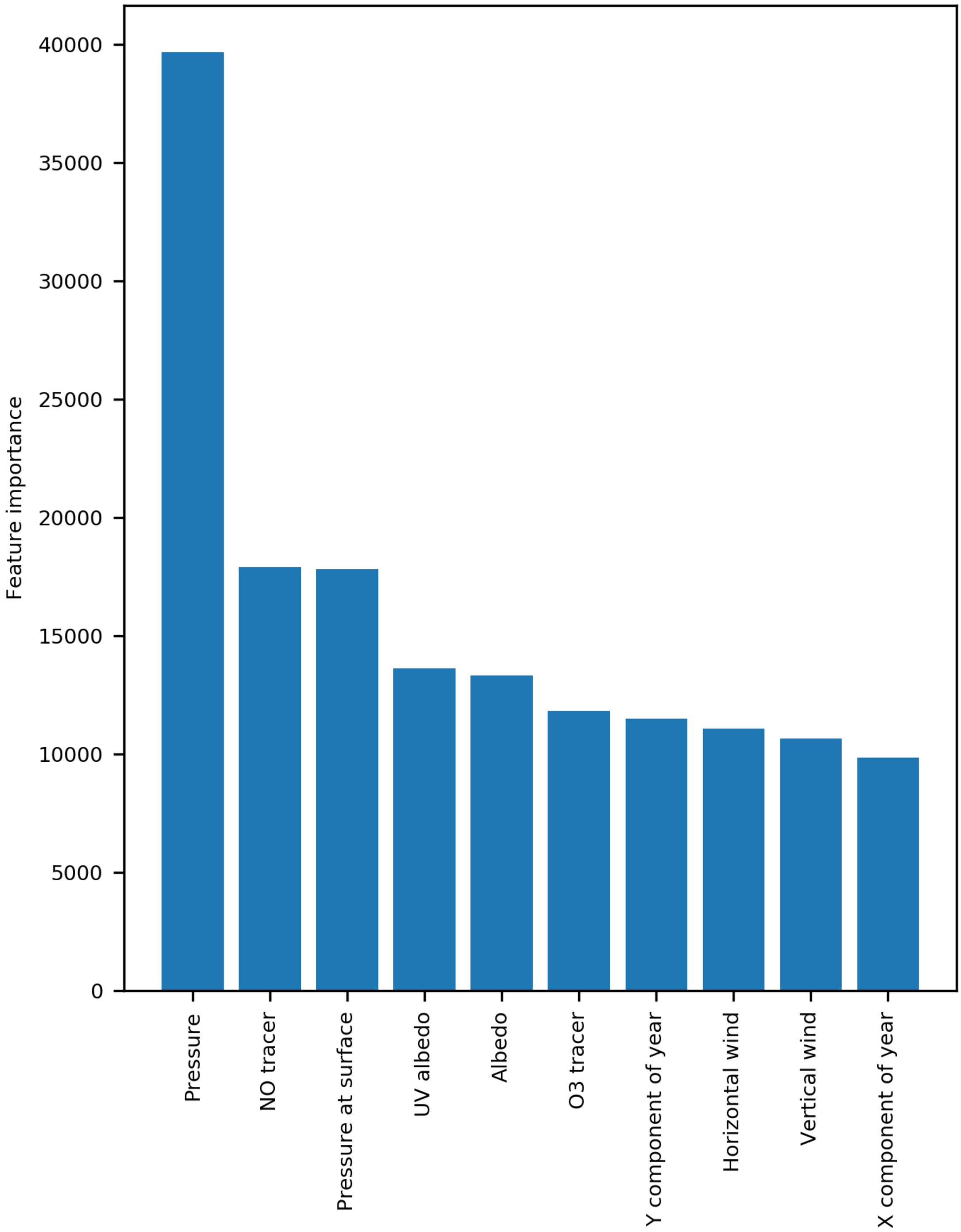

Figure 4.4: Feature importance based on weights (the number of times a feature is used to split he data across all trees).

With the bias-predictor trained and an idea of its structure known, its ability to predict the model bias for a completely different set of years is now tested.

4.7 Testing

The bias-predictor was tested on the period between January 1st 2016 and December 31st 2017 (now referred to the "testing period"). None of the data used to train the predictor was thus used in its testing. Observed bias was calculated in the same way as the training period and included 3,783,303 surface and 78,451 sonde observations. Additionally, the ATom flight data (10,518 observations) was included as an independent testing dataset.

Testing was achieved by evaluating the performance of the model against observation and then comparing the change in performance with the predicted bias removed. Three performance metrics were chosen and are described in Sect. 4.7.1.

4.7.1 Performance metrics

The Root Mean Squared Error (RMSE) measures the average error in the prediction (Equation 1.1). Normalised Mean Bias (NMB) measures the direction of the bias and normalises to the result (Equation 1.2). The Pearson's R correlation coefficient measures the linear relationship between the prediction and the observation (Equation 1.3).

$$RMSE(y, \hat{y}) = [\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2]^{\frac{1}{2}} \tag{Eq. 4.1}$$

$$NMB(y, \hat{y}) = \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)}{\sum_{i=0}^{N} y_i} \tag{Eq. 4.2}$$

$$R(y, \hat{y}) = \frac{\sum_{i=1}^{N}[(y_i - \bar{y}_i)(\hat{y}_i - \hat{\bar{y}}_i)]}{\sum_{i=1}^{N}[(y_i - \bar{y}_i)^2(\hat{y}_i - \hat{\bar{y}}_i)^2]^{\frac{1}{2}}} \tag{Eq. 4.3}$$

Where y is the observed values, $\hat{y}$ is the predicted values and N is the number of samples.

4.7.2 Surface data

A point by point comparison between the testing period surface data and the model with and without the bias correction is shown in Figure 4.5. The bias correction removes virtually all of the model biases (NMB) taking it from 0.29 to -0.04, substantially reduces the error (RMSE) from 16.2 ppb to 7.5 ppb and increases the correlation (Pearson's R) between the model and the measurements from 0.48 to 0.84.

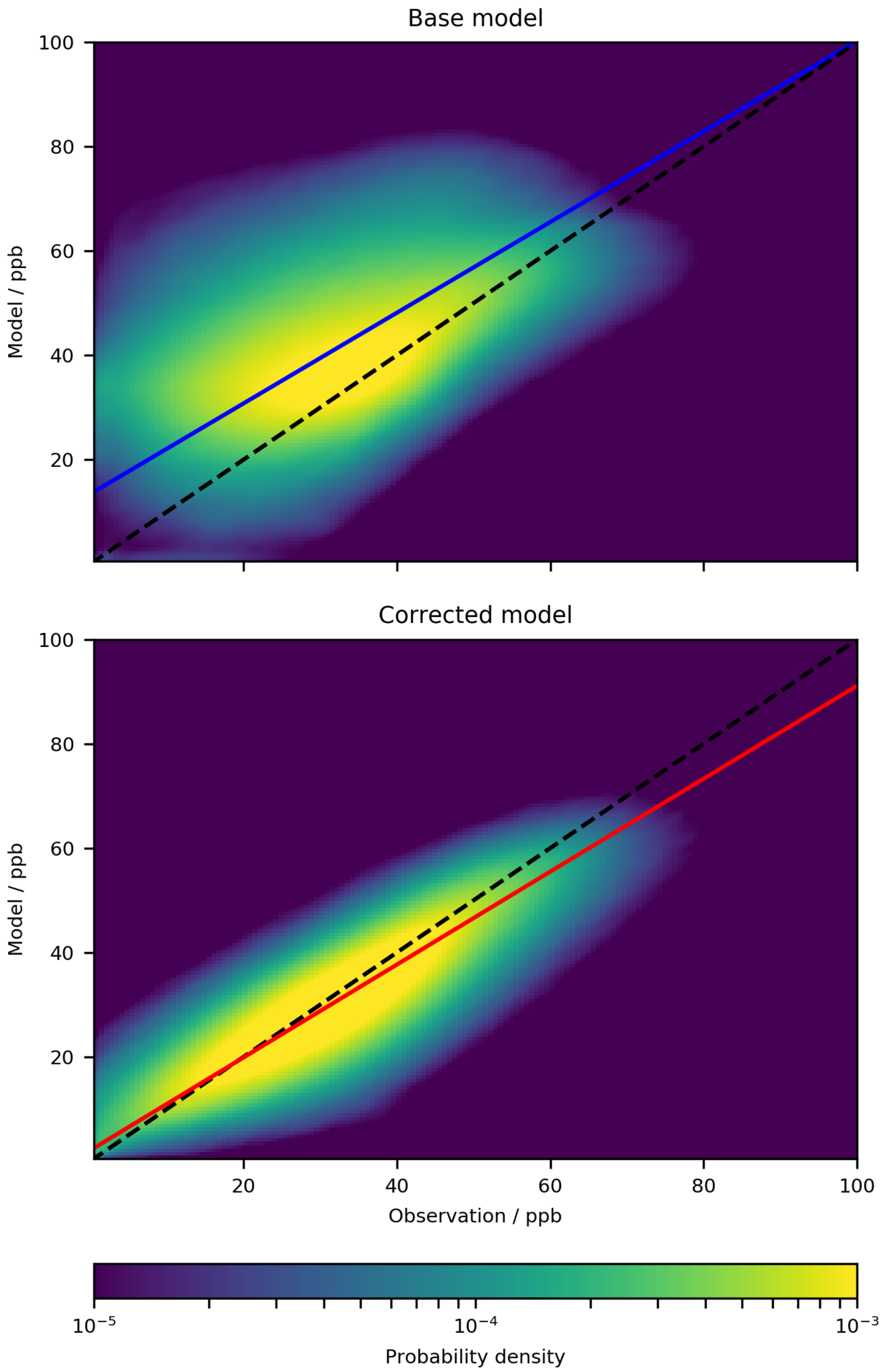

Figure 4.5: Kernel density estimation plot of model verses observation for all ground sites in the base model (upper panel) and corrected model (lower panel) for the testing period (1/1/2016 to 31/12/2017). The dashed black line is the 1:1 line, the coloured line show the orthogonal line of best fit. The plot is made up of 3,783,303 data points.

The bias corrector appears to be doing a good job in removing the bias from the model during the period 2016 to 2017. A more focused analysis now evaluates what the algorithm has done to make this improvement.

Diurnal variation

To probe the performance of the algorithm in more detail, the median diurnal cycles of nine, globally distributed ground sites were examined. Results are shown in Fig. 4.6 with statistics given in Table 4.3. The base model (blue line) shows substantial differences with the observations (black line) for most sites. The removal of the predicted bias from the base model (red line), lead to a significant increase in the fidelity of the simulation with the red line looking substantially more like the black.

Over the USA sites, the base model overestimated at all times, consistent with previous work (Travis et al., 2016), with the largest biases occurring during the night. The bias-corrected model shows a diurnal cycle very similar to that observed. Averaged across the three American sites the NMB reduced from a mean of 0.51 to -0.02, RMSEs reduced from a mean of 15.1 ppb to 1.1 ppb, and Pearson R values increased from a mean of 0.92 to 1.00. The bias correction thus successfully corrected biases seen in the diurnal cycle, notably the significant night-time bias.

Although the base model failure is less evident for the European sites, again consistent with previous work (L. Hu, Keller, et al., 2018), there were small improvements with the inclusion of the bias correction.

The Japanese data shows a differing pattern. Similar to the USA sites, the base model over-estimates the O_3, generating a much smaller diurnal cycle than observed. Although the bias correction improves the median value, it did not entirely correct the diurnal cycle. This is attributed to the coastal nature of Japan. The model grid-box containing the Japanese observations is mainly oceanic, but the observations show a continental diurnal cycle (a significant increase in O_3 during the day similar to those seen in the USA). If there is a fundamental mismatch between the model's description of the site and the reality (ocean verse land), the bias-predictor will form false relationships. Another potential reason for the reduced performance is that Japan experiences a substantial long-range O_3 influence from China. In the current configuration of the model, only local features are included in the training data. Thus the bias-predictor may not identify influences from the surrounding area. Future work with the bias corrector could explore some regional markers for important compounds. These regional markers and thus transportation biases could be incorporated via the inclusion of the surrounding grid box concentrations or via aggregated regional variables such as the mean 5 x 5 grid box value.

For the two clean tropical sites (Cape Verde and Cape Point in South Africa) the base model already performs reasonably (T. Sherwen et al., 2016), so the bias-corrected version improves little and slightly reduced the NMB performance at Cape Verde from 0.03 to 0.04. For the Antarctic site, the notable bias evident in the model (T. Sherwen et al., 2016) was almost entirely removed by the bias correction, but that resulted in a small reduction in the R-value.

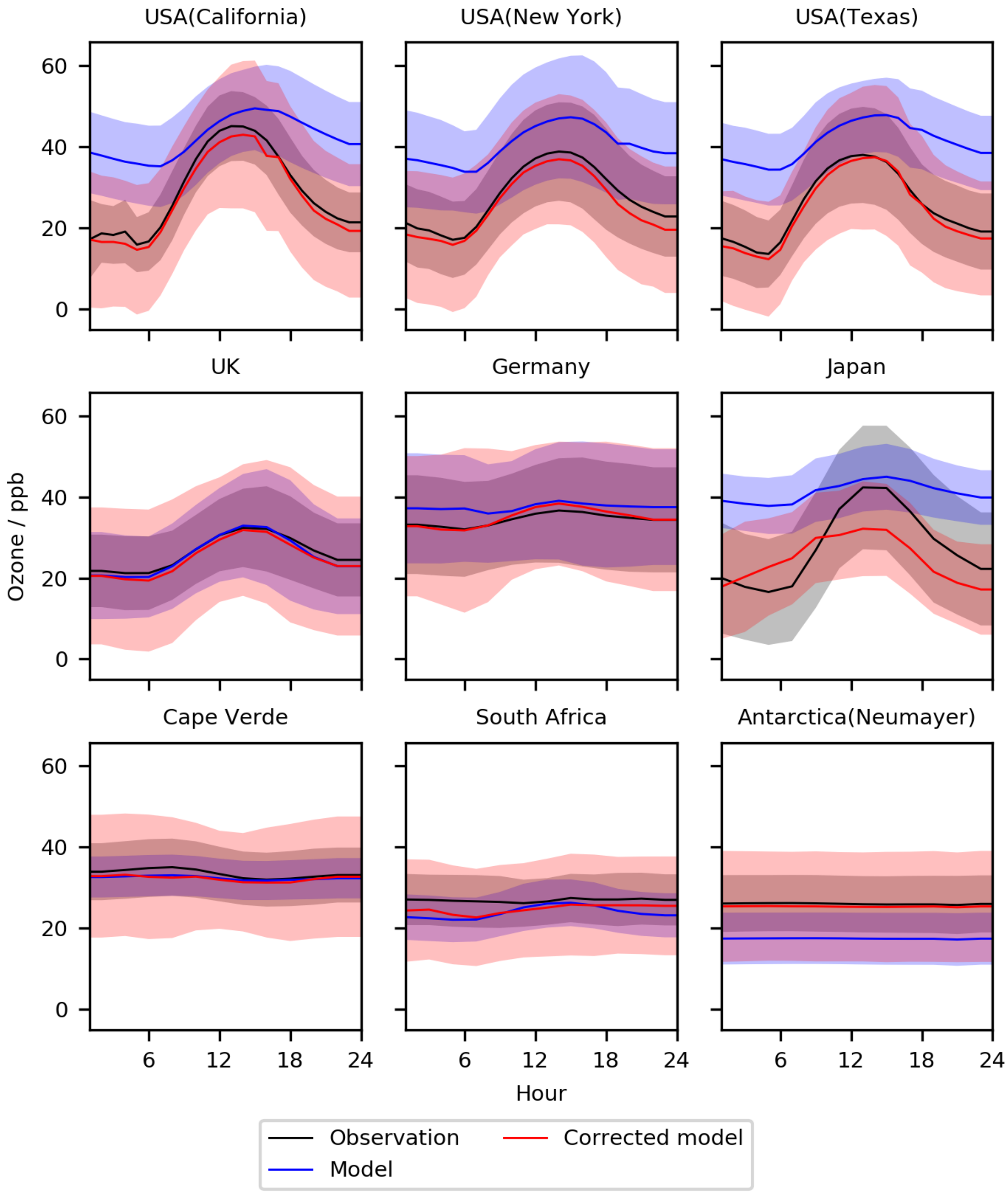

Figure 4.6: Median diurnal cycle for O_3 at nine meta sites in 2016-2017. Shown are the observations, the base model and the model corrected with the bias predictor. The median values are shown as the continuous line and the 25th to 75th percentiles as shaded areas.

Table 4.3: Statistics for diurnal variability at the nine selected sites for the period 1/1/2016-31/12/2017, for the base model (BM) and the model with the bias correction applied (BC)

	Pearson's R		RMSE / ppb		NMB	
Site	BM	BC	BM	BC	BM	BC
USA (California)	0.852	0.997	14.74	1.98	0.46	-0.06
USA (New York)	0.970	0.994	13.12	2.25	0.46	-0.08
USA (Texas)	0.915	0.998	16.29	1.45	0.62	-0.05
UK	0.993	0.998	1.02	1.39	-0.02	-0.05
Germany	0.791	0.991	3.25	0.92	0.09	0.01
Japan	0.98	0.764	14.9	6.94	0.48	-0.12
Cape Verde	0.994	0.812	1.23	1.38	-0.03	-0.04
South Africa (Cape Point)	0.081	0.616	3.32	2.34	-0.11	-0.08
Antarctica (Neumayer)	0.883	0.872	8.57	0.67	-0.33	-0.03

Seasonal variation

The seasonal median comparison for the nine sites can been seen in Fig 4.7 with statistics given in Table 4.4. Over the polluted sites (USA, UK, Germany), biases are effectively removed by the inclusion of bias correction. The performance for Japan was less improved, with the clean tropical sites again showing only small improvements. Over Antarctica, significant bias was removed with the application of bias correction.

Where the performance statistics of the model were already good, such as the RMSE at Cape Verde, or for the NMB at the UK, the inclusion of the bias correction can slightly degrade performance (Table 4.4).

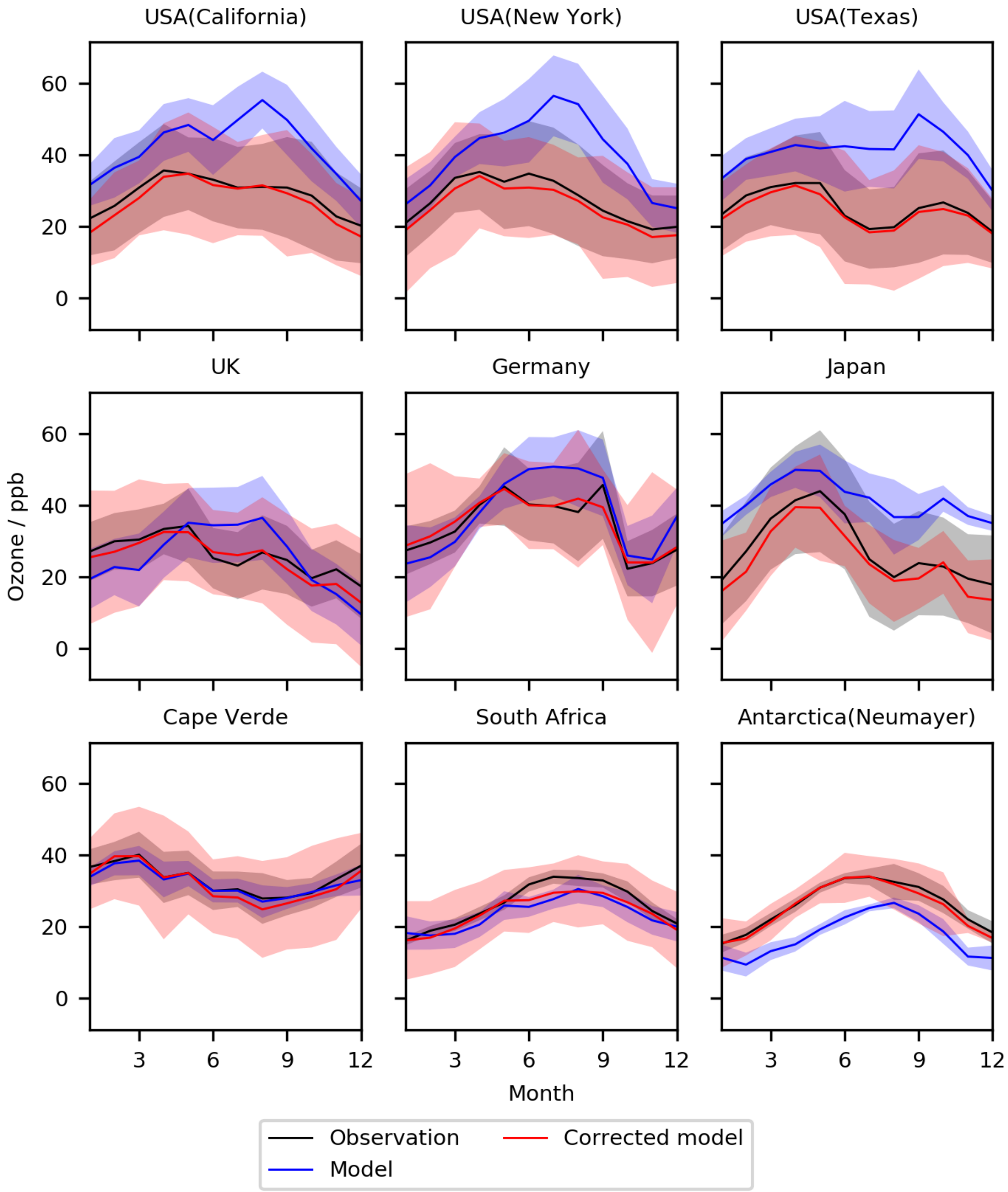

Figure 4.7: Median seasonal cycle for O_3 at nine meta sites in 2016-2017. Shown are the observations, the base model and the model corrected with the bias predictor. The median values are shown as the continuous lines and the 25th to 75th percentiles as shaded areas.

