A Statistical Analysis of Surface Ozone Variability Over the Mediterranean Region During Summer

Dissertation

zur

Erlangung des Doktorgrades (Dr. rer. nat.)

der

Mathematisch-Naturwissenschaftlichen Fakult.t

der

Rheinischen Friedrich-Wilhelms-Universit.t Bonn

vorgelegt von

Najmeh Kaffashzadeh

aus

Estahban, Iran

Bonn, 2018

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakult.t der Rheinischen Friedrich-Wilhelms-Universit.t Bonn

1. Gutachter: PD Dr. Martin G. Schultz

2. Gutachter: Prof. Dr. Andreas Hense

Tag der Promotion: 21 August 2018

Erscheinungsjahr: 2018

Abstract

The Mediterranean (Med) region contains the highest surface ozone (O_3^{surf}) concentrations in Europe, in particular in summer. Although atmospheric chemistry models (ACMs) are able to reproduce the regional O_3^{surf} maximum over the Med region, they generally exhibit a high bias over this region. There are many chemical and physical processes acting on different time scales that act as controlling factors of the O_3^{surf} variability and might not be simulated accurately by the models. Thus, detailed investigation of a single model is needed to understand why (or where) the model is not able to reproduce the observed O_3^{surf} variability.

In this study, we used two datasets to evaluate the O_3^{surf} variability over the Med region on the intraday (ID), diurnal (DU), and synoptic (SY) time scales. Observations from the TOAR database at 76 rural and 109 urban Med stations for summers 2010, 2011, and 2012 were compared to results of a WRF-Chem model simulation at 30 *km* resolution. As a first step, a scale analysis was performed on both datasets to determine the relative importance of the different time scales, i.e. ID, DU, and SY, for the O_3^{surf} variability. In a second step, a model performance indicator was used to quantify the amount of the model error in simulating the O_3^{surf} variability in each time scale. In the third step, a multiple linear regression (MLR) model was established for each component of the variability spectrum to identify the relationship between O_3^{surf} and several independent variables such as AT, RH, SP, U, V, and NO_x in the measurements and in the simulation.

The results of the scale analysis of the measured and simulated O_3^{surf} show that the model simulation is able to capture major features of the O_3^{surf} variability at each time scale for both rural and urban sites. However, the relative contributions of each time scale to the total O_3^{surf} variability differ from those of the observations. The model performance indicator reveals that the major part of the model variance error in simulating the O_3^{surf} variability over the Med region is associated with the model's difficulties in simulating the observed DU and SY variability. From the MLR analysis, it appears that the ID, DU and SY variability of the measured O_3^{surf} over the Med region are predominantly associated with the variability of NO_x, AT, and RH, respectively. However, in the WRF-Chem model, the relationship between O_3^{surf} and NO_x is much stronger than in the measurements for all time scales. Moreover, the model does not capture the observed relationship between O_3^{surf} and RH well. Contrary to prior expectations, the MLR analysis did not show a strong relationship between the observed O_3^{surf} and SP, U, and V, even though it is clear that local circulation patterns can affect measurements for example at a mountain or coastal site. This lack of correlation might be reproduced by the nonlinearity of the relationship between O_3^{surf} and these variables. Therefore, misrepresentation of the true relationship between variability of the O_3^{surf} and several variables, such as NO_x, RH, and AT, can be a possible reason for the model variance error.

To Masumeh and Gholamreza

Contents

Color codes in this thesis are:

- **Rural stations**
- Original data (ORG)
- Intraday spectral component (ID)
- Diurnal spectral component (DU)
- Synoptic spectral component (SY)
- Baseline spectral component (BL)
- **Urban stations**
- Original data (ORG)
- Intraday spectral component (ID)
- Diurnal spectral component (DU)
- Synoptic spectral component (SY)
- Baseline spectral component (BL)

List of Figures

List of Tables

1 Introduction

1.1 Problem statement

Surface ozone (O_3^{surf}) is distinguished as an air pollutant causing negative impacts on human health and vegetation at regional scale (US EPA, 2018: ozone pollution https://www.epa.gov/ozone-pollution, last accessed 13 March 2018). Many physical and chemical processes such as transport, photolysis at various temporal and spatial scales contribute to the O_3^{surf} levels. Air quality models are used to better understand the drivers of air pollution at a regional scale and to evaluate the roles of the interactions between emissions, meteorology, and chemistry (Mar et al., 2016). Atmospheric chemistry models (ACMs) including chemistry transport models (CTMs), which use meteorological fields as inputs (Seinfeld and Pandis, 2006), and coupled chemistry meteorology models (CCMM), which simulate meteorology and atmospheric chemistry jointly (Zhang, 2008; Baklanov et al., 2014) are routinely used for air quality forecasting (Bocquet et al., 2015). Although these models are successful in describing the major features of the O_3^{surf} variabilities, such as seasonal (Wang et al., 2011) and diurnal variability (Loon et al., 2007), the numerical simulations often exhibit a discrepancy with the observations. While these discrepancies can be often explained by the differences between models and measurements characteristics, such as point measurements versus model grid-boxes (Barnes et al., 2016), a more detailed investigation of a single model is needed to understand why the differences arise.