Table 4.4: Statistics for seasonal variability at the nine selected sites for the period 1/1/2016-31/12/2017, for the base model (BM) and the model with the bias correction applied (BC)

	Pearson's R		RMSE / ppb		NMB	
Site	BM	BC	BM	BC	BM	BC
USA (California)	0.833	0.987	14.02	2.19	0.45	-0.06
USA (New York)	0.759	0.992	14.51	2.23	0.46	-0.08
USA (Texas)	0.335	0.991	16.64	1.45	0.62	-0.05
UK	0.519	0.935	7.27	2.51	-0.03	-0.05
Germany	0.848	0.956	6.55	2.42	0.09	0.01
Japan	0.939	0.972	14.0	3.92	0.48	-0.12
Cape Verde	0.956	0.978	1.61	1.73	-0.03	-0.04
South Africa (Cape Point)	0.953	0.976	3.6	2.63	-0.11	-0.08
Antarctica (Neumayer)	0.939	0.993	8.86	1.04	-0.33	-0.03

4.7.3 Sonde data

A point by point comparison between the sonde data in the testing period and the model with and without the bias corrector is shown in Figure 4.8. Compared to observations, the bias correction lowered the model biases (NMB) taking it from 0.03 to 0.02, reduced the error (RMSE) from 12.9 ppb to 10.5 ppb and increased the correlation (Pearson's R) from 0.73 to 0.78.

These improvements are small compared to the improvements seen at the ground. This may be for a number of reasons.

GEOS-Chem already performs better in the free troposphere than at the surface (L. Hu, Keller, et al., 2018) thus there was less overall bias to correct, so any improvement would likely be small.

The methodology assumes that there are no biases between the observational techniques. However, the measurements from the sondes used electrochemical sensors rather than the UV absorption technique at the ground. There is a long history of corrections to electro-chemical sondes to bring them into line with the UV method and some of the differences may also be explained by this (Sterling et al., 2018). The spatial and temporal coverage of the sondes is also limited, with most sonde flights taking place only at noon.

This resulting in a much smaller dataset than for the surface observations. It is not apparent how much information from the surface observations is being used when correcting in the upper troposphere. This imbalance in observations may be resulting in reduced improvements in the sonde data. A possible way to improve on the performance would be to add further free troposphere data. The obvious dataset to include would be those from the In-service aircraft for a global observing system (IAGOS) data. This consists of observations made from commercial flights and provides a significant set of observations that could be used (Nedelec et al., 2015).

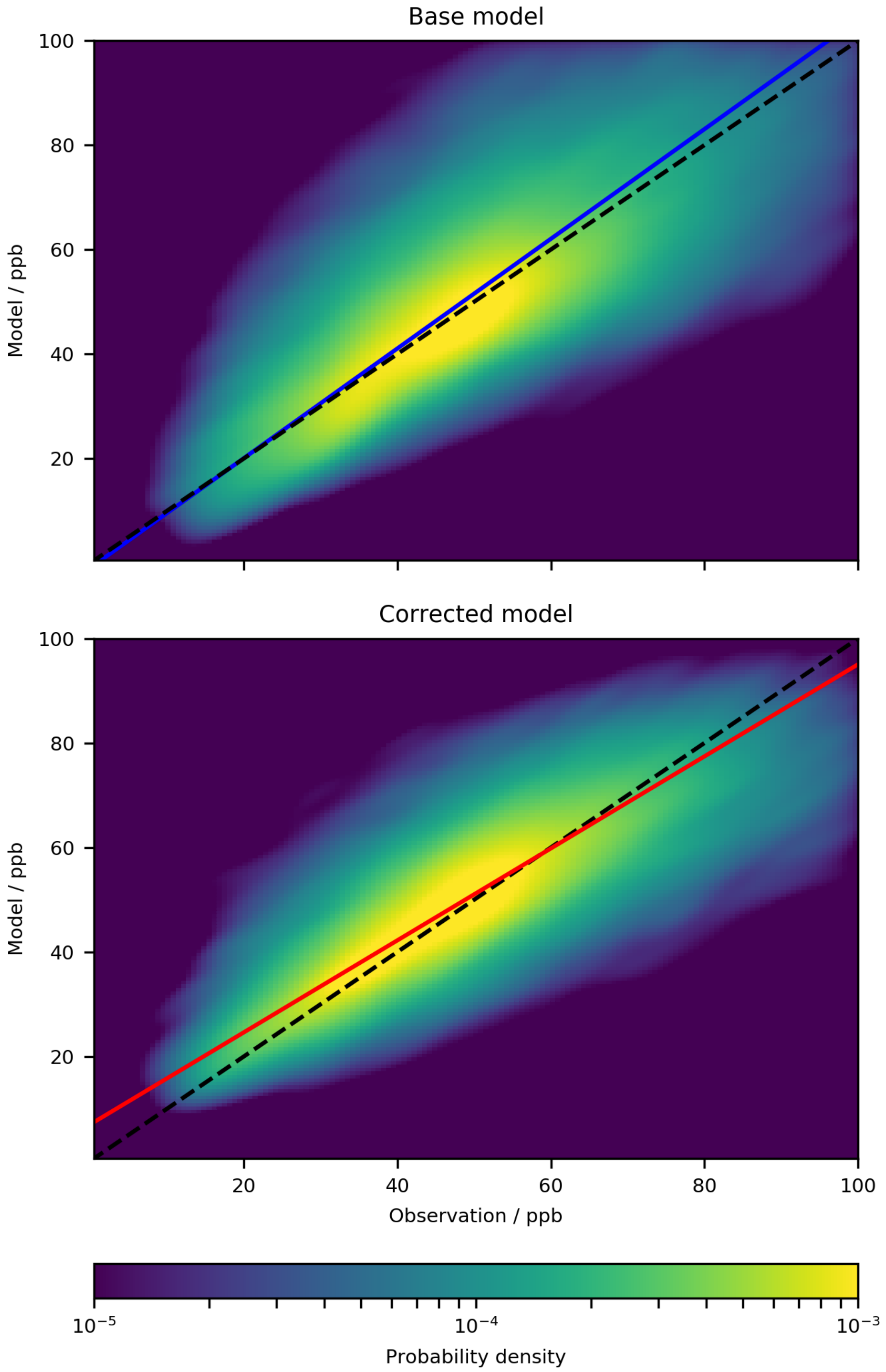

Figure 4.8: Kernel density estimation plot of model verses observation for all sonde sites in the base model (upper panel) and corrected model (lower panel) for the testing period (1/1/2016 to 31/12/2017). The dashed black line is the 1:1 line, the coloured line show the orthogonal line of best fit.

To examine model performance in the free troposphere in more detail, the median vertical O_3 concentration at three sonde locations (central Europe, Colorado and Australia) were examined (Fig. 4.9 with statistics in Table 4.5).

At pressures greater than 400 hPa (below ~7 km) the base model made a reasonable prediction of O_3. Below this pressure the Colorado site was biased low, whereas the Australian site was biased high. Both of which were corrected by the algorithm. At pressures lower than 400 hPa the model started to show a profoundly high bias, which may be a problem with the stratospheric flux (Greenslade et al., 2017). This could result in too much O_3 being entrained from the stratosphere in the model and enriching the concentration in the upper troposphere. The correction algorithm improved all three locations at this lower pressure.

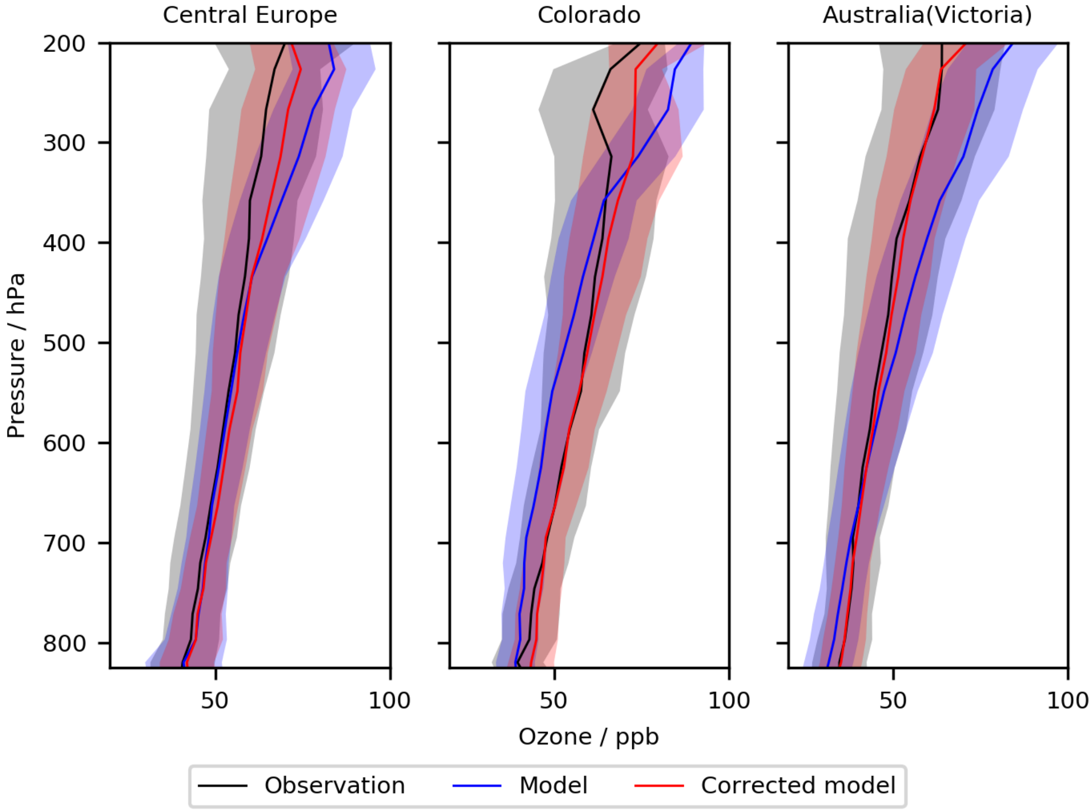

Figure 4.9: Vertical profile for O_3 at sonde sites in 2016-2017. Shown are the observations (black), the base model (blue) and the model corrected with the bias-predictor (red). The median values are shown as the continuous lines and the 25th to 75th percentiles as shaded areas.

Table 4.5: Statistics for the sonde vertical profiles, for the base model (BM) and the bias corrected model (BC).

	Pearson's R		RMSE / ppb	
Site	BM	BC	BM	BC
Central Europe	0.922	0.979	8.54	3.79
Colorado	0.898	0.972	8.35	3.86
Australia	0.923	0.814	11.89	9.58

4.7.4 ATom data

While the evaluation of the bias predictor has been for a different period to the training data, the locations used for testing and training have been kept consistent. It would be preferable to use a completely different dataset to evaluate the performance of the system.

The ATom dataset (described in Section 4.4) provides this independent evaluation. Fig. 4.10 shows the point by point comparison with and without the bias correction. The bias correction lowers the model bias (NMB) taking it from 0.08 to 0.06, reduces the error (RMSE) from 12.1 ppb to 10.5 ppb and increases the correlation (Pearson's R) from 0.76 to 0.79.

These biases and their improvements are similar to those seen with the sondes. Likewise, the improvement is significantly smaller than that seen for the surface data. The RMSE was reduced by only 13% compared to ATom observations compared to 54% for the surface observations. Similarly, Pearson's R only marginally improved with the use of the bias correction. Much of the improvement in the model's performance for the ATom flights will be coming from the observations collected by the sonde network. As mentioned (Section 4.4), the sonde observations are significantly less temporally and spatially resolved than the surface observations. Thus for the bias-predictor to learn the bias, there must be significant volumes of observations under sufficiently diverse conditions. It would appear that the sonde network may not provide that level of information.

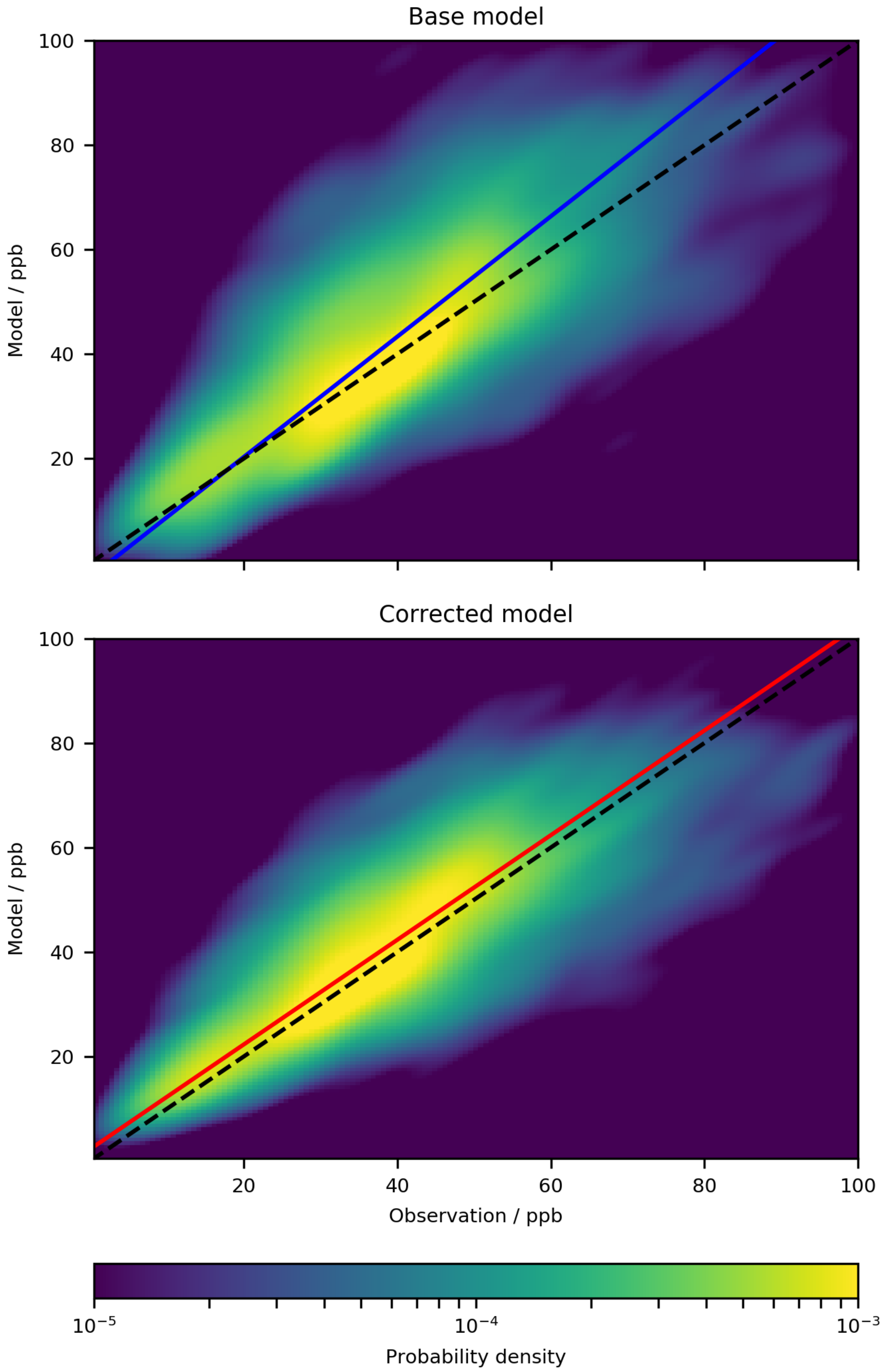

Figure 4.10: Kernel density estimation plot of model verses observation for all ATom locations in the base model (upper panel) and corrected model (lower panel) for the testing period (1/1/2016 to 31/12/2017). The dashed black line is the 1:1 line, the coloured line show the orthogonal line of best fit. The plot is made up of 10,518 data points.

This testing shows that the use of the bias correction improves the performance of the model as compared to independent test data. In the next section, the correction is applied to every grid box, to examine the global distribution of the bias correction. .

4.8 Global Correction

As described in Sect. 4.3, the local model state was output for every grid box for the year 2016. While the original plan was to run the bias-predictor on every grid box for every hour, this produced a vast amount of data (>3 Tb). The size of the dataset resulted in unfeasible running times for the algorithm. To reduce the computational load only the 1st and 15th day of every month was used to get the approximate annual average predicted bias. The predicted bias globally and zonally, can be seen in Fig. 4.11 and Fig. 4.12 respectively.

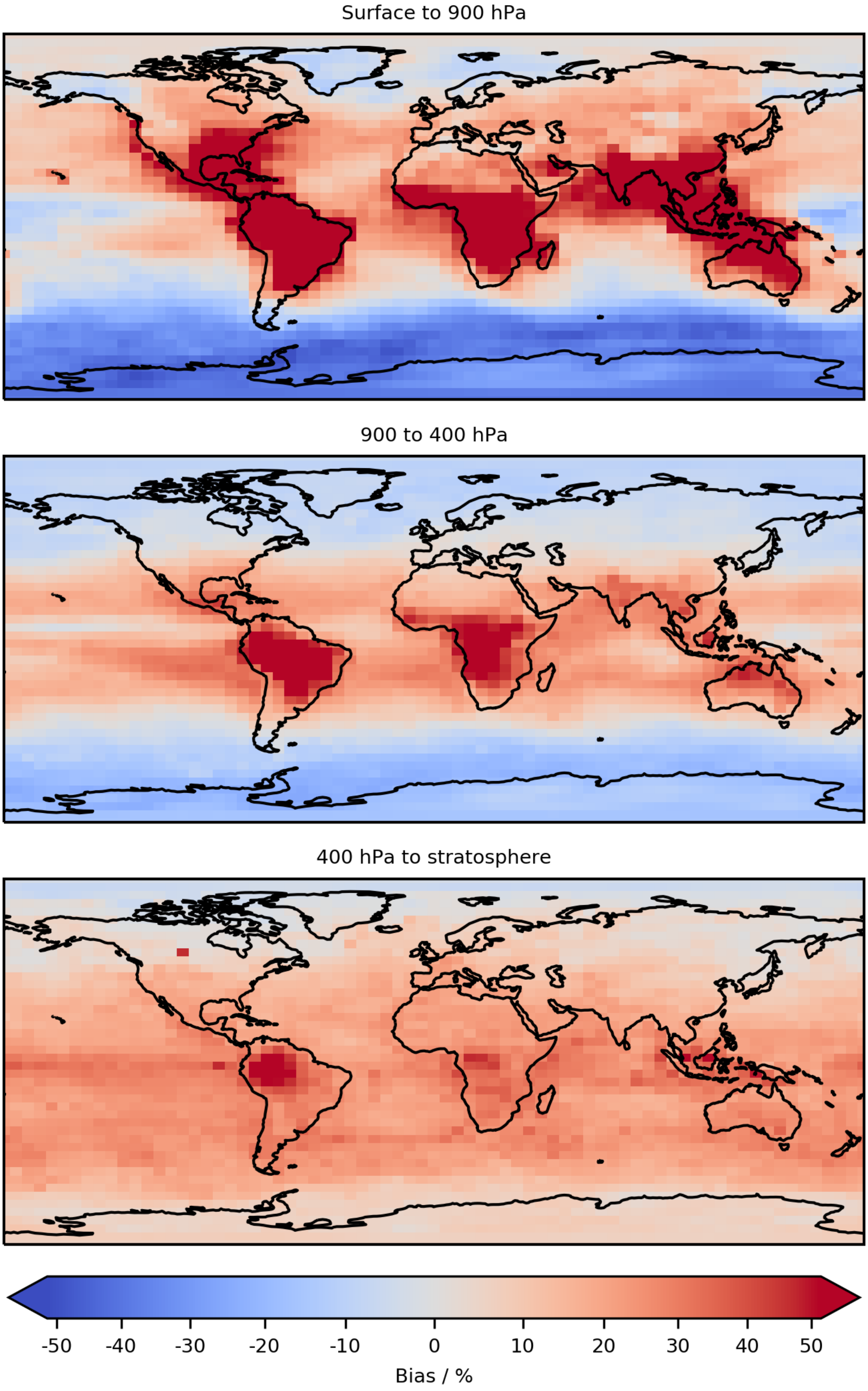

Figure 4.11: Percentage annual mean predicted bias for the lower, middle and upper troposphere.

Many of the features of the predicted O_3 bias have been identified in previous studies. The southeastern United States has been previously recognised as being biased high, with a daytime bias of 0 - 19 ppb (Travis et al., 2016; J. J. Guo et al., 2018). L. Hu, Keller, et al., 2018 also found that the model was biased high at the surface in Europe but with a smaller overall annual bias than the United States. This is consistent with the predictions made in Fig. 4.11. The low O_3 bias in the extra tropic Southern Hemisphere was described in T. Sherwen et al., 2016 and Pound et al., 2019, with the latter blaming the simplistic approximations in the O_3 deposition velocities. Schmidt et al., 2016 identified the Southern Hemisphere low bias extending to approximately 8 km in altitude.

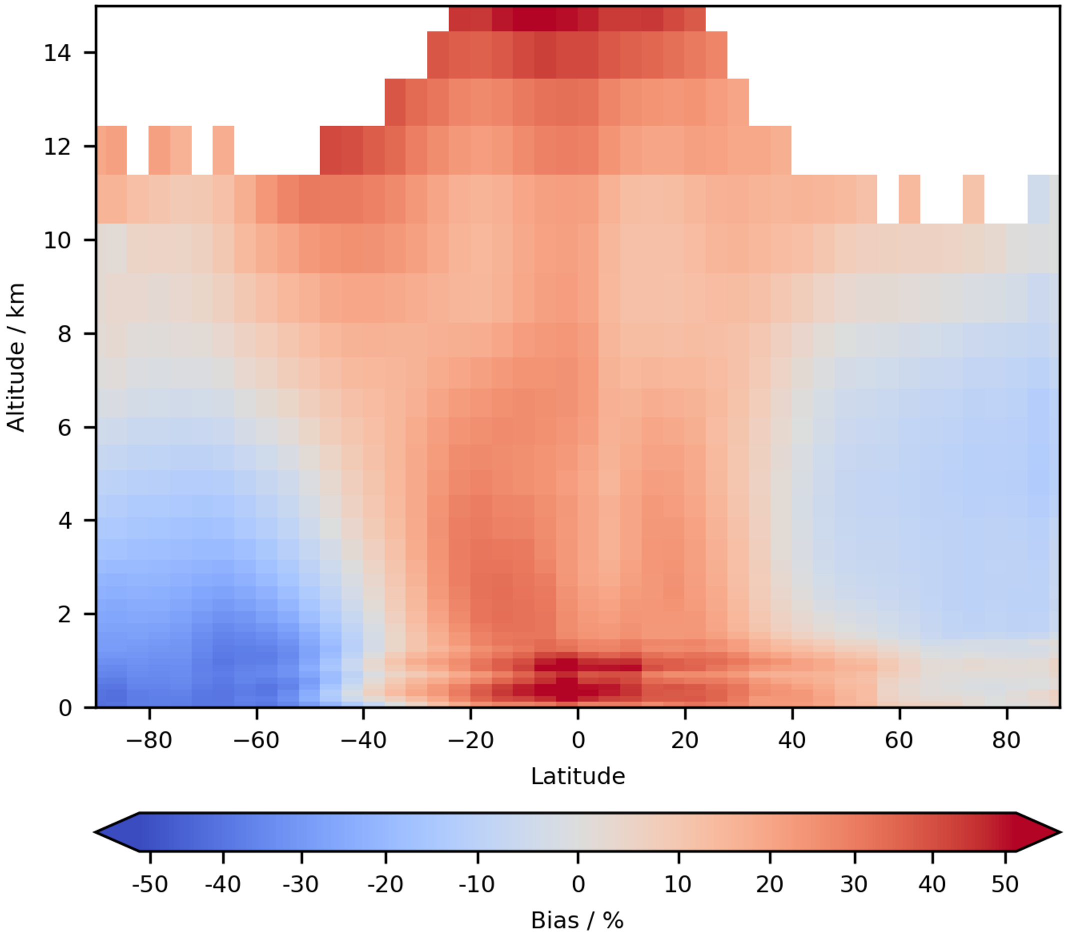

Figure 4.12: Percentage zonal annual mean predicted bias.

Understanding the model bias for regions outside the training data such as south Asia is difficult. The algorithm calculated corrections based on what it learnt from regions with similar chemical or physical conditions. So in the case of south Asia, it appears the correction was based on the correction applied to the southeastern United States. Both these areas have high biogenic and anthropogenic emissions. However, any error in the emissions in Asia is likely to be different from that of the United States. The degree to which the bias corrector can be "trusted" in these regions is thus unknown. Future work is necessary to develop the technique to understand biases in the prediction and their robustness.

One way to test the stability of the prediction is to remove some of the training data and retrain the bias-predictor. A comparison between the different bias predictors gives some information about the robustness of the predictor. This analysis is performed in the next section.

4.8.1 Data denial

To understand the uncertainty in the bias correction a so-called "data denial experiment" is used. This technique assesses the impact of removing individual or regional sets of observations from the training dataset. Two sets of experiments are performed here, first the data from the nine sites examined in Sect. 4.7.2 are removed from the training dataset, and in the second experiment, all western hemisphere points are removed from the training dataset.

Data from the nine meta sites shown in Figures 4.13 and 4.14 (California, New York, Texas, UK, Germany, Japan, Cape Verde, South Africa (Cape Point), Antarctica (Neumayer)) were removed from the training dataset and the bias-predictor was retrained. The retrained bias-predictor was then used to re-correct the testing period data.

In the USA, removing the nine observational data sets did degrade the overall model performance slightly (the green lines in Figures 4.13 and 4.14) compared to the full training dataset (red line). However, this reduction is relatively small, so it appears that the neighbouring sites and other similar sites around the world (notably Europe) were similar enough to the removed sites (CA, NY and TX) to provide sufficient information to almost completely correct the bias. There were different degrees of impact for the other sites.

For the UK, the impact of removing the nine sites from the training dataset was minimal. For Germany, the bias corrections were more substantial and overcompensated the base model during the night and in the summer months. Removal of the Japan information provided a simulation halfway between the simulation with and without the standard bias correction. For remote sites, such as Cape Verde and South Africa, their removal made the bias-corrected model worse than the base model. Similar to Japan, removing the Antarctic site led to a bias correction which is between the standard bias-corrected model and the standard model.

Much of this behaviour is related to the similarity of the removed sites to other sites in the training data. For sites such as the US, and to some extent Europe, removing a few sites had little influence on the bias-predictor, as there were several similar neighbouring sites which could provide similar information. For other locations such as the "clean" Cape Verde and South African sites, there were no other similar sites in the training data. Thus removing those sites removed significant amounts of information. If there were no similar sites for the bias correction to use, an inappropriate correction could be applied, which made the simulation worse. For places such as the Japanese and Antarctic sites, there were some similar sites (Fig. 4.2) in the training data, which provided some improvement over the base model.

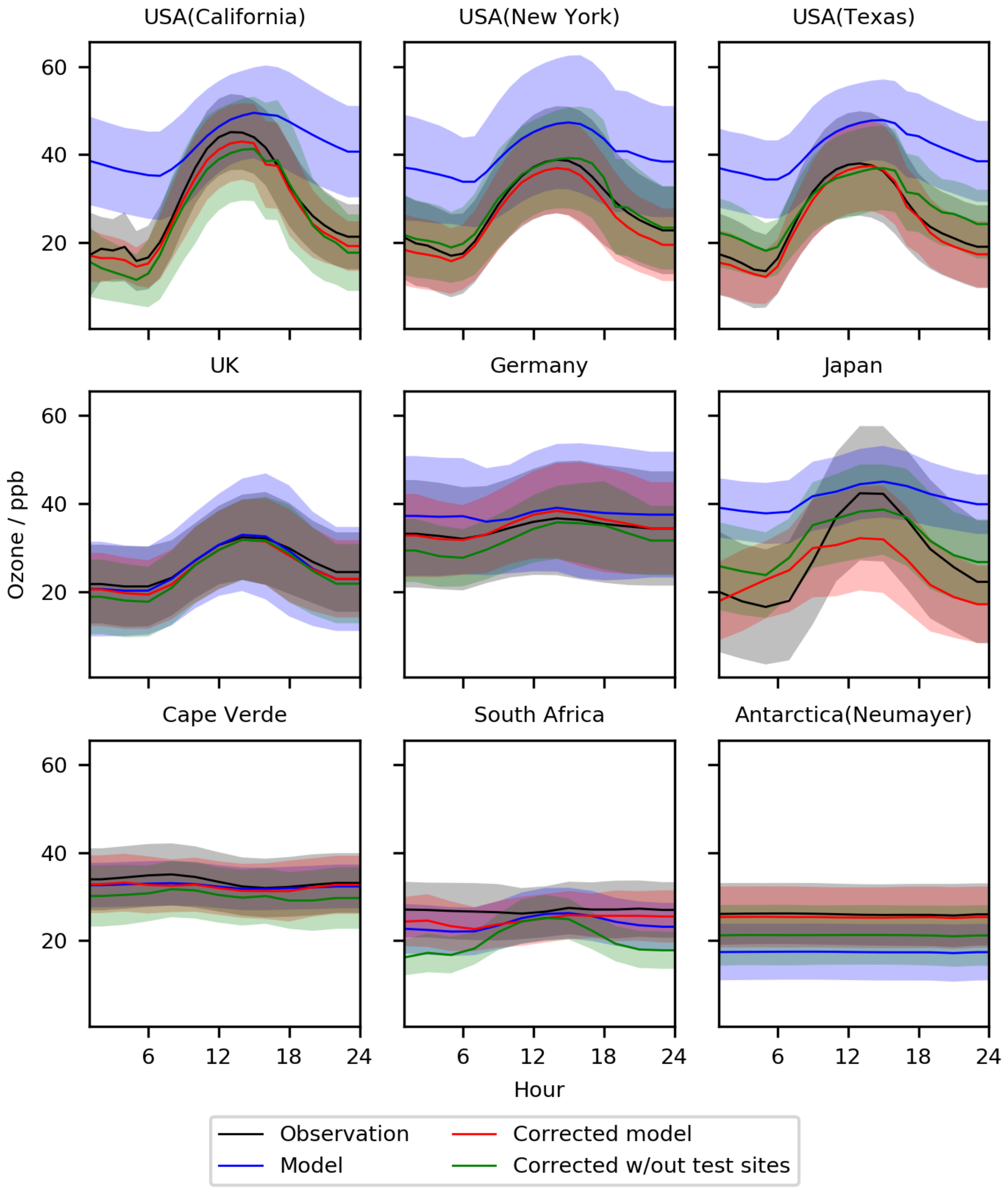

Figure 4.13: Median diurnal cycle for O_3 at nine meta sites in 2016-2017. Shown are the observations (black), the base model (blue), corrected model trained with all observations (red) and the model trained with the nine sites removed (green). The median values are shown as the continuous line and the 25th to 75th percentiles as shaded areas.

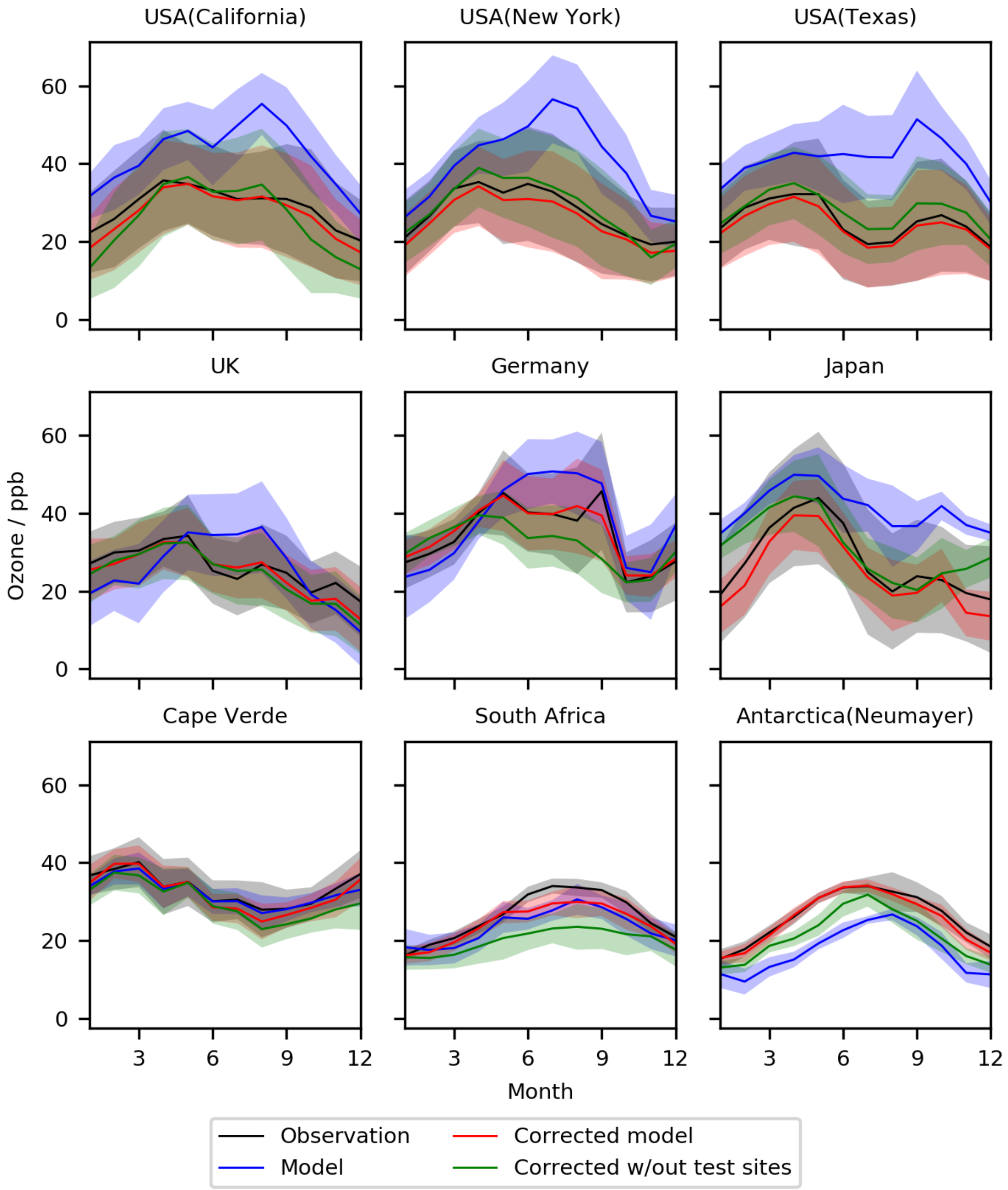

Figure 4.14: Median seasonal cycle for O_3 at nine meta sites in 2016-2017. Shown are the observations (black), the base model (blue), corrected model trained with all observations (red) and the model trained with the nine sites removed (green). The median values are shown as the continuous line and the 25th to 75th percentiles as shaded areas.

Taking the data denial experiment further, observations within North and South America (everything between -180° and -10° East) were removed from the training dataset. Figures 4.15 and 4.16 show the impact of this on the standard nine sites. Surprisingly for New York and Texas, the bias-corrected model performed almost as well without any of North and South America as it did with the full training dataset. The bias corrector predicted roughly the same correction for California as it does for New York and Texas. This over-corrected the daytime concentration for California but simulated the night time and the seasonal cycle much better than without the bias corrector. For the other six sites around the world, the influence of removing North and South America was minimal.

It appears surprising that the corrections applied for North America were so good even though the North American data was not included within the training dataset. This suggests that there are some common reasons for biases between, say North America and Europe. This may indicate a common global source of some of the bias due to errors in the model's chemistry or meteorology, which would be global (or at least northern hemisphere land based) rather than a local source of bias.

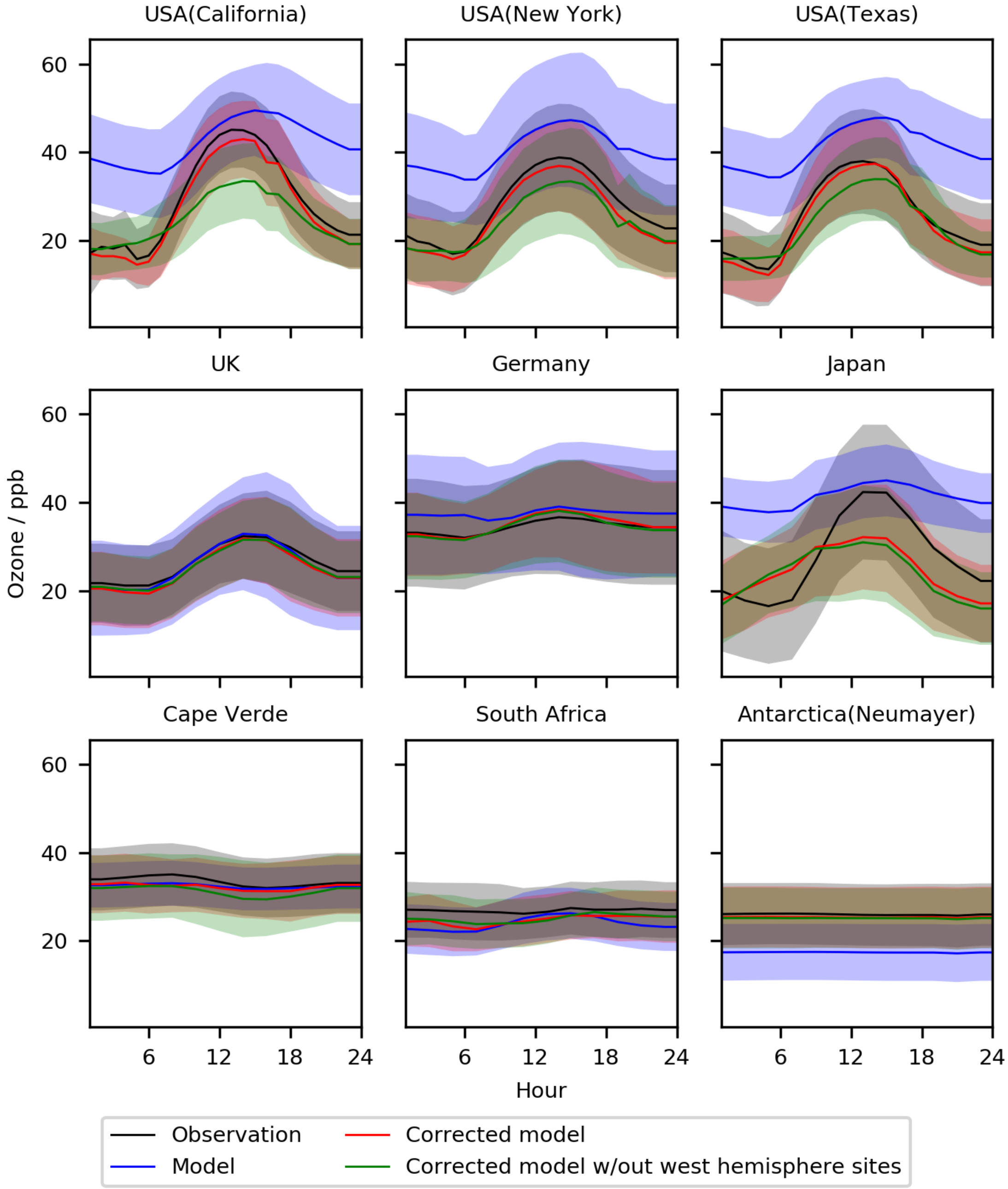

Figure 4.15: Median diurnal cycle for O_3 at nine meta sites in 2016-2017. Shown are the observations (black), the base model (blue), corrected model trained with all observations (red) and the model trained with Western Hemisphere (west of -20^{o}E) data removed (green). The median values are shown as the continuous line and the 25th to 75th percentiles as shaded areas.

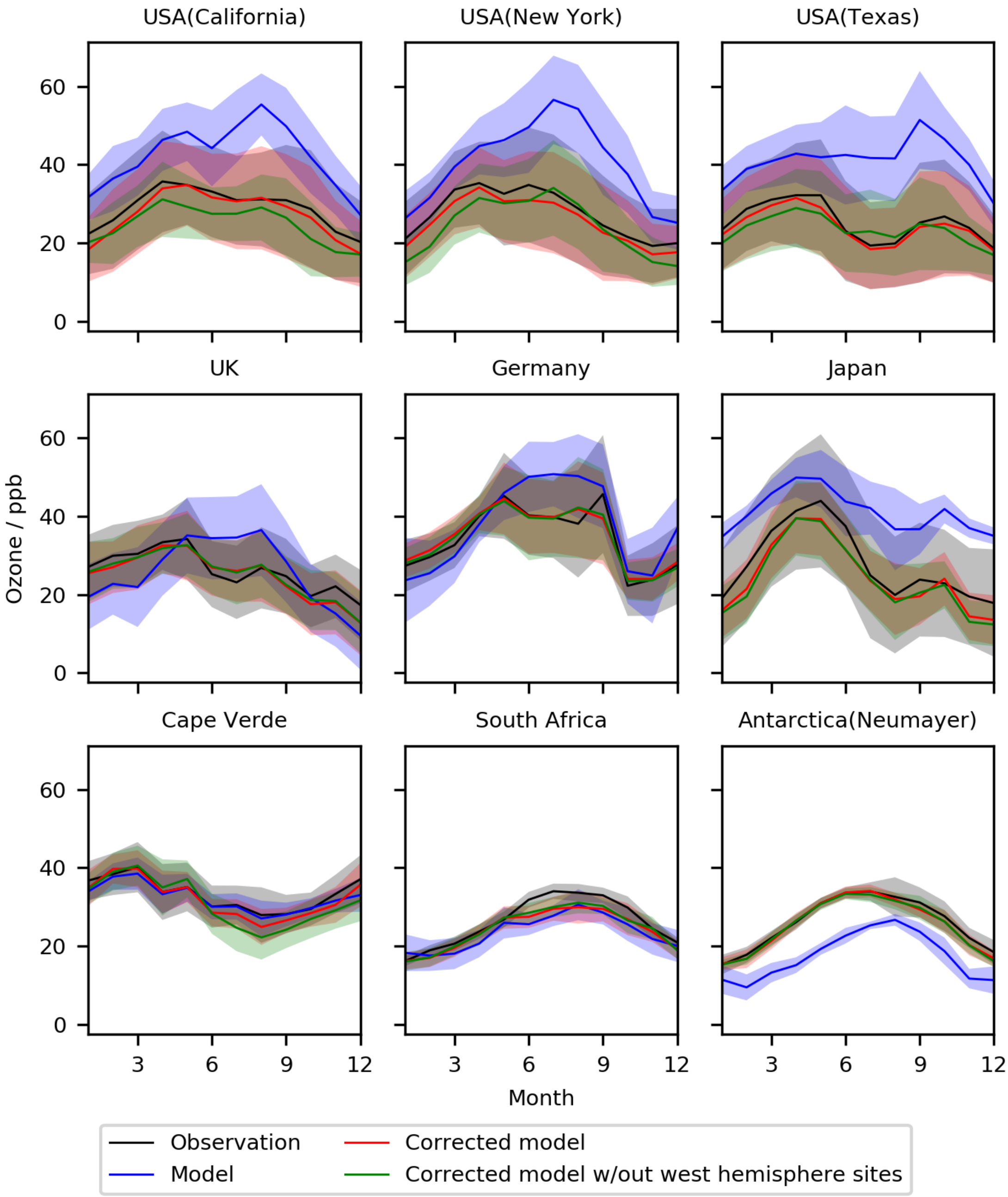

Figure 4.16: Median seasonal cycle for O_3 at nine meta sites in 2016-2017. Shown are the observations (black), the base model (blue), corrected model trained with all observations (red) and the model trained with Western Hemisphere (west of -20^{o}E) data removed (green). The median values are shown as the continuous line and the 25th to 75th percentiles as shaded areas.

The retrained bias-predictor with no Western Hemisphere data was used to predict the mean global percentage bias as was done in Section 4.8. The difference in the predicted absolute bias between the two bias predictors, one with all the data and one with western hemisphere only is shown in Fig. 4.17.

For most sites where there are remaining observations (Europe and GAW sites) changes in the bias prediction is minimal. However removing the Western Hemisphere sites changed the prediction of some sites that still have observations in the training set, particularly Spain. The algorithm was making some of its decisions based on similar biases observed at sites in America. Spain being a polluted subtropical environment, is chemically similar to sites in North America. The changes were much more profound in areas that had no observations of their own. Removing the Western Hemisphere reduces the number of unique environments the algorithm could learn from.

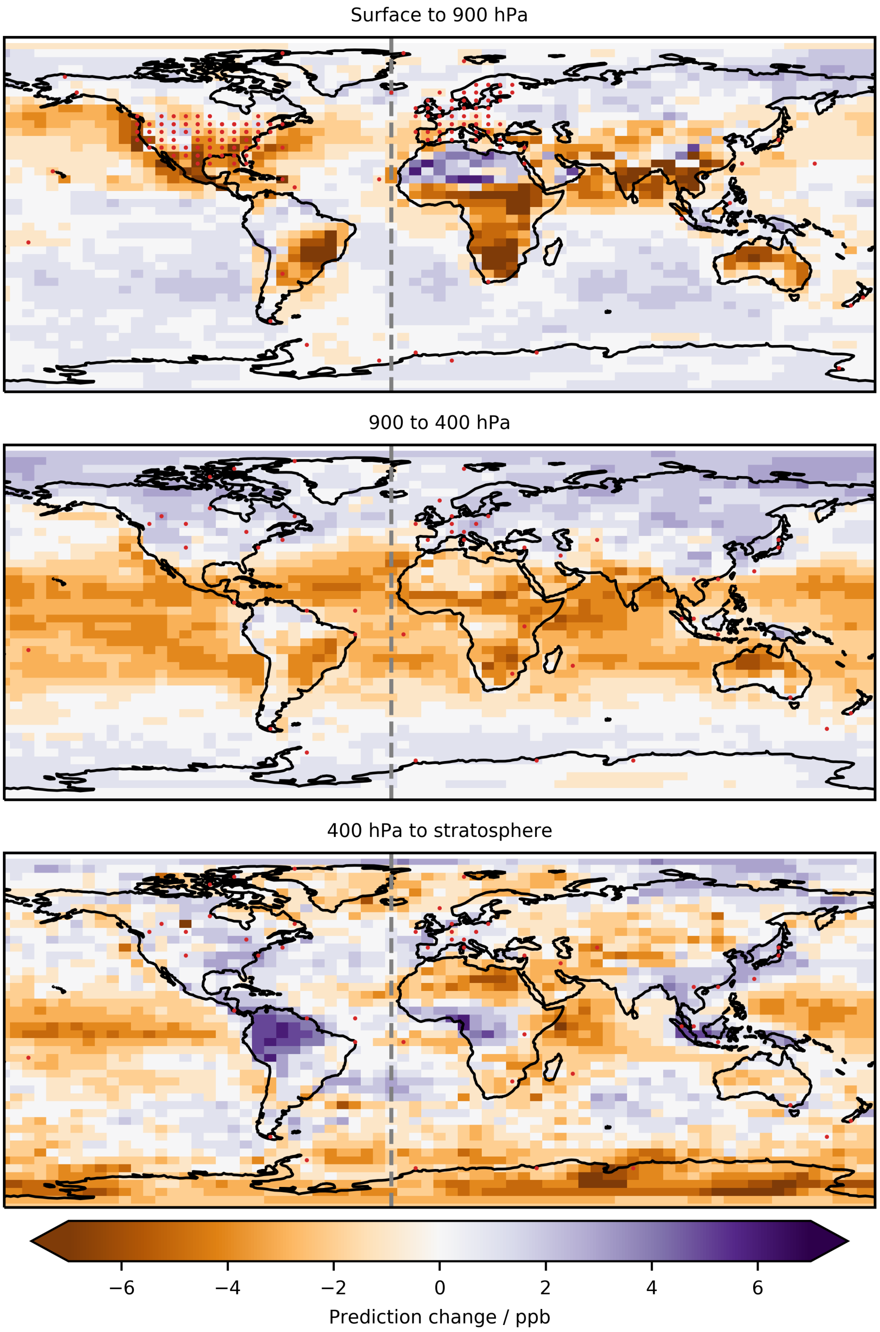

Figure 4.17: Difference in the global mean annual surface O_3 prediction between a predictor trained with western hemisphere observation data (west of -20^{o}E) removed and a predictor trained without this data. Red dots show locations of ground sites in the surface to 900 hPa plot, and sonde locations in the other two plots. The dashed line shows the -20^{o}E line

These types of data denial experiments may provide for an ability to look for explanations of model failings which could be used to help improve the process level representation within models. Another form of data denial is a reduction in the temporal length of the training set, which will be investigated next.

4.9 Training duration

In Section 4.5 the bias-predictor was trained using six years of data (2010-2015). When incorporating certain observational data sets, this long training time provides a challenge. For some critical locations such as China or India the observational record is not that long. For high-resolution model data (e.g. The NASA GEOS global 12.5 km simulation (L. Hu, Keller, et al., 2018)) managing and processing 73 parameters for six years could be computationally burdensome for globally distributed data. Being able to reduce the number of years of data while maintaining the usefulness of the approach would, therefore, be useful.

The bias-predictor was retrained with varying lengths of training data. The end time was kept as 31st of December 2015, while the start date was steadily moved back in time. Figure 4.18 shows the change in the global performance of the model metrics for surface O_3 compared to the number of training months used (the diagnostics are the same as those described in Sect. 4.7.1). Training with only a month of data showed some small benefit, but significant benefit arose once at least eight months of training data had been included. Using the bias-predictor trained with a year of observational data increased the performance of the base model, halved the RMSE, removed most of the NMB and increased the Pearson's R by 60%.

Much of the variability in the power-spectrum of surface O_3 is captured by timescales of a year or less (Bowdalo, Mathew J. Evans, and Sofen, 2016) thus a timescale of a year appears to be the best balance between computational burden and utility for an operational system such as air quality forecasting.

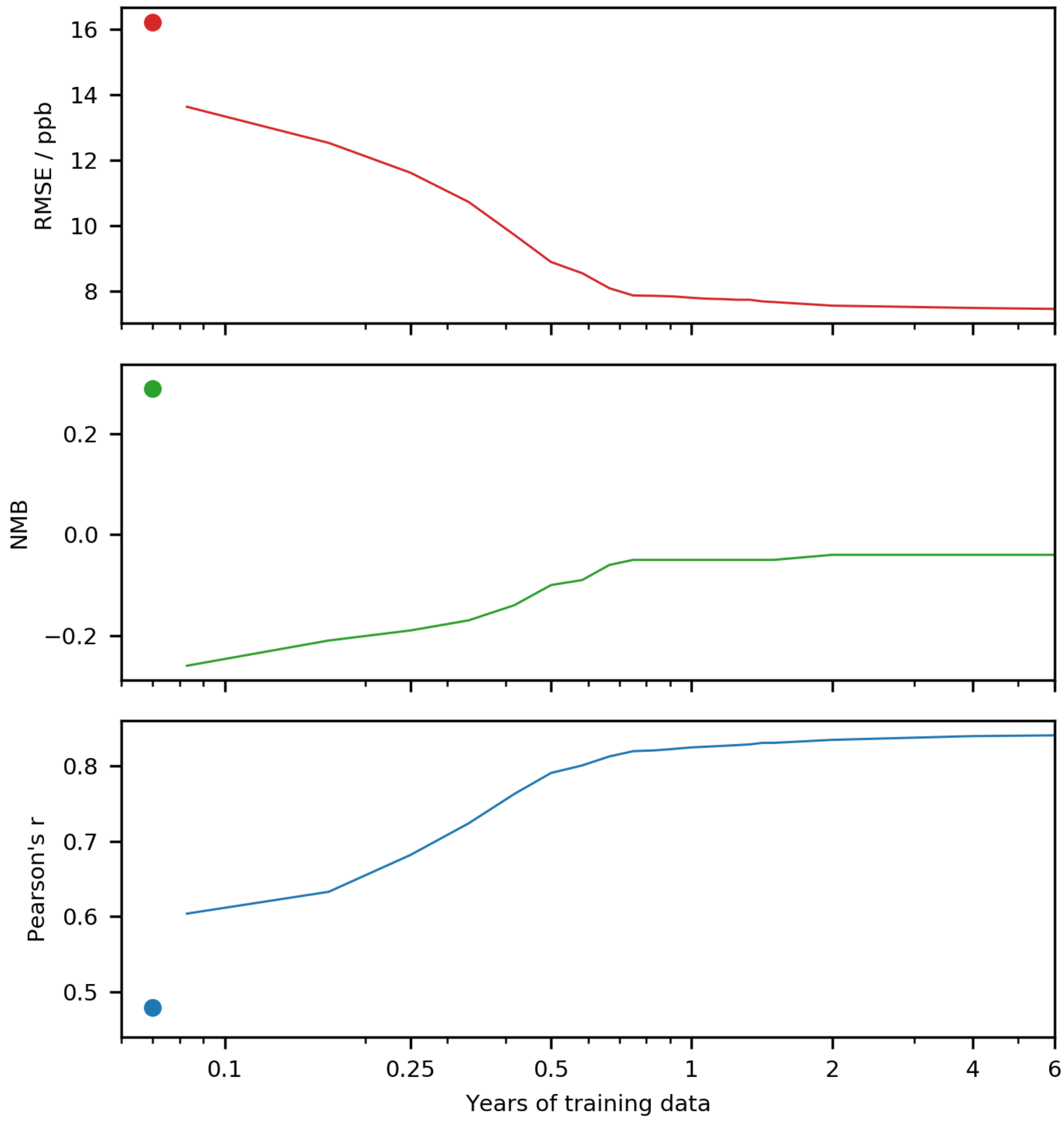

Figure 4.18: Performance change with increasing length of training data. The dot in each plot represents the uncorrected model performance.

Up to this point, the focus has been on varying what data was used to train the predictor. In the next section, an alternative method for correction is tested.

4.10 Direct O_3 concentration prediction

The bias correction method described so far, attempts to predict the model bias ($[O3]_{model}$ / $[O3]_{measured}$). An alternative approach would be to predict $[O_3]$ directly, based on the same chemical and physical input variables. An algorithm to do this, given the same model local state information, was trained on the standard six years of the training period. For the testing period, the direct prediction of surface O_3 performed marginally better than the bias correction for most metrics (Table. 4.6). RMSE decreased to 7.1 ppb from 7.5 ppbv, NMB decreased to 0.00 from -0.04, and Pearson's R increased to 0.85 from 0.84, but for some metrics, the performance is less good.

Table 4.6: Statistical performance for the period 1/1/2016-31/12/2017 of the base model, model with a predicted bias correction applied, and directly predicted O_3 concentration for the surface.

	RMSE / ppb	NMB	Pearson's R
Base O_3	16.21	0.29	0.479
Corrected O_3	7.48	-0.04	0.841
Predicted O_3	7.11	0.00	0.850

For the ATom dataset, the bias predictor performs better than the direct prediction (Table 4.7). This is interpreted here to mean that for locations where observations were included in training (surface sites and sondes), directly predicting $[O_3]$ has benefits. However, for sites where no observations were used in the training, it is better to use the bias corrector approach. Further work is necessary to advance understanding in the form of the prediction that is necessary to provide the most useful enhancement of the system.

Table 4.7: Statistical performance for the period 1/1/2016-31/12/2017 of the base model, model with a predicted bias correction applied, and directly predicted O_3 concentration for the ATom data.

	RMSE / ppb	NMB	Pearson's R
Base O_3	12.11	0.08	0.761
Corrected O_3	10.50	0.06	0.792
Predicted O_3	10.92	0.11	0.797

4.11 Conclusions

The bias in O_3 concentration calculated by a chemistry transport model can be reduced through the use of a machine learning algorithm. The results appeared robust to data denial and training length experiments. For activities such as air quality forecasting for sites with a long observational record, bias-prediction appeared to offer a route to significant improvements in the fidelity of the forecasts without having to improve the process-level understanding. This work offers some practical advantages over data assimilation. The observations do not necessarily need to be available in real-time. The training of the bias-predictor can be made using past observations and applied to a forecast without the need for the latest observations. The bias-predictor approach may also be applied to regions where observational data is not available. Although necessitating care, the temporary lack of available data is much less of a problem for this approach than for other methods of data assimilation.

Further work on this approach could include running the model at a higher resolution in order to resolve a more significant number of observation sites and allow for the evaluation of the averaging performed in this study. Further experimentation with other machine learning techniques such as recurrent and convolutional neural network architectures could also prove useful. These more complex architectures would allow additional spatial and temporal information to be incorporated into the prediction. Another area of interest would be to predict bias in other species such as NO_x and particulate matter. Finally, there appears to be significant scope for exploring coupling such bias correction methods with an operational air quality forecasting system to provide improved air quality forecasts.

Chapter 5

Conclusions

5.1 Conclusions

This thesis has explored some of our understanding of the processes controlling O_3 on a regional to global scale. This topic has been the subject of intense scientific interest over the last 40 years. However, there are still challenging science problems to be explored. The focus of the first two science chapters was on the city of Beijing. China is now the focus of much of the world's research into air pollution, given the high population and the highly polluted air experienced. The Atmospheric Pollution and Human Health in a Chinese Mega-city project (APHH-Beijing) campaign provided a comprehensive observational dataset to explore our understanding of air pollution in Beijing during the summer of 2017.

This observational dataset, together with the GEOS-Chem chemistry transport model, was used to explore the processes controlling O_3 in the city of Beijing. After tuning the primary emissions in the model to fit the observations (which identified a number of failings in the model emissions), the focus turned to an evaluation of the short lived species in the model. These radical species play a fundamental role in determining the chemistry of the atmosphere, yet are rarely evaluated within a chemistry transport model.

This evaluation identified some key failures in the model's ability to simulate the radical chemistry.

- Failure to reproduce NO to NO_2 ratio
- Failure to reproduce day time HONO concentrations
- Underestimation of OH reactivity
- Underestimation of RO_2 concentrations
- Overestimation of $j(O(^1D)$

Some of these failures were evident in box model simulations of the same data by Whalley et al. (2020) and so these problems are not unique to the GEOS-Chem model. These failures limit our confidence in our ability to understand the processes controlling O_3 in urban environments and so future work to address these problems is fundamental to developing more robust policy options in the future.

Despite these problems with the short-lived radical chemistry, the model was able to reproduce much of the variability observed in the O_3 concentration throughout the campaign. This suggest that the some of the errors in the model, may, to some extend be compensating. The model may be simulating the right O_3 for the wrong reasons.

Ozone isopleths have been used for decades as a tool to understand the photochemical environment in a city. They have typically been calculated using box models of various complexity but given the computational resources now available, the chemistry transport model was used to calculate O_3 isopleths for the city. These isopleths showed the typical structure of a NO_x and VOC limited regimes. In the mean, Beijing was found to be in the VOC limited regime during the field campaign in 2017. In that situation a reduction in NO_x would lead to an increase in the O_3 concentration. The magnitude of this increase was found to reduce, depending on the metric used to assess the impact. The maximum O_3 during the campaign would increase very little if NO_x emissions were reduced, whereas the mean O_3 could increase by 7 ppbv.

Emissions of the aromatics (notably the xylenes) and isoprene (mainly from outside of the city) were found to be the most important VOCs for O_3 production. Reducing the total VOC emissions, and notably the concentrations of the xylenes and other aromatics is most likely the best route to reducing O_3 in Beijing. Reducing the regional emissions of isoprene would be difficult to achieve without devastation of the regional forests.

Uptake of HO_2 onto aerosol surfaces was found to be an important inhibitor of O_3 production under some circumstances. By diagnosing both the model and the observations, the magnitude of the different radical sinks (OH+NO2, RO_x+RO_x and HO_2 uptake) could be evaluated. Ternary diagrams were developed to show the relative importance of the different sink processes both in the model and the observations and these showed that despite the problems in the model's fast photo-chemistry, the balance between the sinks was approximately correct in the model.

Novel O_3 isopleths were developed which varied NO_x emission and and the magnitude of aerosol uptake of HO_2. This identified 3 photochemical regimes:

- Aerosol insensitive (High NOx). In this regime the $OH+NO_2$ reaction is overwhelmingly the dominant radical sink. Variations in HO_2 uptake don't make a significant impact on the radical concentration so O_3 is insensitive to the aerosol surface area density.

- Aerosol insensitive (Low NOx). In this regime the RO_x+RO_x reaction are overwhelmingly the dominant radical sink. Variation in HO_2 uptake does not make a significant impact on the radical concentration so O_3 is again insensitive to the aerosol surface area density.

- Aerosol sensitive. In this regime there is a more equitable balance between the different radical sinks and so a change in the rate of HO_2 uptake onto aerosols can have an impact on the radical concentration and so the O_3 concentration.

By diagnosing the different radical sink processes it was possible to explore which of the different dominant sinks was occurring spatially and temporally over the domain. While the radical loss in Beijing is predominantly NO_2 + OH dominated, prevailing winds can shift this to a more RO_x + RO_x or HO_2 uptake dominated regime. Large areas of the domain were in each of the different loss regimes. Thus assessing our understanding of each of these regimes is important.

Up to this point the focus has been on assessing and improving the processes represented in the model. However, substantial model biases still occur. For some applications (air quality forecasting, understanding the spatial distribution of O_3) this is problematic. The final chapter, invoked a machine learning based approach to remove biases without improving process level understanding.

The machine learning approach learned the bias from a large dataset of observations that were compared to the model result. The algorithm was able to learn the bias from a six year training period and then improve the results of a two year test period. The algorithm was able to learn some degree of bias for all nine globally distributed test sites. Further to this, the algorithm was able to identify published known O_3 biases in a variety of locations.

Additionally, the technique displayed robustness to a reduction in observational data. This technique could have significant potential use in systems that do not require a process level understanding such as air quality forecasts.

To conclude, chemical transport models should be evaluated against short lived species. There are still significant challenges ahead with regards to simulating O_3. Underestimation of the RO_2 and HONO concentration, as well as the missing OH reactivity, could lead to a substantially increased model O_3 production and as of yet, there is not a clear candidate for the removal of this added O_3.

Machine learning will likely revolutionize our ability to use and understand large datasets and provide computationally efficient solutions for solving known unknowns. Future work will be required to identify how best to incorporate machine learning techniques into science.

5.1.1 Closing remarks

This thesis sought to improve and increase understanding of how a chemical transport model reproduces observed O_3. Through the inclusion of improved chemistry and revised emissions, the performance of the model was improved. This improvement was explicit, meaning the exact changes in the model were known. The results were analysed, including developing new investigatory and visualisation techniques resulting in a thorough understanding of what chemical processes were governing observed trends. However, the evaluation was time consuming, required an extensive local data set and was only evaluated for one location at one time of year.

From a performance metric perspective, an algorithmic approach also improved the representation of O_3 within the model. While the algorithm only offered inferred ways of identifying model behaviour, it provided a substantial increase in the model's ability to replicate O_3 across the world, at all layers of the troposphere and all times of the year.

Process-based improvements to models are ideal, in a perfect world, every process that governs the atmosphere would be replicated mathematically in the model. Models represent our current understanding of atmospheric science and will continue to improve. However, due to increasing computational power, increasing access to large datasets, and developing more advanced machine learning techniques, algorithmically augmented models will likely have a substantial role in science in the future.

Bibliography

Acharya, U. R. et al. (2018). “Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals”. In: *Computers in Biology and Medicine* 100, pp. 270–278. DOI: `10.1016/j.compbiomed.2017.09.017`.

Akimoto, Hajime et al. (2019). “Comparison of surface ozone simulation among selected regional models in MICS-Asia III - effects of chemistry and vertical transport for the causes of difference”. In: *Atmospheric Chemistry and Physics* 19.1, pp. 603–615. DOI: `10.5194/acp-19-603-2019`.

Alexander, B. et al. (2005). “Sulfate formation in sea-salt aerosols: Constraints from oxygen isotopes”. In: *Journal of Geophysical Research-Atmospheres* 110.D10. DOI: `10.1029/2004jd005659`.

Aliwell, S. R. and R. L. Jones (1998). “Measurements of tropospheric NO3 at midlatitude”. In: *Journal of Geophysical Research-Atmospheres* 103.D5, pp. 5719–5727. DOI: `10.1029/97jd03119`.

Altshuller, A. P. (1977). “EYE IRRITATION AS AN EFFECT OF PHOTOCHEMICAL AIR-POLLUTION”. In: *Journal of the Air Pollution Control Association* 27.11, pp. 1125–1126. DOI: `10.1080/00022470.1977.10470538`.

Amos, H. M. et al. (2012). “Gas-particle partitioning of atmospheric Hg(II) and its effect on global mercury deposition”. In: *Atmospheric Chemistry and Physics* 12.1, pp. 591–603. DOI: `10.5194/acp-12-591-2012`.

Andreae, M. O. and P. Merlet (2001). “Emission of trace gases and aerosols from biomass burning”. In: *Global Biogeochemical Cycles* 15.4, pp. 955–966. DOI: `10.1029/2000gb001382`.

Archer, D. (2005). “Fate of fossil fuel CO2 in geologic time”. In: *Journal of Geophysical Research-Oceans* 110.C9. DOI: `10.1029/2004jc002625`.

Astitha, M. et al. (2012). “Parameterization of dust emissions in the global atmospheric chemistry-climate model EMAC: impact of nudging and soil properties”. In: *Atmospheric Chemistry and Physics* 12.22, pp. 11057–11083. DOI: `10.5194/acp-12-11057-2012`.

Atkinson, R. (2000). “Atmospheric chemistry of VOCs and NOx”. In: *Atmospheric Environment* 34.12-14, pp. 2063–2101. DOI: `10.1016/s1352-2310(99)00460-4`.

Atkinson, R. and J. Arey (2003a). “Atmospheric degradation of volatile organic compounds”. In: *Chemical Reviews* 103.12, pp. 4605–4638. DOI: `10.1021/cr0206420`.

— (2003b). “Gas-phase tropospheric chemistry of biogenic volatile organic compounds: a review”. In: *Atmospheric Environment* 37, S197–S219. DOI: `10.1016/s1352-2310(03)00391-1`.

Atkinson, R., A. M. Winer, and J. N. Pitts (1986). “ESTIMATION OF NIGHTTIME N2O5 CONCENTRATIONS FROM AMBIENT NO2 AND NO3 RADICAL CONCENTRATIONS AND THE ROLE OF N2O5 IN NIGHTTIME CHEMISTRY”. In: *Atmospheric Environment* 20.2, pp. 331–339. DOI: `10.1016/0004-6981(86)90035-1`.

Ballabio, D., F. Grisoni, and R. Todeschini (2018). “Multivariate comparison of classification performance measures”. In: *Chemometrics and Intelligent Laboratory Systems* 174, pp. 33–44. DOI: `10.1016/j.chemolab.2017.12.004`.

Bates, K. H. and D. J. Jacob (2019). “A new model mechanism for atmospheric oxidation of isoprene: global effects on oxidants, nitrogen oxides, organic products, and secondary organic aerosol”. In: *Atmospheric Chemistry and Physics* 19.14, pp. 9613–9640. DOI: `10.5194/acp-19-9613-2019`.

— (2020). “An Expanded Definition of the Odd Oxygen Family for Tropospheric Ozone Budgets: Implications for Ozone Lifetime and Stratospheric Influence”. In: *Geophysical Research Letters* 47.4. DOI: `10.1029/2019gl084486`.

Battin, Tom J. et al. (2009). “The boundless carbon cycle”. In: *Nature Geoscience* 2.9, pp. 598–600. DOI: `10.1038/ngeo618`.

Bauer, Peter, Alan Thorpe, and Gilbert Brunet (2015). “The quiet revolution of numerical weather prediction”. In: *Nature* 525.7567, pp. 47–55. DOI: `10.1038/nature14956`.

Bauer, S. E. et al. (2004). “Global modeling of heterogeneous chemistry on mineral aerosol surfaces: Influence on tropospheric ozone chemistry and comparison to observations”. In: *Journal of Geophysical Research-Atmospheres* 109.D2. DOI: `10.1029/2003jd003868`.

Bell, M. L., D. L. Davis, and T. Fletcher (2004). “A retrospective assessment of mortality from the London smog episode of 1952: The role of influenza and pollution”. In: *Environmental Health Perspectives* 112.1, pp. 6–8. DOI: `10.1289/ehp.6539`.

Bengio, Y. (2000). “Gradient-based optimization of hyperparameters”. In: *Neural Computation* 12.8, pp. 1889–1900. DOI: `10.1162/089976600300015187`.

Bernstein, Jonathan A. et al. (2008). “The health effects of nonindustrial indoor air pollution”. In: *Journal of Allergy and Clinical Immunology* 121.3, pp. 585–591. DOI: `10.1016/j.jaci.2007.10.045`.

Bey, I., B. Aumont, and G. Toupance (2001). “A modeling study of the nighttime radical chemistry in the lower continental troposphere 1. Development of a detailed chemical mechanism including nighttime chemistry”. In: *Journal of Geophysical Research-Atmospheres* 106.D9, pp. 9959–9990. DOI: `10.1029/2000jd900347`.

Bey, I., D. J. Jacob, et al. (2001). “Global modeling of tropospheric chemistry with assimilated meteorology: Model description and evaluation”. In: *Journal of Geophysical Research-Atmospheres* 106.D19, pp. 23073–23095. DOI: `10.1029/2001jd000807`.

Bian, H. S. and M. J. Prather (2002). “Fast-J2: Accurate simulation of stratospheric photolysis in global chemical models”. In: *Journal of Atmospheric Chemistry* 41.3, pp. 281–296. DOI: `10.1023/a:1014980619462`.

Blanchard, C. L. and D. Fairley (2001). “Spatial mapping of VOC and NOx-limitation of ozone formation in central California”. In: *Atmospheric Environment* 35.22, pp. 3861–3873. DOI: `10.1016/s1352-2310(01)00153-4`.

Bloss, C. et al. (2005). “Development of a detailed chemical mechanism (MCMv3.1) for the atmospheric oxidation of aromatic hydrocarbons”. In: *Atmospheric Chemistry and Physics* 5, pp. 641–664. DOI: `10.5194/acp-5-641-2005`.

Bloss, W. J. et al. (2005). “The oxidative capacity of the troposphere: Coupling of field measurements of OH and a global chemistry transport model”. In: *Faraday Discussions* 130, pp. 425–436. DOI: `10.1039/b419090d`.

Bocquet, M. et al. (2015). “Data assimilation in atmospheric chemistry models: current status and future prospects for coupled chemistry meteorology models”. In: *Atmospheric Chemistry and Physics* 15.10, pp. 5325–5358. DOI: `10.5194/acp-15-5325-2015`.

Bofill, J. M. et al. (1999). “The mechanism of methoxy radical oxidation by O(2) in the gas phase. Computational evidence for direct H atom transfer assisted by an intermolecular noncovalent O center dot center dot center dot O bonding interaction”. In: *Journal of the American Chemical Society* 121.6, pp. 1337–1347. DOI: `10.1021/ja981926y`.

Bohn, B. et al. (2016). “Characterisation and improvement of j((OD)-D-1) filter radiometers”. In: *Atmospheric Measurement Techniques* 9.7, pp. 3455–3466. DOI: `10.5194/amt-9-3455-2016`.

Bottenheim, J. W., K. A. Brice, and K. G. Anlauf (1984). “DISCUSSION OF A LAGRANGIAN TRAJECTORY MODEL DESCRIBING LONG-RANGE TRANSPORT OF OXIDES OF NITROGEN, THE INCORPORATION OF PAN IN THE CHEMICAL MECHANISM, AND SUPPORTING MEASUREMENTS OF PAN AND NITRATE SPECIES AT RURAL SITES IN ONTARIO, CANADA”. In: *Atmospheric Environment* 18.12, pp. 2609–2619. DOI: `10.1016/0004-6981(84)90325-1`.

Bowdalo, Dene R., Mathew J. Evans, and Eric D. Sofen (2016). “Spectral analysis of atmospheric composition: application to surface ozone model-measurement comparisons”. In: *Atmospheric Chemistry and Physics* 16.13, pp. 8295–8308. DOI: `10.5194/acp-16-8295-2016`.

Breiman, L. (2001). “Random forests”. In: *Machine Learning* 45.1, pp. 5–32. DOI: `10.1023/a:1010933404324`.

Brimblecombe, P. (1976). “ATTITUDES AND RESPONSES TOWARDS AIR-POLLUTION IN MEDIEVAL ENGLAND”. In: *Journal of the Air Pollution Control Association* 26.10, pp. 941–945. DOI: `10.1080/00022470.1976.10470341`.

Brown, S. S., J. A. Neuman, et al. (2006). “Nocturnal odd-oxygen budget and its implications for ozone loss in the lower troposphere”. In: *Geophysical Research Letters* 33.8. DOI: `10.1029/2006gl025900`.

Brown, S. S., R. K. Talukdar, and A. R. Ravishankara (1999). “Rate constants for the reaction OH+NO2+M -> HNO3+M under atmospheric conditions”. In: *Chemical Physics Letters* 299.3-4, pp. 277–284. DOI: `10.1016/s0009-2614(98)01283-4`.

Brunekreef, Bert et al. (2009). “Effects of long-term exposure to traffic-related air pollution on respiratory and cardiovascular mortality in the Netherlands: the NLCS-AIR study”. In: *Research report (Health Effects Institute)* 139, 5–71, discussion 73–89.

Carter, W. P. L. (1990). “A DETAILED MECHANISM FOR THE GAS-PHASE ATMOSPHERIC REACTIONS OF ORGANIC-COMPOUNDS”. In: *Atmospheric Environment Part a-General Topics* 24.3, pp. 481–518. DOI: `10.1016/0960-1686(90)90005-8`.

— (2010). “Development of a condensed SAPRC-07 chemical mechanism”. In: *Atmospheric Environment* 44.40, pp. 5336–5345. DOI: `10.1016/j.atmosenv.2010.01.024`.

Chameides, W. L. et al. (1992). “OZONE PRECURSOR RELATIONSHIPS IN THE AMBIENT ATMOSPHERE”. In: *Journal of Geophysical Research-Atmospheres* 97.D5, pp. 6037–6055. DOI: `10.1029/91jd03014`.

Chang, W. L. et al. (2016). “Evaluating N2O5 heterogeneous hydrolysis parameterizations for CalNex 2010”. In: *Journal of Geophysical Research-Atmospheres* 121.9, pp. 5051–5070. DOI: `10.1002/2015jd024737`.

Chatani, S. et al. (2014). “Photochemical roles of rapid economic growth and potential abatement strategies on tropospheric ozone over South and East Asia in 2030”. In: *Atmospheric Chemistry and Physics* 14.17, pp. 9259–9277. DOI: `10.5194/acp-14-9259-2014`.

Chen, Tianqi and Carlos Guestrin (2016). “XGBoost: A Scalable Tree Boosting System”. In: *CoRR* abs/1603.02754, pp. 785–794. DOI: `10.1145/2939672.2939785`.

Chen, Tze-Ming et al. (2007). “Outdoor air pollution: Nitrogen dioxide, sulfur dioxide, and carbon monoxide health effects”. In: *American Journal of the Medical Sciences* 333.4, pp. 249–256. DOI: `10.1097/MAJ.0b013e31803b900f`.

Chen, X. et al. (2019). “On the sources and sinks of atmospheric VOCs: an integrated analysis of recent aircraft campaigns over North America”. In: *Atmospheric Chemistry and Physics* 19.14, pp. 9097–9123. DOI: `10.5194/acp-19-9097-2019`.

Chen, Y. et al. (2016). “Long-term variation of black carbon and PM2.5 in Beijing, China with respect to meteorological conditions and governmental measures”. In: *Environmental Pollution* 212, pp. 269–278. DOI: `10.1016/j.envpol.2016.01.008`.

Chen, Y. Z. et al. (2015). “Assessment of SAPRC07 with updated isoprene chemistry against outdoor chamber experiments”. In: *Atmospheric Environment* 105, pp. 109–120. DOI: `10.1016/j.atmosenv.2015.01.042`.

Chen, Ying et al. (2018). “A parameterization of the heterogeneous hydrolysis of N2O5 for mass-based aerosol models: improvement of particulate nitrate prediction”. In: *Atmospheric Chemistry and Physics* 18.2, pp. 673–689. DOI: `10.5194/acp-18-673-2018`.

Cheng, H. R. et al. (2010). “Assessing photochemical ozone formation in the Pearl River Delta with a photochemical trajectory model”. In: *Atmospheric Environment* 44.34, pp. 4199–4208. DOI: `10.1016/j.atmosenv.2010.07.019`.

Chin, M. et al. (1994). “RELATIONSHIP OF OZONE AND CARBON-MONOXIDE OVER NORTH-AMERICA”. In: *Journal of Geophysical Research-Atmospheres* 99.D7, pp. 14565–14573. DOI: `10.1029/94jd00907`.

Cohan, D. S. et al. (2005). “Nonlinear response of ozone to emissions: Source apportionment and sensitivity analysis”. In: *Environmental Science and Technology* 39.17, pp. 6739–6748. DOI: `10.1021/es048664m`.

Cohen, Aaron J. et al. (2017). “Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: an analysis of data from the Global Burden of

Diseases Study 2015". In: *Lancet* 389.10082, pp. 1907–1918. DOI: `10.1016/s0140-6736(17)30505-6`.

Coleman, M. D. et al. (1995). "CARBON ALLOCATION AND PARTITIONING IN ASPEN CLONES VARYING IN SENSITIVITY TO TROPOSPHERIC OZONE". In: *Tree Physiology* 15.9, pp. 593–604. DOI: `10.1093/treephys/15.9.593`.

Couzo, Evan et al. (2015). "Impacts of heterogeneous HONO formation on radical sources and ozone chemistry in Houston, Texas". In: *Atmospheric Environment* 112, pp. 344–355. DOI: `10.1016/j.atmosenv.2015.04.048`.

Cox, P. M. et al. (2000). "Acceleration of global warming due to carbon-cycle feedbacks in a coupled climate model". In: *Nature* 408.6809, pp. 184–187. DOI: `10.1038/35041539`.

Crane, Bret et al. (2018). "China's special economic zones: an analysis of policy to reduce regional disparities". In: *Regional Studies Regional Science* 5.1, pp. 98–107. DOI: `10.1080/21681376.2018.1430612`.

Crippa, M. et al. (2018). "Gridded emissions of air pollutants for the period 1970-2012 within EDGAR v4.3.2". In: *Earth System Science Data* 10.4, pp. 1987–2013. DOI: `10.5194/essd-10-1987-2018`.

Cryer, D. R. (2016). "Measurements of hydroxyl radical reactivity and formaldehyde in the atmosphere, University of Leeds, Thesis".

Cynthia, Rudin (2019). "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead". In: *Nat Mach Intell* 1, pp. 206–215. DOI: `https://doi.org/10.1038/s42256-019-0048-x`.

Damian, V. et al. (2002). "The kinetic preprocessor KPP - a software environment for solving chemical kinetics". In: *Computers and Chemical Engineering* 26.11, pp. 1567–1579. DOI: `10.1016/s0098-1354(02)00128-x`.

Dang, R. J. and H. Liao (2019). "Radiative Forcing and Health Impact of Aerosols and Ozone in China as the Consequence of Clean Air Actions over 2012-2017". In: *Geophysical Research Letters* 46.21, pp. 12511–12519. DOI: `10.1029/2019gl084605`.

Dehghani, Mansooreh et al. (2018). “Characteristics and health effects of BTEX in a hot spot for urban pollution”. In: *Ecotoxicology and Environmental Safety* 155, pp. 133–143. DOI: `10.1016/j.ecoenv.2018.02.065`.

Delmas, R., D. Serca, and C. Jambert (1997). “Global inventory of NOx sources”. In: *Nutrient Cycling in Agroecosystems* 48.1-2, pp. 51–60. DOI: `10.1023/a:1009793806086`.

Demore, W. B. and O. F. Raper (1966). “PRIMARY PROCESSES IN OZONE PHOTOLYSIS”. In: *Journal of Chemical Physics* 44.5, 1780–and. DOI: `10.1063/1.1726939`.

Deng, F. Y. et al. (2020). “A big data approach to improving the vehicle emission inventory in China”. In: *Nature Communications* 11.1. DOI: `10.1038/s41467-020-16579-w`.

Derwent, R. G., M. E. Jenkin, and S. M. Saunders (1996). “Photochemical ozone creation potentials for a large number of reactive hydrocarbons under European conditions”. In: *Atmospheric Environment* 30.2, pp. 181–199. DOI: `10.1016/1352-2310(95)00303-g`.

Donkelaar, A. van et al. (2008). “Analysis of aircraft and satellite measurements from the Intercontinental Chemical Transport Experiment (INTEX-B) to quantify long-range transport of East Asian sulfur to Canada”. In: *Atmospheric Chemistry and Physics* 8.11, pp. 2999–3014. DOI: `10.5194/acp-8-2999-2008`.

Duan, J. C. et al. (2008). “Concentration, sources and ozone formation potential of volatile organic compounds (VOCs) during ozone episode in Beijing”. In: *Atmospheric Research* 88.1, pp. 25–35. DOI: `10.1016/j.atmosres.2007.09.004`.

Duce, R. A. (1978). “SPECULATIONS ON BUDGET OF PARTICULATE AND VAPOR-PHASE NON-METHANE ORGANIC-CARBON IN GLOBAL TROPOSPHERE”. In: *Pure and Applied Geophysics* 116.2-3, pp. 244–273. DOI: `10.1007/bf01636883`.

Eastham, Sebastian D., Debra K. Weisenstein, and Steven R. H. Barrett (2014). “Development and evaluation of the unified tropospheric-stratospheric chemistry extension (UCX) for the global chemistry-transport model GEOS-Chem”. In: *Atmospheric Environment* 89, pp. 52–63. DOI: `10.1016/j.atmosenv.2014.02.001`.

Fairlie, T. D., D. J. Jacob, and R. J. Park (2007). “The impact of transpacific transport of mineral dust in the United States”. In: *Atmospheric Environment* 41.6, pp. 1251–1266. DOI: `10.1016/j.atmosenv.2006.09.048`.

Farber, R. J., P. R. Welsing, and C. Rozzi (1994). “PM10 AND OZONE CONTROL STRATEGY TO IMPROVE VISIBILITY IN THE LOS-ANGELES BASIN”. In: *Atmospheric Environment* 28.20, pp. 3277–3283. DOI: `10.1016/1352-2310(94)00160-m`.

Farman, J. C., B. G. Gardiner, and J. D. Shanklin (1985). “LARGE LOSSES OF TOTAL OZONE IN ANTARCTICA REVEAL SEASONAL CLOX/NOX INTERACTION”. In: *Nature* 315.6016, pp. 207–210. DOI: `10.1038/315207a0`.

Feng, L. and W. J. Liao (2016). “Legislation, plans, and policies for prevention and control of air pollution in China: achievements, challenges, and improvements”. In: *Journal of Cleaner Production* 112, pp. 1549–1558. DOI: `10.1016/j.jclepro.2015.08.013`.

Finlaysonpitts, B. J. and J. N. Pitts (1993). “ATMOSPHERIC CHEMISTRY OF TROPOSPHERIC OZONE FORMATION - SCIENTIFIC AND REGULATORY IMPLICATIONS”. In: *Journal of the Air and Waste Management Association* 43.8, pp. 1091–1100.

Fiore, A. M., L. W. Horowitz, et al. (2005). “Evaluating the contribution of changes in isoprene emissions to surface ozone trends over the eastern United States”. In: *Journal of Geophysical Research-Atmospheres* 110.D12. DOI: `10.1029/2004jd005485`.

Fiore, A. M., D. J. Jacob, et al. (2002). “Linking ozone pollution and climate change: The case for controlling methane”. In: *Geophysical Research Letters* 29.19. DOI: `10.1029/2002gl015601`.

Fischer, E. V. et al. (2014). “Atmospheric peroxyacetyl nitrate (PAN): a global budget and source attribution”. In: *Atmospheric Chemistry and Physics* 14.5, pp. 2679–2698. DOI: `10.5194/acp-14-2679-2014`.

Fisher, J. A., E. L. Atlas, et al. (2018). “Methyl, Ethyl, and Propyl Nitrates: Global Distribution and Impacts on Reactive Nitrogen in Remote Marine Environments”. In: *Jour-*

nal of Geophysical Research-Atmospheres 123.21, pp. 12429–12451. DOI: `10.1029/2018jd029046`.

Fisher, J. A., D. J. Jacob, et al. (2011). "Sources, distribution, and acidity of sulfate-ammonium aerosol in the Arctic in winter-spring". In: *Atmospheric Environment* 45.39, pp. 7301–7318. DOI: `10.1016/j.atmosenv.2011.08.030`.

Flynn, P. F. et al. (2000). "The inevitability of engine-out NOx emissions from spark-ignited and diesel engines". In: *Proceedings of the Combustion Institute* 28, pp. 1211–1218. DOI: `10.1016/s0082-0784(00)80332-x`.

Fountoukis, C. and A. Nenes (2007). "ISORROPIA II: a computationally efficient thermodynamic equilibrium model for K+-Ca2+-Mg2+-Nh(4)(+)-Na+-SO42–NO3–Cl–H2O aerosols". In: *Atmospheric Chemistry and Physics* 7.17, pp. 4639–4659. DOI: `10.5194/acp-7-4639-2007`.

Foy, Benjamin de et al. (2015). "Estimates of power plant NOx emissions and lifetimes from OMI NO2 satellite retrievals". In: *Atmospheric Environment* 116, pp. 1–11. DOI: `10.1016/j.atmosenv.2015.05.056`.

Frery, Jordan et al. (2017). "Efficient Top Rank Optimization with Gradient Boosting for Supervised Anomaly Detection". In: *Machine Learning and Knowledge Discovery in Databases, Ecml Pkdd 2017, Pt I* 10534, pp. 20–35. DOI: `10.1007/978-3-319-71249-9\_2`.

Friend, A. L. and P. T. Tomlinson (1992). "MILD OZONE EXPOSURE ALTERS C-14 DYNAMICS IN FOLIAGE OF PINUS-TAEDA L". In: *Tree Physiology* 11.3, pp. 215–227. DOI: `10.1093/treephys/11.3.215`.

Gardner, M. W. and S. R. Dorling (1998). "Artificial neural networks (the multilayer perceptron) - A review of applications in the atmospheric sciences". In: *Atmospheric Environment* 32.14-15, pp. 2627–2636. DOI: `10.1016/s1352-2310(97)00447-0`.

— (2000). "Statistical surface ozone models: an improved methodology to account for non-linear behaviour". In: *Atmospheric Environment* 34.1, pp. 21–34. DOI: `10.1016/s1352-2310(99)00359-3`.

Geyer, A. et al. (2003). “Nighttime formation of peroxy and hydroxyl radicals during the BERLIOZ campaign: Observations and modeling studies”. In: *Journal of Geophysical Research-Atmospheres* 108.D4. DOI: `10.1029/2001jd000656`.

Ghahramani, Z. (2015). “Probabilistic machine learning and artificial intelligence”. In: *Nature* 521.7553, pp. 452–459. DOI: `10.1038/nature14541`.

Giglio, L., J. T. Randerson, and G. R. van der Werf (2013). “Analysis of daily, monthly, and annual burned area using the fourth-generation global fire emissions database (GFED4)”. In: *Journal of Geophysical Research-Biogeosciences* 118.1, pp. 317–328. DOI: `10.1002/jgrg.20042`.

Goliff, W. S., W. R. Stockwell, and C. V. Lawson (2013). “The regional atmospheric chemistry mechanism, version 2”. In: *Atmospheric Environment* 68, pp. 174–185. DOI: `10.1016/j.atmosenv.2012.11.038`.

Goyer, R. A. (1990). “TRANSPLACENTAL TRANSPORT OF LEAD”. In: *Environmental Health Perspectives* 89, pp. 101–105. DOI: `10.2307/3430905`.

Greenslade, Jesse W. et al. (2017). “Stratospheric ozone intrusion events and their impacts on tropospheric ozone in the Southern Hemisphere”. In: *Atmospheric Chemistry and Physics* 17.17, pp. 10269–10290. DOI: `10.5194/acp-17-10269-2017`.

Guaita, Rosana et al. (2011). “Short-term impact of particulate matter (PM2.5) on respiratory mortality in Madrid”. In: *International Journal of Environmental Health Research* 21.4, pp. 260–274. DOI: `10.1080/09603123.2010.544033`.

Guenther, A. et al. (2006). “Estimates of global terrestrial isoprene emissions using MEGAN (Model of Emissions of Gases and Aerosols from Nature)”. In: *Atmospheric Chemistry and Physics* 6, pp. 3181–3210. DOI: `10.5194/acp-6-3181-2006`.

Guenther, A. B. et al. (2012). “The Model of Emissions of Gases and Aerosols from Nature version 2.1 (MEGAN2.1): an extended and updated framework for modeling biogenic emissions”. In: *Geoscientific Model Development* 5.6, pp. 1471–1492. DOI: `10.5194/gmd-5-1471-2012`.

Guo, J. et al. (2019). “Theoretical evaluation of different factors affecting the HO2 uptake coefficient driven by aqueous-phase first-order loss reaction”. In: *Science of the Total Environment* 683, pp. 146–153. DOI: `10.1016/j.scitotenv.2019.05.237`.

Guo, J. J. et al. (2018). “Average versus high surface ozone levels over the continental USA: model bias, background influences, and interannual variability”. In: *Atmospheric Chemistry and Physics* 18.16, pp. 12123–12140. DOI: `10.5194/acp-18-12123-2018`.

Haagensmit, A. J. (1952). “CHEMISTRY AND PHYSIOLOGY OF LOS-ANGELES SMOG”. In: *Industrial and Engineering Chemistry* 44.6, pp. 1342–1346. DOI: `10.1021/ie50510a045`.

Haagensmit, A. J., C. E. Bradley, and M. M. Fox (1953). “OZONE FORMATION IN PHOTOCHEMICAL OXIDATION OF ORGANIC SUBSTANCES”. In: *Industrial and Engineering Chemistry* 45.9, pp. 2086–2089. DOI: `10.1021/ie50525a044`.

Halonen, Jaana I. et al. (2009). “Particulate Air Pollution and Acute Cardiorespiratory Hospital Admissions and Mortality Among the Elderly”. In: *Epidemiology* 20.1, pp. 143–153. DOI: `10.1097/EDE.0b013e31818c7237`.

Hamer, P. D. et al. (2015). “The impact of observing characteristics on the ability to predict ozone under varying polluted photochemical regimes”. In: *Atmospheric Chemistry and Physics* 15.18, pp. 10645–10667. DOI: `10.5194/acp-15-10645-2015`.

Hampton, C. V. et al. (1983). “HYDROCARBON GASES EMITTED FROM VEHICLES ON THE ROAD .2. DETERMINATION OF EMISSION RATES FROM DIESEL AND SPARK-IGNITION VEHICLES”. In: *Environmental Science and Technology* 17.12, pp. 699–708. DOI: `10.1021/es00118a003`.

Harlass, J., M. Latif, and W. Park (2015). “Improving climate model simulation of tropical Atlantic sea surface temperature: The importance of enhanced vertical atmosphere model resolution”. In: *Geophysical Research Letters* 42.7, pp. 2401–2408. DOI: `10.1002/2015gl063310`.

Heard, D. E. and M. J. Pilling (2003). “Measurement of OH and HO2 in the troposphere”. In: *Chemical Reviews* 103.12, pp. 5163–5198. DOI: `10.1021/cr020522s`.

Heath, Robert L., Allen S. Lefohn, and Robert C. Musselman (2009). “Temporal processes that contribute to nonlinearity in vegetation responses to ozone exposure and dose”. In: *Atmospheric Environment* 43.18, pp. 2919–2928. DOI: `10.1016/j.atmosenv.2009.03.011`.

Henze, D. K., J. H. Seinfeld, N. L. Ng, et al. (2008). “Global modeling of secondary organic aerosol formation from aromatic hydrocarbons: high- vs. low-yield pathways”. In: *Atmospheric Chemistry and Physics* 8.9, pp. 2405–2420. DOI: `10.5194/acp-8-2405-2008`.

Henze, D. K., J. H. Seinfeld, and D. T. Shindell (2009). “Inverse modeling and mapping US air quality influences of inorganic PM2.5 precursor emissions using the adjoint of GEOS-Chem”. In: *Atmospheric Chemistry and Physics* 9.16, pp. 5877–5903. DOI: `10.5194/acp-9-5877-2009`.

Ho, K. E. et al. (2004). “Seasonal and diurnal variations of volatile organic compounds (VOCs) in the atmosphere of Hong Kong”. In: *Science of the Total Environment* 322.1-3, pp. 155–166. DOI: `10.1016/j.scitotenv.2003.10.004`.

Ho, K. F. et al. (2010). “Dicarboxylic acids, ketocarboxylic acids, alpha-dicarbonyls, fatty acids, and benzoic acid in urban aerosols collected during the 2006 Campaign of Air Quality Research in Beijing (CAREBeijing-2006)”. In: *Journal of Geophysical Research-Atmospheres* 115. DOI: `10.1029/2009jd013304`.

Hoekman, S. K. and C. Robbins (2012). “Review of the effects of biodiesel on NOx emissions”. In: *Fuel Processing Technology* 96, pp. 237–249. DOI: `10.1016/j.fuproc.2011.12.036`.

Hoesly, R. M. et al. (2018). “Historical (1750-2014) anthropogenic emissions of reactive gases and aerosols from the Community Emissions Data System (CEDS)”. In: *Geoscientific Model Development* 11.1, pp. 369–408. DOI: `10.5194/gmd-11-369-2018`.

Holmes, C. D. et al. (2019). “The Role of Clouds in the Tropospheeric NOx Cycle: A New Modeling Approach for Cloud Chemistry and Its Global Implications”. In: *Geophysical Research Letters* 46.9, pp. 4980–4990. DOI: `10.1029/2019gl081990`.

Hong-li, Wang et al. (2017). “Volatile organic compounds (VOCs) source profiles of on-road vehicle emissions in China”. In: *Science of the Total Environment* 607, pp. 253–261. DOI: `10.1016/j.scitotenv.2017.07.001`.

Hopkins, J. R., C. E. Jones, and A. C. Lewis (2011). “A dual channel gas chromatograph for atmospheric analysis of volatile organic compounds including oxygenated and monoterpene compounds”. In: *Journal of Environmental Monitoring* 13.8, pp. 2268–2276. DOI: `10.1039/c1em10050e`.

Horner, J. M. (2000). “Anthropogenic emissions of carbon monoxide”. In: *Reviews on environmental health* 15.3, pp. 289–98.

Hu, L., C. A. Keller, et al. (2018). “Global simulation of tropospheric chemistry at 12.5 km resolution: performance and evaluation of the GEOS-Chem chemical module (v10-1) within the NASA GEOS Earth system model (GEOS-5 ESM)”. In: *Geoscientific Model Development* 11.11, pp. 4603–4620. DOI: `10.5194/gmd-11-4603-2018`.

Hu, L., D. B. Millet, et al. (2015). “Emissions of C-6-C-8 aromatic compounds in the United States: Constraints from tall tower and aircraft measurements”. In: *Journal of Geophysical Research-Atmospheres* 120.2, pp. 826–842. DOI: `10.1002/2014jd022627`.

Huang, Z. H. et al. (2016). “Real-time monitoring of respiratory absorption factors of volatile organic compounds in ambient air by proton transfer reaction time-of-flight mass spectrometry”. In: *Journal of Hazardous Materials* 320, pp. 547–555. DOI: `10.1016/j.jhazmat.2016.08.064`.

Huberman, E. et al. (1976). “IDENTIFICATION OF MUTAGENIC METABOLITES OF BENZO A PYRENE IN MAMMALIAN-CELLS”. In: *Proceedings of the National Academy of Sciences of the United States of America* 73.2, pp. 607–611. DOI: `10.1073/pnas.73.2.607`.

Hudman, R. C. et al. (2012). “Steps towards a mechanistic model of global soil nitric oxide emissions: implementation and space based-constraints”. In: *Atmospheric Chemistry and Physics* 12.16, pp. 7779–7795. DOI: `10.5194/acp-12-7779-2012`.

Imhoff, R. E. et al. (1995). “THE PRODUCTION OF O-3 IN AN URBAN PLUME - AIRBORNE SAMPLING OF THE ATLANTA URBAN PLUME”. In: *Atmospheric Environment* 29.17, pp. 2349–2358. DOI: `10.1016/1352-2310(94)00305-5`.

Irving, P. M. (1983). “ACIDIC PRECIPITATION EFFECTS ON CROPS - A REVIEW AND ANALYSIS OF RESEARCH”. In: *Journal of Environmental Quality* 12.4, pp. 442–453. DOI: `10.2134/jeq1983.00472425001200040002x`.

Ivatt, P. D. and M. J. Evans (2020). “Improving the prediction of an atmospheric chemistry transport model using gradient-boosted regression trees”. In: *Atmospheric Chemistry and Physics* 20.13, pp. 8063–8082. DOI: `10.5194/acp-20-8063-2020`.

Jacob, D. J. (2000). “Heterogeneous chemistry and tropospheric ozone”. In: *Atmospheric Environment* 34.12-14, pp. 2131–2159. DOI: `10.1016/s1352-2310(99)00462-8`.

Jacob, D. J. and S. C. Wofsy (1988). “PHOTOCHEMISTRY OF BIOGENIC EMISSIONS OVER THE AMAZON FOREST”. In: *Journal of Geophysical Research-Atmospheres* 93.D2, pp. 1477–1486. DOI: `10.1029/JD093iD02p01477`.

Jaegle, L., D. J. Jacob, et al. (1998). “Sources and chemistry of NOx in the upper troposphere over the United States”. In: *Geophysical Research Letters* 25.10, pp. 1705–1708. DOI: `10.1029/97gl03591`.

Jaegle, L., P. K. Quinn, et al. (2011). “Global distribution of sea salt aerosols: new constraints from in situ and remote sensing observations”. In: *Atmospheric Chemistry and Physics* 11.7, pp. 3137–3157. DOI: `10.5194/acp-11-3137-2011`.

Jaffe, L. S. (1967). “BIOLOGICAL EFFECTS OF PHOTOCHEMICAL AIR POLLUTANTS ON MAN AND ANIMALS”. In: *American Journal of Public Health and the Nations Health* 57.8, pp. 1269–+. DOI: `10.2105/ajph.57.8.1269`.

Jaishankar, Monisha et al. (2014). “Toxicity, mechanism and health effects of some heavy metals”. In: *Interdisciplinary toxicology* 7.2, pp. 60–72. DOI: `10.2478/intox-2014-0009`.

Janechek, N. J., K. M. Hansen, and C. O. Stanier (2017). “Comprehensive atmospheric modeling of reactive cyclic siloxanes and their oxidation products”. In: *Atmospheric Chemistry and Physics* 17.13, pp. 8357–8370. DOI: `10.5194/acp-17-8357-2017`.

Jariyasopit, Narumol et al. (2014). “Heterogeneous Reactions of Particulate Matter-Bound PAHs and NPAHs with NO3/N2O5, OH Radicals, and O-3 under Simulated Long-Range Atmospheric Transport Conditions: Reactivity and Mutagenicity”. In: *Environmental Science and Technology* 48.17, pp. 10155–10164. DOI: `10.1021/es5015407`.

Jenkin, M. E., S. M. Saunders, and M. J. Pilling (1997). “The tropospheric degradation of volatile organic compounds: A protocol for mechanism development”. In: *Atmospheric Environment* 31.1, pp. 81–104. DOI: `10.1016/s1352-2310(96)00105-7`.

Jenkins, G. (2014). “A comparison between two types of widely used weather stations”. In: *Weather* 69.4, pp. 105–110. DOI: `10.1002/wea.2158`.

Jiang, Shuhui et al. (2019). “Deep Decision Tree Transfer Boosting”. In: *IEEE transactions on neural networks and learning systems*. DOI: `10.1109/tnnls.2019.2901273`.

Jing, B. Y. et al. (2016). “Development of a vehicle emission inventory with high temporal-spatial resolution based on NRT traffic data and its impact on air pollution in Beijing - Part 1: Development and evaluation of vehicle emission inventory”. In: *Atmospheric Chemistry and Physics* 16.5, pp. 3161–3170. DOI: `10.5194/acp-16-3161-2016`.

Kaiser, J. W. et al. (2012). “Biomass burning emissions estimated with a global fire assimilation system based on observed fire radiative power”. In: *Biogeosciences* 9.1, pp. 527–554. DOI: `10.5194/bg-9-527-2012`.

Karnosky, D. F. et al. (2002). “Interacting elevated CO2 and tropospheric O-3 predisposes aspen (Populus tremuloides Michx.) to infection by rust (Melampsora medusae f. sp tremuloidae)”. In: *Global Change Biology* 8.4, pp. 329–338. DOI: `10.1046/j.1354-1013.2002.00479.x`.

Karnosky, David F. et al. (2007). “Perspectives regarding 50 years of research on effects of tropospheric ozone air pollution on US forests”. In: *Environmental Pollution* 147.3, pp. 489–506. DOI: `10.1016/j.envpol.2006.08.043`.

Kasischke, E. S. and L. P. Bruhwiler (2002). “Emissions of carbon dioxide, carbon monoxide, and methane from boreal forest fires in 1998”. In: *Journal of Geophysical Research-Atmospheres* 108.D1. DOI: `10.1029/2001jd000461`.

Kavanaugh, M. (1987). “ESTIMATES OF FUTURE CO, N2O AND NOX EMISSIONS FROM ENERGY COMBUSTION”. In: *Atmospheric Environment* 21.3, pp. 463–468. DOI: `10.1016/0004-6981(87)90028-x`.

Kelishadi, Roya and Parinaz Poursafa (2010). “Air pollution and non-respiratory health hazards for children”. In: *Archives of Medical Science* 6.4, pp. 483–495. DOI: `10.5114/aoms.2010.14458`.

Keller, C. A. et al. (2014). “HEMCO v1.0: a versatile, ESMF-compliant component for calculating emissions in atmospheric models”. In: *Geoscientific Model Development* 7.4, pp. 1409–1417. DOI: `10.5194/gmd-7-1409-2014`.

Khalil, M. A. K. and R. A. Rasmussen (1990). “THE GLOBAL CYCLE OF CARBON-MONOXIDE - TRENDS AND MASS BALANCE”. In: *Chemosphere* 20.1-2, pp. 227–242. DOI: `10.1016/0045-6535(90)90098-e`.

Kim, Y. M., S. Harrad, and R. M. Harrison (2001). “Concentrations and sources of VOCs in urban domestic and public microenvironments”. In: *Environmental Science and Technology* 35.6, pp. 997–1004. DOI: `10.1021/es000192y`.

King, M. D., C. E. Canosa-Mas, and R. P. Wayne (2001). “Gas-phase reactions between RO2 and NO, HO2 or CH3O2: correlations between rate constants and the SOMO energy of the peroxy (RO2) radical”. In: *Atmospheric Environment* 35.12, pp. 2081–2088. DOI: `10.1016/s1352-2310(00)00501-x`.

Kingsford, C. and S. L. Salzberg (2008). “What are decision trees?” In: *Nature Biotechnology* 26.9, pp. 1011–1013. DOI: `10.1038/nbt0908-1011`.

Kirchner, F. and W. R. Stockwell (1996). “Effect of peroxy radical reactions on the predicted concentrations of ozone, nitrogenous compounds, and radicals”. In: *Journal of Geophysical Research-Atmospheres* 101.D15, pp. 21007–21022. DOI: `10.1029/96jd01519`.

Kleinermanns, K. and A. C. Luntz (1982). “THE CHEMICAL-DYNAMICS OF HYDROGEN-ATOM ABSTRACTION FROM UNSATURATED-HYDROCARBONS BY O(P-3)”. In: *Journal of Chemical Physics* 77.7, pp. 3533–3536. DOI: `10.1063/1.444253`.

Krol, M., P. J. van Leeuwen, and J. Lelieveld (1998). "Global OH trend inferred from methylchloroform measurements". In: *Journal of Geophysical Research-Atmospheres* 103.D9, pp. 10697–10711. DOI: `10.1029/98jd00459`.

Kroll, J. H. et al. (2002). "Gas-phase ozonolysis of alkenes: Formation of OH from anti carbonyl oxides". In: *Journal of the American Chemical Society* 124.29, pp. 8518–8519. DOI: `10.1021/ja0266060`.

Krupa, S. V., M. Nosal, and A. H. Legge (1998). "A numerical analysis of the combined open-top chamber data from the USA and Europe on ambient ozone and negative crop responses". In: *Environmental Pollution* 101.1, pp. 157–160. DOI: `10.1016/s0269-7491(98)00019-0`.

Kuhns, H., E. M. Knipping, and J. M. Vukovich (2005). "Development of a United States-Mexico emissions inventory for the Big Bend Regional Aerosol and Visibility Observational (BRAVO) Study". In: *Journal of the Air and Waste Management Association* 55.5, pp. 677–692. DOI: `10.1080/10473289.2005.10464648`.

Kulle, T. J. et al. (1985). "OZONE RESPONSE RELATIONSHIPS IN HEALTHY NON-SMOKERS". In: *American Review of Respiratory Disease* 132.1, pp. 36–41.

Kulmala, M. et al. (2013). "Direct Observations of Atmospheric Aerosol Nucleation". In: *Science* 339.6122, pp. 943–946. DOI: `10.1126/science.1227385`.

Kumar, V., B. P. Chandra, and V. Sinha (2018). "Large unexplained suite of chemically reactive compounds present in ambient air due to biomass fires". In: *Scientific Reports* 8. DOI: `10.1038/s41598-017-19139-3`.

Lakey, P. S. J. et al. (2015). "Measurements of the HO2 Uptake Coefficients onto Single Component Organic Aerosols". In: *Environmental Science and Technology* 49.8, pp. 4878–4885. DOI: `10.1021/acs.est.5b00948`.

Lang, J. L. et al. (2017). "Trends of PM2.5 and Chemical Composition in Beijing, 2000-2015". In: *Aerosol and Air Quality Research* 17.2, pp. 412–425. DOI: `10.4209/aaqr.2016.07.0307`.

Latimer, R. N. C. and R. V. Martin (2019). "Interpretation of measured aerosol mass scattering efficiency over North America using a chemical transport model". In: *Atmo-*

spheric Chemistry and Physics 19.4, pp. 2635–2653. DOI: `10.5194/acp-19-2635-2019`.

Lee, C. et al. (2011). "SO2 emissions and lifetimes: Estimates from inverse modeling using in situ and global, space-based (SCIAMACHY and OMI) observations". In: *Journal of Geophysical Research-Atmospheres* 116. DOI: `10.1029/2010jd014758`.

Li, J. et al. (2018). "Radiative and heterogeneous chemical effects of aerosols on ozone and inorganic aerosols over East Asia". In: *Science of the Total Environment* 622, pp. 1327–1342. DOI: `10.1016/j.scitotenv.2017.12.041`.

Li, K. et al. (2019). "Anthropogenic drivers of 2013-2017 trends in summer surface ozone in China". In: *Proceedings of the National Academy of Sciences of the United States of America* 116.2, pp. 422–427. DOI: `10.1073/pnas.1812168116`.

Li, M., H. Liu, et al. (2017). "Anthropogenic emission inventories in China: a review". In: *National Science Review* 4.6, pp. 834–866. DOI: `10.1093/nsr/nwx150`.

Li, M., Q. Zhang, et al. (2017). "MIX: a mosaic Asian anthropogenic emission inventory under the international collaboration framework of the MICS-Asia and HTAP". In: *Atmospheric Chemistry and Physics* 17.2, pp. 935–963. DOI: `10.5194/acp-17-935-2017`.

Li, Q. Q. et al. (2020). "An investigation into the role of VOCs in SOA and ozone production in Beijing, China". In: *Science of the Total Environment* 720. DOI: `10.1016/j.scitotenv.2020.137536`.

Li, Shenshen et al. (2019). "Comparisons of the vertical distributions of aerosols in the CALIPSO and GEOS-Chem datasets in China". In: *Atmospheric Environment-X* 3. DOI: `10.1016/j.aeaoa.2019.100036`.

Li, Yinghui et al. (2020). "Source apportionment of atmospheric volatile aromatic compounds (BTEX) by stable carbon isotope analysis: A case study during heating period in Taiyuan, northern China". In: *Atmospheric Environment* 225. DOI: `10.1016/j.atmosenv.2020.117369`.

Liang, Y. T. et al. (2017). “Revisiting nitrous acid (HONO) emission from on-road vehicles: A tunnel study with a mixed fleet”. In: *Journal of the Air and Waste Management Association* 67.7, pp. 797–805. DOI: `10.1080/10962247.2017.1293573`.

Lim, Stephen S. et al. (2012). “A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010”. In: *Lancet* 380.9859, pp. 2224–2260. DOI: `10.1016/s0140-6736(12)61766-8`.

Lin, H. P. et al. (2020). “WRF-GC (v1.0): online coupling of WRF (v3.9.1.1) and GEOS-Chem (v12.2.1) for regional atmospheric chemistry modeling - Part 1: Description of the one-way model”. In: *Geoscientific Model Development* 13.7, pp. 3241–3265. DOI: `10.5194/gmd-13-3241-2020`.

Lin, J. T. and M. B. McElroy (2010). “Impacts of boundary layer mixing on pollutant vertical profiles in the lower troposphere: Implications to satellite remote sensing”. In: *Atmospheric Environment* 44.14, pp. 1726–1739. DOI: `10.1016/j.atmosenv.2010.02.009`.

Lin, S. J. and R. B. Rood (1996). “Multidimensional flux-form semi-Lagrangian transport schemes”. In: *Monthly Weather Review* 124.9, pp. 2046–2070. DOI: `10.1175/1520-0493(1996)124<2046:mffslt>2.0.co;2`.

Lippmann, M. (1989). “HEALTH-EFFECTS OF OZONE - A CRITICAL-REVIEW”. In: *Japca-the Journal of the Air and Waste Management Association* 39.5, pp. 672–695. DOI: `10.1080/08940630.1989.10466554`.

Liu, D. et al. (2017). “Concentration, source identification, and exposure risk assessment of PM2.5-bound parent PAHs and nitro-PAHs in atmosphere from typical Chinese cities”. In: *Scientific Reports* 7. DOI: `10.1038/s41598-017-10623-4`.

Liu, D. T. et al. (2017). “Black-carbon absorption enhancement in the atmosphere determined by particle mixing state”. In: *Nature Geoscience* 10.3, 184–U132. DOI: `10.1038/ngeo2901`.

Liu, H. Y. et al. (2001). “Constraints from Pb-210 and Be-7 on wet deposition and transport in a global three-dimensional chemical tracer model driven by assimilated meteorological fields”. In: *Journal of Geophysical Research-Atmospheres* 106.D11, pp. 12109–12128. DOI: `10.1029/2000jd900839`.

Logan, W. P. D. (1953). “MORTALITY IN THE LONDON FOG INCIDENT, 1952”. In: *Lancet* 264.FEB14, pp. 336–338.

Londahl, Jakob et al. (2006). “A set-up for field studies of respiratory tract deposition of fine and ultrafine particles in humans”. In: *Journal of Aerosol Science* 37.9, pp. 1152–1163. DOI: `10.1016/j.jaerosci.2005.11.004`.

Long, M. S. et al. (2015). “Development of a grid-independent GEOS-Chem chemical transport model (v9-02) as an atmospheric chemistry module for Earth system models”. In: *Geoscientific Model Development* 8.3, pp. 595–602. DOI: `10.5194/gmd-8-595-2015`.

Lou, S. J., H. Liao, and B. Zhu (2014). “Impacts of aerosols on surface-layer ozone concentrations in China through heterogeneous reactions and changes in photolysis rates”. In: *Atmospheric Environment* 85, pp. 123–138. DOI: `10.1016/j.atmosenv.2013.12.004`.

Lu, Keding et al. (2010). “Oxidant (O-3 + NO2) production processes and formation regimes in Beijing”. In: *Journal of Geophysical Research-Atmospheres* 115. DOI: `10.1029/2009jd012714`.

Lu, X. et al. (2019). “Exploring 2016-2017 surface ozone pollution over China: source contributions and meteorological influences”. In: *Atmospheric Chemistry and Physics* 19.12, pp. 8339–8361. DOI: `10.5194/acp-19-8339-2019`.

Lu, X. C. et al. (2018). “Evidence of heterogeneous HONO formation from aerosols and the regional photochemical impact of this HONO source”. In: *Environmental Research Letters* 13.11. DOI: `10.1088/1748-9326/aae492`.

Mahajan, A. S. et al. (2010). “Measurement and modelling of tropospheric reactive halogen species over the tropical Atlantic Ocean”. In: *Atmospheric Chemistry and Physics* 10.10, pp. 4611–4624. DOI: `10.5194/acp-10-4611-2010`.

Malley, C. S. et al. (2017). “Updated Global Estimates of Respiratory Mortality in Adults >= 30 Years of Age Attributable to Long-Term Ozone Exposure”. In: *Environmental Health Perspectives* 125.8. doi: `10.1289/ehp1390`.

Manisalidis, I. et al. (2020). “Environmental and Health Impacts of Air Pollution: A Review”. In: *Frontiers in Public Health* 8. doi: `10.3389/fpubh.2020.00014`.

Mao, J., S. Fan, et al. (2013). “Radical loss in the atmosphere from Cu-Fe redox coupling in aerosols”. In: *Atmospheric Chemistry and Physics* 13.2, pp. 509–519. doi: `10.5194/acp-13-509-2013`.

Mao, J., D. J. Jacob, et al. (2010). “Chemistry of hydrogen oxide radicals (HOx) in the Arctic troposphere in spring”. In: *Atmospheric Chemistry and Physics* 10.13, pp. 5823–5838. doi: `10.5194/acp-10-5823-2010`.

Martin, A. (1984). “ESTIMATED WASHOUT COEFFICIENTS FOR SULFUR-DIOXIDE, NITRIC-OXIDE, NITROGEN-DIOXIDE AND OZONE”. In: *Atmospheric Environment* 18.9, pp. 1955–1961. doi: `10.1016/0004-6981(84)90373-1`.

Martin, M. Val et al. (2008). “Seasonal variation of nitrogen oxides in the central North Atlantic lower free troposphere”. In: *Journal of Geophysical Research-Atmospheres* 113.D17. doi: `10.1029/2007jd009688`.

Marvin, Margaret R. et al. (2017). “Impact of evolving isoprene mechanisms on simulated formaldehyde: An inter-comparison supported by in situ observations from SENEX”. In: *Atmospheric Environment* 164, pp. 325–336. doi: `10.1016/j.atmosenv.2017.05.049`.

Maxwell, A. E., T. A. Warner, and F. Fang (2018). “Implementation of machine-learning classification in remote sensing: an applied review”. In: *International Journal of Remote Sensing* 39.9, pp. 2784–2817. doi: `10.1080/01431161.2018.1433343`.

McDuffie, E. E. et al. (2018). “Heterogeneous N2O5 Uptake During Winter: Aircraft Measurements During the 2015 WINTER Campaign and Critical Evaluation of Current Parameterizations”. In: *Journal of Geophysical Research-Atmospheres* 123.8, pp. 4345–4372. doi: `10.1002/2018jd028336`.

McLinden, C. A. et al. (2000). “Stratospheric ozone in 3-D models: A simple chemistry and the cross-tropopause flux”. In: *Journal of Geophysical Research-Atmospheres* 105.D11, pp. 14653–14665. DOI: `10.1029/2000jd900124`.

Meisner, Jan and Johannes Kaestner (2016). “Reaction rates and kinetic isotope effects of H-2 + OH -> H2O + H”. In: *Journal of Chemical Physics* 144.17. DOI: `10.1063/1.4948319`.

Mertens, Mariano et al. (2020). “Are contributions of emissions to ozone a matter of scale? - a study using MECO(n) (MESSy v2.50)”. In: *Geoscientific Model Development* 13.1, pp. 363–383. DOI: `10.5194/gmd-13-363-2020`.

Meusel, H. et al. (2018). “Emission of nitrous acid from soil and biological soil crusts represents an important source of HONO in the remote atmosphere in Cyprus”. In: *Atmospheric Chemistry and Physics* 18.2, pp. 799–813. DOI: `10.5194/acp-18-799-2018`.

Michelsen, H. A. et al. (1994). “PRODUCTION OF O(D-1) FROM PHOTOLYSIS OF O-3”. In: *Geophysical Research Letters* 21.20, pp. 2227–2230. DOI: `10.1029/94gl02052`.

Millet, D. B. et al. (2015). “A large and ubiquitous source of atmospheric formic acid”. In: *Atmospheric Chemistry and Physics* 15.11, pp. 6283–6304. DOI: `10.5194/acp-15-6283-2015`.

Millet, Dylan B. et al. (2008). “Spatial distribution of isoprene emissions from North America derived from formaldehyde column measurements by the OMI satellite sensor”. In: *Journal of Geophysical Research-Atmospheres* 113.D2. DOI: `10.1029/2007jd008950`.

Mo, Z. W. et al. (2018). “Species-specified VOC emissions derived from a gridded study in the Pearl River Delta, China”. In: *Scientific Reports* 8. DOI: `10.1038/s41598-018-21296-y`.

Mo, Ziwei et al. (2018). “Evaluation of biogenic isoprene emissions and their contribution to ozone formation by ground-based measurements in Beijing, China”. In: *Science of the Total Environment* 627, pp. 1485–1494. DOI: `10.1016/j.scitotenv.2018.01.336`.

Molod, A. et al. (2015). “Development of the GEOS-5 atmospheric general circulation model: evolution from MERRA to MERRA2”. In: *Geoscientific Model Development* 8.5, pp. 1339–1356. DOI: `10.5194/gmd-8-1339-2015`.

Monks, P. S. (2005). “Gas-phase radical chemistry in the troposphere”. In: *Chemical Society Reviews* 34.5, pp. 376–395. DOI: `10.1039/b307982c`.

Montzka, Stephen A. et al. (2018). “An unexpected and persistent increase in global emissions of ozone-depleting CFC-11”. In: *Nature* 557.7705, pp. 413–+. DOI: `10.1038/s41586-018-0106-2`.

Moortgat, G. K. and P. Warneck (1979). “CO AND H-2 QUANTUM YIELDS IN THE PHOTO-DECOMPOSITION OF FORMALDEHYDE IN AIR”. In: *Journal of Chemical Physics* 70.8, pp. 3639–3651. DOI: `10.1063/1.437956`.

Moxim, W. J., H. Levy, and P. S. Kasibhatla (1996). “Simulated global tropospheric PAN: Its transport and impact on NOx”. In: *Journal of Geophysical Research-Atmospheres* 101.D7, pp. 12621–12638. DOI: `10.1029/96jd00338`.

Murray, L. T., D. J. Jacob, et al. (2012). “Optimized regional and interannual variability of lightning in a global chemical transport model constrained by LIS/OTD satellite data”. In: *Journal of Geophysical Research-Atmospheres* 117. DOI: `10.1029/2012jd017934`.

Murray, L. T., J. A. Logan, and D. J. Jacob (2013). “Interannual variability in tropical tropospheric ozone and OH: The role of lightning”. In: *Journal of Geophysical Research-Atmospheres* 118.19, pp. 11468–11480. DOI: `10.1002/jgrd.50857`.

Nedelec, P. et al. (2015). “Instrumentation on commercial aircraft for monitoring the atmospheric composition on a global scale: the IAGOS system, technical overview of ozone and carbon monoxide measurements”. In: *Tellus Series B-Chemical and Physical Meteorology* 67. DOI: `10.3402/tellusb.v67.27791`.

Newsome, B. and M. Evans (2017). “Impact of uncertainties in inorganic chemical rate constants on tropospheric composition and ozone radiative forcing”. In: *Atmospheric Chemistry and Physics* 17.23, pp. 14333–14352. DOI: `10.5194/acp-17-14333-2017`.

Noxon, J. F. (1976). “ATMOSPHERIC NITROGEN-FIXATION BY LIGHTNING”. In: *Geophysical Research Letters* 3.8, pp. 463–465. doi: `10.1029/GL003i008p00463`.

Owaida, Muhsen et al. (2017). “Scalable Inference of Decision Tree Ensembles: Flexible Design for CPU-FPGA Platforms”. In: *27th International Conference on Field Programmable Logic and Applications (FPL)*. International Conference on Field Programmable and Logic Applications.

Pai, S. J. et al. (2020). “An evaluation of global organic aerosol schemes using airborne observations”. In: *Atmospheric Chemistry and Physics* 20.5, pp. 2637–2665. doi: `10.5194/acp-20-2637-2020`.

Pan, L. L. et al. (2004). “Definitions and sharpness of the extratropical tropopause: A trace gas perspective”. In: *Journal of Geophysical Research-Atmospheres* 109.D23. doi: `10.1029/2004jd004982`.

Park, R. J. et al. (2004). “Natural and transboundary pollution influences on sulfate-nitrate-ammonium aerosols in the United States: Implications for policy”. In: *Journal of Geophysical Research-Atmospheres* 109.D15. doi: `10.1029/2003jd004473`.

Parungo, F., C. Nagamoto, and R. Madel (1987). “A STUDY OF THE MECHANISMS OF ACID-RAIN FORMATION”. In: *Journal of the Atmospheric Sciences* 44.21, pp. 3162–3174. doi: `10.1175/1520-0469(1987)044<3162:asotmo>2.0.co;2`.

Pell, E. J. et al. (1999). “Differential response of four tree species to ozone-induced acceleration of foliar senescence”. In: *Plant Cell and Environment* 22.7, pp. 779–790. doi: `10.1046/j.1365-3040.1999.00449.x`.

Perez, Laura et al. (2012). “Saharan dust, particulate matter and cause-specific mortality: A case-crossover study in Barcelona (Spain)”. In: *Environment International* 48, pp. 150–155. doi: `10.1016/j.envint.2012.07.001`.

Pilidis, Georgios A. et al. (2009). “Measurements of benzene and formaldehyde in a medium sized urban environment. Indoor/outdoor health risk implications on special population groups”. In: *Environmental Monitoring and Assessment* 150.1-4, pp. 285–294. doi: `10.1007/s10661-008-0230-9`.

Platt, U. et al. (1990). “PEROXY-RADICALS FROM NIGHTTIME REACTION OF NO3 WITH ORGANIC-COMPOUNDS”. In: *Nature* 348.6297, pp. 147–149. DOI: `10.1038/348147a0`.

Popoola, L. T., S. A. Adebanjo, and B. K. Adeoye (2018). “Assessment of atmospheric particulate matter and heavy metals: a critical review”. In: *International Journal of Environmental Science and Technology* 15.5, pp. 935–948. DOI: `10.1007/s13762-017-1454-4`.

Porter, W. C., S. A. Safieddine, and C. L. Heald (2017). “Impact of aromatics and monoterpenes on simulated tropospheric ozone and total OH reactivity”. In: *Atmospheric Environment* 169, pp. 250–257. DOI: `10.1016/j.atmosenv.2017.08.048`.

Pound, R. J. et al. (2019). “Influences of oceanic ozone deposition on tropospheric photochemistry”. In: *Atmospheric Chemistry and Physics Discussions* 2019, pp. 1–25. DOI: `10.5194/acp-2019-1043`.

Pratapas, J. M. and J. Calcagni (1983). “OZONE CONTROL STRATEGIES IN THE UNITED-STATES”. In: *Environment International* 9.6, pp. 529–538. DOI: `10.1016/0160-4120(83)90009-0`.

Pye, H. O. T. et al. (2010). “Global modeling of organic aerosol: the importance of reactive nitrogen (NOx and NO3)”. In: *Atmospheric Chemistry and Physics* 10.22, pp. 11261–11276. DOI: `10.5194/acp-10-11261-2010`.

Ravishankara, A. R. (1988). “KINETICS OF RADICAL REACTIONS IN THE ATMOSPHERIC OXIDATION OF CH4”. In: *Annual Review of Physical Chemistry* 39, pp. 367–394. DOI: `10.1146/annurev.physchem.39.1.367`.

Ravishankara, A. R. and R. L. Thompson (1983). “KINETIC-STUDY OF THE REACTION OF OH WITH CO FROM 250-K TO 1040-K”. In: *Chemical Physics Letters* 99.5-6, pp. 377–381. DOI: `10.1016/0009-2614(83)80158-4`.

Reed, Chris et al. (2016). “Interferences in photolytic NO2 measurements: explanation for an apparent missing oxidant?” In: *Atmospheric Chemistry and Physics* 16.7, pp. 4707–4724. DOI: `10.5194/acp-16-4707-2016`.

Reich, P. B. and J. P. Lassoie (1985). “INFLUENCE OF LOW CONCENTRATIONS OF OZONE ON GROWTH, BIOMASS PARTITIONING AND LEAF SENESCENCE IN YOUNG HYBRID POPLAR PLANTS”. In: *Environmental Pollution Series a-Ecological and Biological* 39.1, pp. 39–51. DOI: `10.1016/0143-1471(85)90061-3`.

Rhew, R. C. et al. (2017). “Ethene, propene, butene and isoprene emissions from a ponderosa pine forest measured by relaxed eddy accumulation”. In: *Atmospheric Chemistry and Physics* 17.21, pp. 13417–13438. DOI: `10.5194/acp-17-13417-2017`.

Richmond-Bryant, Jennifer et al. (2017). “Estimation of on-road NO2 concentrations, NO2/NOX ratios, and related roadway gradients from near-road monitoring data”. In: *Air Quality Atmosphere and Health* 10.5, pp. 611–625. DOI: `10.1007/s11869-016-0455-7`.

Ridley, D. A., C. L. Heald, and B. Ford (2012). “North African dust export and deposition: A satellite and model perspective”. In: *Journal of Geophysical Research-Atmospheres* 117. DOI: `10.1029/2011jd016794`.

Riemer, N. et al. (2003). “Impact of the heterogeneous hydrolysis of N2O5 on chemistry and nitrate aerosol formation in the lower troposphere under photosmog conditions”. In: *Journal of Geophysical Research-Atmospheres* 108.D4. DOI: `10.1029/2002jd002436`.

Ripperton, L. A. and F. M. Vukovich (1971). “GAS PHASE DESTRUCTION OF TROPOSPHERIC OZONE”. In: *Journal of Geophysical Research* 76.30, pp. 7328–+. DOI: `10.1029/JC076i030p07328`.

Robinson, E. and R. C. Robbins (1970). “GASEOUS NITROGEN COMPOUND POLLUTANTS FROM URBAN AND NATURAL SOURCES”. In: *Journal of the Air Pollution Control Association* 20.5, pp. 303–306.

Rodhe, H., F. Dentener, and M. Schulz (2002). “The global distribution of acidifying wet deposition”. In: *Environmental Science and Technology* 36.20, pp. 4382–4388. DOI: `10.1021/es020057g`.

Russell, W. T. (1924). “The influence of fog on mortality from respiratory diseases”. In: *Lancet* 2, pp. 335–339.

Rypdal, K. and W. Winiwarter (2001). “Uncertainties in greenhouse gas emission inventories — evaluation, comparability and implications”. In: *Environmental Science and Policy* 4.2-3, pp. 107–116. DOI: `https://doi.org/10.1016/S1462-9011(00)00113-1.`.

Samet, J. M., M. C. Marbury, and J. D. Spengler (1987). “HEALTH-EFFECTS AND SOURCES OF INDOOR AIR-POLLUTION .1”. In: *American Review of Respiratory Disease* 136.6, pp. 1486–1508. DOI: `10.1164/ajrccm/136.6.1486`.

Samoli, Evangelia et al. (2008). “Acute Effects of Ambient Particulate Matter on Mortality in Europe and North America: Results from the APHENA Study”. In: *Environmental Health Perspectives* 116.11, pp. 1480–1486. DOI: `10.1289/ehp.11345`.

Sander, R. (2015). “Compilation of Henry’s law constants (version 4.0) for water as solvent”. In: *Atmospheric Chemistry and Physics* 15.8, pp. 4399–4981. DOI: `10.5194/acp-15-4399-2015`.

Saunders, S. M. et al. (2003). “Protocol for the development of the Master Chemical Mechanism, MCM v3 (Part A): tropospheric degradation of non-aromatic volatile organic compounds”. In: *Atmospheric Chemistry and Physics* 3, pp. 161–180. DOI: `10.5194/acp-3-161-2003`.

Saunois, Marielle et al. (2016). “The global methane budget 2000-2012”. In: *Earth System Science Data* 8.2, pp. 697–751. DOI: `10.5194/essd-8-697-2016`.

Schmidt, J. A. et al. (2016). “Modeling the observed tropospheric BrO background: Importance of multiphase chemistry and implications for ozone, OH, and mercury”. In: *Journal of Geophysical Research-Atmospheres* 121.19, pp. 11819–11835. DOI: `10.1002/2015jd024229`.

Schneidemesser, Erika von et al. (2015). “Chemistry and the Linkages between Air Quality and Climate Change”. In: *Chemical Reviews* 115.10, pp. 3856–3897. DOI: `10.1021/acs.chemrev.5b00089`.

Schuetzle, D. et al. (1994). “THE RELATIONSHIP BETWEEN GASOLINE COMPOSITION AND VEHICLE HYDROCARBON EMISSIONS - A REVIEW OF CUR-

RENT STUDIES AND FUTURE-RESEARCH NEEDS". In: *Environmental Health Perspectives* 102, pp. 3–12. DOI: 10.2307/3431925.

Schuh, Andrew E. et al. (2019). "Quantifying the Impact of Atmospheric Transport Uncertainty on CO2 Surface Flux Estimates". In: *Global Biogeochemical Cycles* 33.4, pp. 484–500. DOI: 10.1029/2018gb006086.

Schumann, U. and H. Huntrieser (2007). "The global lightning-induced nitrogen oxides source". In: *Atmospheric Chemistry and Physics* 7.14, pp. 3823–3907. DOI: 10.5194/acp-7-3823-2007.

Schwantes, Rebecca H. et al. (2020). "Comprehensive isoprene and terpene gas-phase chemistry improves simulated surface ozone in the southeastern US". In: *Atmospheric Chemistry and Physics* 20.6, pp. 3739–3776. DOI: 10.5194/acp-20-3739-2020.

Schwartz, S. E. (1984). "GAS-PHASE AND AQUEOUS-PHASE CHEMISTRY OF HO2 IN LIQUID WATER CLOUDS". In: *Journal of Geophysical Research-Atmospheres* 89.ND7, pp. 1589–1598. DOI: 10.1029/JD089iD07p11589.

Seinfeld, J. H. (1989). "URBAN AIR-POLLUTION - STATE OF THE SCIENCE". In: *Science* 243.4892, pp. 745–752. DOI: 10.1126/science.243.4892.745.

Selin, N. E. and D. J. Jacob (2008). "Seasonal and spatial patterns of mercury wet deposition in the United States: Constraints on the contribution from North American anthropogenic sources". In: *Atmospheric Environment* 42.21, pp. 5193–5204. DOI: 10.1016/j.atmosenv.2008.02.069.

Shallcross, D. E. et al. (2005). "A semi-empirical correlation for the rate coefficients for cross- and self-reactions of peroxy radicals in the gas-phase". In: *Atmospheric Environment* 39.4, pp. 763–771. DOI: 10.1016/j.atmosenv.2004.09.072.

Sherwen, T. et al. (2016). "Iodine's impact on tropospheric oxidants: a global model study in GEOS-Chem". In: *Atmospheric Chemistry and Physics* 16.2, pp. 1161–1186. DOI: 10.5194/acp-16-1161-2016.

Sherwen, Tomas et al. (2016). "Global impacts of tropospheric halogens (Cl, Br, I) on oxidants and composition in GEOS-Chem". In: *Atmospheric Chemistry and Physics* 16.18, pp. 12239–12271. DOI: 10.5194/acp-16-12239-2016.

Shetter, R. E. et al. (2003). “Photolysis frequency of NO2: Measurement and modeling during the International Photolysis Frequency Measurement and Modeling Intercomparison (IPMMI)”. In: *Journal of Geophysical Research-Atmospheres* 108.D16. DOI: `10.1029/2002jd002932`.

Shi, Z. B. et al. (2019). “Introduction to the special issue "In-depth study of air pollution sources and processes within Beijing and its surrounding region (APHH-Beijing)"”. In: *Atmospheric Chemistry and Physics* 19.11, pp. 7519–7546. DOI: `10.5194/acp-19-7519-2019`.

Shiraiwa, M. et al. (2017). “Aerosol Health Effects from Molecular to Global Scales”. In: *Environmental Science and Technology* 51.23, pp. 13545–13567. DOI: `10.1021/acs.est.7b04417`.

Sillman, S. (1999). “The relation between ozone, NOx and hydrocarbons in urban and polluted rural environments”. In: *Atmospheric Environment* 33.12, pp. 1821–1845. DOI: `10.1016/s1352-2310(98)00345-8`.

Sillman, S. and D. Y. He (2002). “Some theoretical results concerning O-3-NOx-VOC chemistry and NOx-VOC indicators”. In: *Journal of Geophysical Research-Atmospheres* 107.D22. DOI: `10.1029/2001jd001123`.

Sillman, S., J. A. Logan, and S. C. Wofsy (1990). “THE SENSITIVITY OF OZONE TO NITROGEN-OXIDES AND HYDROCARBONS IN REGIONAL OZONE EPISODES”. In: *Journal of Geophysical Research-Atmospheres* 95.D2, pp. 1837–1851. DOI: `10.1029/JD095iD02p01837`.

Silva, G. da (2010). “Hydroxyl radical regeneration in the photochemical oxidation of glyoxal: kinetics and mechanism of the HC(O)CO + O-2 reaction”. In: *Physical Chemistry Chemical Physics* 12.25, pp. 6698–6705. DOI: `10.1039/b927176g`.

Silva, Raquel A. et al. (2013). “Global premature mortality due to anthropogenic outdoor air pollution and the contribution of past climate change”. In: *Environmental Research Letters* 8.3. DOI: `10.1088/1748-9326/8/3/034005`.

Simpson, D. et al. (1995). “BIOGENIC EMISSIONS IN EUROPE .1. ESTIMATES AND UNCERTAINTIES”. In: *Journal of Geophysical Research-Atmospheres* 100.D11, pp. 22875–22890. DOI: `10.1029/95jd02368`.

Singh, Anita and Madhoolika Agrawal (2008). “Acid rain and its ecological consequences”. In: *Journal of Environmental Biology* 29.1, pp. 15–24.

Singh, H. B. et al. (2010). “Pollution influences on atmospheric composition and chemistry at high northern latitudes: Boreal and California forest fire emissions”. In: *Atmospheric Environment* 44.36, pp. 4553–4564. DOI: `10.1016/j.atmosenv.2010.08.026`.

Situ, S. et al. (2014). “Uncertainties of isoprene emissions in the MEGAN model estimated for a coniferous and broad-leaved mixed forest in Southern China”. In: *Atmospheric Environment* 98, pp. 105–110. DOI: `10.1016/j.atmosenv.2014.08.023`.

Sommariva, R. et al. (2020). “AtChem (version 1), an open-source box model for the Master Chemical Mechanism”. In: *Geoscientific Model Development* 13.1, pp. 169–183. DOI: `10.5194/gmd-13-169-2020`.

Sonnemann, G. R. and M. Grygalashvyly (2014). “Global annual methane emission rate derived from its current atmospheric mixing ratio and estimated lifetime”. In: *Annales Geophysicae* 32.3, pp. 277–283. DOI: `10.5194/angeo-32-277-2014`.

Spengler, J. D. and K. Sexton (1983). “INDOOR AIR-POLLUTION - A PUBLIC-HEALTH PERSPECTIVE”. In: *Science* 221.4605, pp. 9–17. DOI: `10.1126/science.6857273`.

Squires, F. A. et al. (2020). “Measurements of traffic-dominated pollutant emissions in a Chinese megacity”. In: *Atmospheric Chemistry and Physics* 20.14, pp. 8737–8761. DOI: `10.5194/acp-20-8737-2020`.

Stark, R. W. et al. (1968). “INCIDENCE OF BARK BEETLE INFESTATION IN INJURED TREES”. In: *Hilgardia* 39.6, 121–and. DOI: `10.3733/hilg.v39n06p121`.

Stavrakou, T. et al. (2009). “Evaluating the performance of pyrogenic and biogenic emission inventories against one decade of space-based formaldehyde columns”. In: *Atmospheric Chemistry and Physics* 9.3, pp. 1037–1060. DOI: `10.5194/acp-9-1037-2009`.

Sterling, Chance W. et al. (2018). “Homogenizing and estimating the uncertainty in NOAA’s long-term vertical ozone profile records measured with the electrochemical concentration cell ozonesonde”. In: *Atmospheric Measurement Techniques* 11.6, pp. 3661–3687. DOI: `10.5194/amt-11-3661-2018`.

Stockwell, D. Z. et al. (1999). “Modelling NOx from lightning and its impact on global chemical fields”. In: *Atmospheric Environment* 33.27, pp. 4477–4493. DOI: `10.1016/s1352-2310(99)00190-9`.

Sumner, A. L. and P. B. Shepson (1999). “Snowpack production of formaldehyde and its effect on the Arctic troposphere”. In: *Nature* 398.6724, pp. 230–233. DOI: `10.1038/18423`.

Sun, Y. L. et al. (2016). “Primary and secondary aerosols in Beijing in winter: sources, variations and processes”. In: *Atmospheric Chemistry and Physics* 16.13, pp. 8309–8329. DOI: `10.5194/acp-16-8309-2016`.

Tai, Amos P. K. and Maria Val Martin (2017). “Impacts of ozone air pollution and temperature extremes on crop yields: Spatial variability, adaptation and implications for future food security”. In: *Atmospheric Environment* 169, pp. 11–21. DOI: `10.1016/j.atmosenv.2017.09.002`.

Taketani, F., Y. Kanaya, and H. Akimoto (2008). “Kinetics of heterogeneous reactions of HO2 radical at ambient concentration levels with (NH4)(2)SO4 and NaCl aerosol particles”. In: *Journal of Physical Chemistry A* 112.11, pp. 2370–2377. DOI: `10.1021/jp0769936`.

Taketani, F., Y. Kanaya, P. Pochanart, et al. (2012). “Measurement of overall uptake coefficients for HO2 radicals by aerosol particles sampled from ambient air at Mts. Tai and Mang (China)”. In: *Atmospheric Chemistry and Physics* 12.24, pp. 11907–11916. DOI: `10.5194/acp-12-11907-2012`.

Tan, Z. F., A. Hofzumahaus, et al. (2020). “No Evidence for a Significant Impact of Heterogeneous Chemistry on Radical Concentrations in the North China Plain in Summer 2014”. In: *Environmental Science and Technology* 54.10, pp. 5973–5979. DOI: `10.1021/acs.est.0c00525`.

Tan, Z. F., F. Rohrer, et al. (2018). “Wintertime photochemistry in Beijing: observations of ROx radical concentrations in the North China Plain during the BEST-ONE campaign”. In: *Atmospheric Chemistry and Physics* 18.16, pp. 12391–12411. DOI: `10.5194/acp-18-12391-2018`.

Tan, Zhaofeng et al. (2019). “Experimental budgets of OH, HO2, and RO2 radicals and implications for ozone formation in the Pearl River Delta in China 2014”. In: *Atmospheric Chemistry and Physics* 19.10, pp. 7129–7150. DOI: `10.5194/acp-19-7129-2019`.

Taylor, O. C. (1969). “IMPORTANCE OF PEROXY ACETYL NITRATE AS A PHYTO TOXIC AIR POLLUTANT”. In: *Journal of the Air Pollution Control Association* 19.5, pp. 347–351.

Thielmann, A. et al. (2001). “Empirical ozone isopleths as a tool to identify ozone production regimes”. In: *Geophysical Research Letters* 28.12, pp. 2369–2372. DOI: `10.1029/2000gl012787`.

Thornton, J. and J. P. D. Abbatt (2005). “Measurements of HO2 uptake to aqueous aerosol: Mass accommodation coefficients and net reactive loss”. In: *Journal of Geophysical Research-Atmospheres* 110.D8. DOI: `10.1029/2004jd005402`.

Tian, R. et al. (2019). “Aerosol radiative effects on tropospheric photochemistry with GEOS-Chem simulations”. In: *Atmospheric Environment* 208, pp. 82–94. DOI: `10.1016/j.atmosenv.2019.03.032`.

Torlay, L. et al. (2017). “Machine learning-XGBoost analysis of language networks to classify patients with epilepsy”. In: *Brain informatics* 4.3, pp. 159–169. DOI: `10.1007/s40708-017-0065-7`.

Travis, K. R. et al. (2016). “Why do models overestimate surface ozone in the Southeast United States?” In: *Atmospheric Chemistry and Physics* 16.21, pp. 13561–13577. DOI: `10.5194/acp-16-13561-2016`.

Tzompa-Sosa, Z. A. et al. (2017). “Revisiting global fossil fuel and biofuel emissions of ethane”. In: *Journal of Geophysical Research-Atmospheres* 122.4, pp. 2493–2512. DOI: `10.1002/2016jd025767`.

Vaghjiani, G. L. and A. R. Ravishankara (1991). "NEW MEASUREMENT OF THE RATE COEFFICIENT FOR THE REACTION OF OH WITH METHANE". In: *Nature* 350.6317, pp. 406–409. DOI: `10.1038/350406a0`.

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

Viard, W. et al. (2004). "Integrated assessment of heavy metal (Pb, Zn, Cd) highway pollution: bioaccumulation in soil, Graminaceae and land snails". In: *Chemosphere* 55.10, pp. 1349–1359. DOI: `10.1016/j.chemosphere.2004.01.003`.

Waldron, H. A. (1983). "A BRIEF-HISTORY OF SCROTAL CANCER". In: *British Journal of Industrial Medicine* 40.4, pp. 390–401.

Wang, B. et al. (2010). "Variation of ambient non-methane hydrocarbons in Beijing city in summer 2008". In: *Atmospheric Chemistry and Physics* 10.13, pp. 5911–5923. DOI: `10.5194/acp-10-5911-2010`.

Wang, M. et al. (2014). "A temporally and spatially resolved validation of emission inventories by measurements of ambient volatile organic compounds in Beijing, China". In: *Atmospheric Chemistry and Physics* 14.12, pp. 5871–5891. DOI: `10.5194/acp-14-5871-2014`.

Wang, Peng et al. (2019). "Source apportionment of summertime ozone in China using a source-oriented chemical transport model". In: *Atmospheric Environment* 211, pp. 79–90. DOI: `10.1016/j.atmosenv.2019.05.006`.

Wang, Q. et al. (2011). "Sources of carbonaceous aerosols and deposited black carbon in the Arctic in winter-spring: implications for radiative forcing". In: *Atmospheric Chemistry and Physics* 11.23, pp. 12453–12473. DOI: `10.5194/acp-11-12453-2011`.

Wang, Q. Q. et al. (2014). "Global budget and radiative forcing of black carbon aerosol: Constraints from pole-to-pole (HIPPO) observations across the Pacific". In: *Journal of Geophysical Research-Atmospheres* 119.1, pp. 195–206. DOI: `10.1002/2013jd020824`.

Wang, S. X. and J. M. Hao (2012). “Air quality management in China: Issues, challenges, and options”. In: *Journal of Environmental Sciences-China* 24.1, pp. 2–13. DOI: `10.1016/s1001-0742(11)60724-9`.

Wang, T. et al. (2017). “Ozone pollution in China: A review of concentrations, meteorological influences, chemical precursors, and effects”. In: *Science of the Total Environment* 575, pp. 1582–1596. DOI: `10.1016/j.scitotenv.2016.10.081`.

Wang, X. et al. (2019). “The role of chlorine in global tropospheric chemistry”. In: *Atmospheric Chemistry and Physics* 19.6, pp. 3981–4003. DOI: `10.5194/acp-19-3981-2019`.

Wang, X. M. et al. (2002). “Urban roadside aromatic hydrocarbons in three cities of the Pearl River Delta, People’s Republic of China”. In: *Atmospheric Environment* 36.33, pp. 5141–5148. DOI: `10.1016/s1352-2310(02)00640-4`.

Wang, Y. H., D. J. Jacob, and J. A. Logan (1998). “Global simulation of tropospheric O-3-NOx-hydrocarbon chemistry 1. Model formulation”. In: *Journal of Geophysical Research-Atmospheres* 103.D9, pp. 10713–10725. DOI: `10.1029/98jd00158`.

Wang, Y. H., R. Zheng, et al. (2016). “The impact of fuel compositions on the particulate emissions of direct injection gasoline engine”. In: *Fuel* 166, pp. 543–552. DOI: `10.1016/j.fuel.2015.11.019`.

Watson, J. G., J. C. Chow, and E. M. Fujita (2001). “Review of volatile organic compound source apportionment by chemical mass balance”. In: *Atmospheric Environment* 35.9, pp. 1567–1584. DOI: `10.1016/s1352-2310(00)00461-1`.

Wei, H. H. (2019). “Impacts of China’s national vehicle fuel standards and subway development on air pollution”. In: *Journal of Cleaner Production* 241. DOI: `10.1016/j.jclepro.2019.118399`.

Wei, S. J. C. et al. (1991). “DOSE-DEPENDENT DIFFERENCES IN THE PROFILE OF MUTATIONS INDUCED BY AN ULTIMATE CARCINOGEN FROM BENZO A PYRENE”. In: *Proceedings of the National Academy of Sciences of the United States of America* 88.24, pp. 11227–11230. DOI: `10.1073/pnas.88.24.11227`.

Whalley, L. K. et al. (2010). “The chemistry of OH and HO2 radicals in the boundary layer over the tropical Atlantic Ocean”. In: *Atmospheric Chemistry and Physics* 10.4, pp. 1555–1576. DOI: `10.5194/acp-10-1555-2010`.

Whalley, Lisa K. et al. (2016). “Atmospheric OH reactivity in central London: observations, model predictions and estimates of in situ ozone production”. In: *Atmospheric Chemistry and Physics* 16.4, pp. 2109–2122. DOI: `10.5194/acp-16-2109-2016`.

Wiedinmyer, C. et al. (2011). “The Fire INventory from NCAR (FINN): a high resolution global model to estimate the emissions from open burning”. In: *Geoscientific Model Development* 4.3, pp. 625–641. DOI: `10.5194/gmd-4-625-2011`.

Winer, A. M., R. Atkinson, and J. N. Pitts (1984). “GASEOUS NITRATE RADICAL - POSSIBLE NIGHTTIME ATMOSPHERIC SINK FOR BIOGENIC ORGANIC-COMPOUNDS”. In: *Science* 224.4645, pp. 156–159. DOI: `10.1126/science.224.4645.156`.

Wittrock, Folkard et al. (2006). “Simultaneous global observations of glyoxal and formaldehyde from space”. In: *Geophysical Research Letters* 33.16. DOI: `10.1029/2006gl026310`.

Wofsy, S.C. et al. (2018). “ATom: Merged Atmospheric Chemistry, Trace Gases, and Aerosols”. In: DOI: `10.3334/ORNLDAAC/1581`.

Wu, X. Q. et al. (2007). “Impact of a modified convective scheme on the Madden-Julian Oscillation and El Nino-Southern Oscillation in a coupled climate model”. In: *Geophysical Research Letters* 34.16. DOI: `10.1029/2007gl030637`.

Wu, Z. J. et al. (2016). “Particle hygroscopicity and its link to chemical composition in the urban atmosphere of Beijing, China, during summertime”. In: *Atmospheric Chemistry and Physics* 16.2, pp. 1123–1138. DOI: `10.5194/acp-16-1123-2016`.

Wuerth, Christian et al. (2011). “Comparison of Methods and Achievable Uncertainties for the Relative and Absolute Measurement of Photoluminescence Quantum Yields”. In: *Analytical Chemistry* 83.9, pp. 3431–3439. DOI: `10.1021/ac2000303`.

Xiao, H. and B. Zhu (2003). “Modelling study of photochemical ozone creation potential of non-methane hydrocarbon”. In: *Water Air and Soil Pollution* 145.1, pp. 3–16. DOI: `10.1023/a:1023604007059`.

Xiao, Y. P. et al. (2008). “Global budget of ethane and regional constraints on US sources”. In: *Journal of Geophysical Research-Atmospheres* 113.D21. doi: `10.1029/2007jd009415`.

Xie, Xin et al. (2008). “Estimate of initial isoprene contribution to ozone formation potential in Beijing, China”. In: *Atmospheric Environment* 42.24, pp. 6000–6010. doi: `10.1016/j.atmosenv.2008.03.035`.

Xing, Jia et al. (2018). “Quantification of the enhanced effectiveness of NOx control from simultaneous reductions of VOC and NH3 for reducing air pollution in the Beijing-Tianjin-Hebei region, China”. In: *Atmospheric Chemistry and Physics* 18.11, pp. 7799–7814. doi: `10.5194/acp-18-7799-2018`.

Xu, J. et al. (2012). “Aerosol effects on ozone concentrations in Beijing: A model sensitivity study”. In: *Journal of Environmental Sciences* 24.4, pp. 645–656. doi: `10.1016/s1001-0742(11)60811-5`.

Yan, Y. Y. et al. (2019). “Global tropospheric effects of aromatic chemistry with the SAPRC-11 mechanism implemented in GEOS-Chem version 9-02”. In: *Geoscientific Model Development* 12.1, pp. 111–130. doi: `10.5194/gmd-12-111-2019`.

Yang, Y. D. et al. (2017). “How the OH reactivity affects the ozone production efficiency: case studies in Beijing and Heshan, China”. In: *Atmospheric Chemistry and Physics* 17.11, pp. 7127–7142. doi: `10.5194/acp-17-7127-2017`.

Ye, C. X., H. L. Gao, et al. (2016). “Photolysis of Nitric Acid and Nitrate on Natural and Artificial Surfaces”. In: *Environmental Science and Technology* 50.7, pp. 3530–3536. doi: `10.1021/acs.est.5b05032`.

Ye, C. X., X. L. Zhou, et al. (2016). “Rapid cycling of reactive nitrogen in the marine boundary layer”. In: *Nature* 532.7600, pp. 489–491. doi: `10.1038/nature17195`.

Yerramilli, A. et al. (2012). “Simulation of surface ozone pollution in the Central Gulf Coast region during summer synoptic condition using WRF/Chem air quality model”. In: *Atmospheric Pollution Research* 3.1, pp. 55–71. doi: `10.5094/apr.2012.005`.

Zhang, Huina et al. (2020). “Vertical profiles of biogenic volatile organic compounds as observed online at a tower in Beijing”. In: *Journal of environmental sciences (China)* 95, pp. 33–42. doi: `10.1016/j.jes.2020.03.032`.

Zhang, L. et al. (2016). “Potential sources of nitrous acid (HONO) and their impacts on ozone: A WRF-Chem study in a polluted subtropical region”. In: *Journal of Geophysical Research-Atmospheres* 121.7, pp. 3645–3662. DOI: `10.1002/2015jd024468`.

Zhang, Lingli et al. (2019). “Short-term and long-term effects of PM2.5 on acute nasopharyngitis in 10 communities of Guangdong, China”. In: *Science of the Total Environment* 688, pp. 136–142. DOI: `10.1016/j.scitotenv.2019.05.470`.

Zhang, Q. Q. et al. (2019). “Bias in ammonia emission inventory and implications on emission control of nitrogen oxides over North China Plain”. In: *Atmospheric Environment* 214. DOI: `10.1016/j.atmosenv.2019.116869`.

Zhang, Y., Z. Q. Li, and A. Macke (2002). “Retrieval of surface solar radiation budget under ice cloud sky: Uncertainty analysis and parameterization”. In: *Journal of the Atmospheric Sciences* 59.20, pp. 2951–2965. DOI: `10.1175/1520-0469(2002)059<2951:rossrb>2.0.co;2`.

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

Zheng, B. et al. (2018). “Trends in China’s anthropogenic emissions since 2010 as the consequence of clean air actions”. In: *Atmospheric Chemistry and Physics* 18.19, pp. 14095–14111. DOI: `10.5194/acp-18-14095-2018`.

Zhong, Qirui et al. (2017). “Global estimates of carbon monoxide emissions from 1960 to 2013”. In: *Environmental Science and Pollution Research* 24.1, pp. 864–873. DOI: `10.1007/s11356-016-7896-2`.

Zhu, J. and H. Liao (2016). “Future ozone air quality and radiative forcing over China owing to future changes in emissions under the Representative Concentration Pathways (RCPs)”. In: *Journal of Geophysical Research-Atmospheres* 121.4, pp. 1978–2001. DOI: `10.1002/2015jd023926`.

Zhu, L. et al. (2013). "Constraining U.S. ammonia emissions using TES remote sensing observations and the GEOS-Chem adjoint model". In: *Journal of Geophysical Research-Atmospheres* 118.8, pp. 3355–3368. DOI: `10.1002/jgrd.50166`.