In recent years, the Weather Research and Forecasting model coupled with a Chemistry model (WRF-Chem, Grell et al., 2005) has been increasingly used for the air quality modeling studies over Europe (EU) (Solazzo et al., 2012; Tuccella et al., 2012; Žabkar et al., 2015; Mar et al., 2016). Two of these studies (Tuccella et al., 2012; Mar et al., 2016) have been dedicated to the evaluation of the WRF-Chem-simulated meteorology and chemistry over the European domain. In the recent study by Mar et al. (2016) it appears that although the WRF-Chem model shows strength in simulating the O_3^{surf} and meteorological variables over the Northern Europe, it exhibits weaknesses for the Southern Europe, including the Mediterranean region (Med).

Summertime O_3 concentrations over the Med region is a factor of 2.5–3 higher than in the hemispheric background troposphere in the boundary layer and up to 4 km altitude (Lelieveld et al., 2002). Results from the measurements at monitoring stations in Greece, Italy and Spain showed summer average O_3 values of about 60–70 ppb, significantly higher than values in Northern and Western EU (Kalabokas and Repapis, 2004; Paoletti, 2006; Millan et al., 2000). Results of the O_3 simulation from ACMs indicated higher O_3 concentrations over the Med Sea in comparison to that over the rest of EU (e.g. Johnson et al., 2001). Thus, both modeling and measurements studies confirmed the summertime high O_3 concentration over the MED region. However, various proofs have been suggested as an agent of high O_3 concentration over the region. For instance, the Med region hosts high pollution sources because of covering several megacities cities, such as Istanbul, and large urban agglomerations, such as Athene (Im et al., 2012). In addition, there is a transport of O_3 and its precursors, such as carbon monoxide (CO), from European countries to the region (Lelieveld et. al., 2002). The importance of the transport is not limited to the transport from the European countries. On a local scale, sea (land) breeze, i.e. wind towards the land (sea) in daytime (nighttime), has a large influences the O_3^{surf} concentration over the Med region, in particular over the Western Med (Millan et al., 1999; Velchev et al., 2011). On the other hand, there is a big concern in the simulation of lower-layer pollutants at the complex-terrain costal sites, where a ACM with an inadequate scale might not able to resolve the entire meso-scale interactions (Palau et al., 2005). In this thesis, advanced statistical techniques are employed to enable a deeper understanding of how the WRF-Chem model simulates the variability of O_3^{surf} mixing ratio at different stations over the Med region.

1.2 Factors controlling O_3^{surf} levels

1.2.1 Chemistry

Ozone (O_3) in the lower troposphere is produced by photochemical oxidation of its precursors such as volatile organic compounds (VOCs), methane (CH_4) and carbon monoxide (CO) in the presence of nitrogen oxides (NO_x = nitrogen dioxide (NO_2) + nitric oxide (NO)) (Crutzen et al., 1988; Lelieveld et al., 2000; Cooper et al., 2014; Monk et al., 2015).

In the O_3-NO_x photochemical cycle, O_3 forms via NO_2 photolysis within a few minutes (*R1*), while it is quickly removed after reacting with NO (*R2*) (Seinfeld et al., 1996).

(*R1*) $NO_2 + O_2 \xrightarrow{hv(\lambda<424nm)} NO + O_3$

(*R2*) $NO + O_3 \xrightarrow{k} NO_2 + O_2$

Although *R1* only takes place in sunlight, *R2* is independent of sunlight. *R2* is known as NO titration and it is an important process of O_3^{surf} removal associated with directly emitted NO (Sillman et al., 1999). The NO titration process only removes at most one O_3 per emitted NO (up to 1.5 O_3 per NO at night), whereas the process of O_3 formation typically produces four or more O_3 per emitted NO_x (Lin et al., 1988; Liu et al., 1987; Trainer et al., 1993; Sillman et al., 1998 and 1999). Nevertheless, the process of NO titration has a large impact in three situations: during nighttime, in winter and in vicinity of large power plants, since the process of O_3 formation is very slow (Sillman et al., 1999). This cycle (*R1-R2*) appears as a quasi-steady-state cycle, in which O_3 achieves a steady concentration, determined by the initial concentration of NO, the photolysis rate of NO_2 and the rate constant of the reaction between NO and O_3 (Seinfeld et al., 1996) (see Figure 1.1).

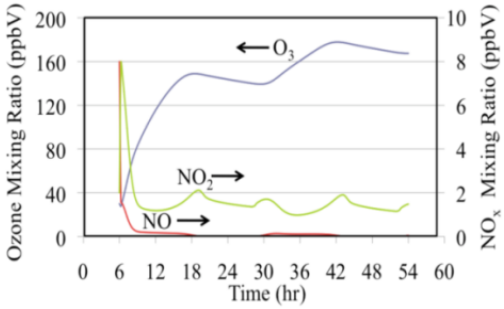

Figure 1.1 The mixing ratios of O_3, NO, and NO_2 for an ideal polluted urban atmosphere. (Stockwell et al., 2012)

In reality, the O_3^{surf} production arises when NO_2 levels increases due to reactions involving CO or volatile organic compounds (VOCs) in the presence of NO. These reaction sequences are usually initiated by reactions of CO or hydrocarbons (RHs) with OH. The reaction of the OH radical with CO produces a hydro-peroxy radical (HO_2) (*R3*) (see Figure 1.2). Similarly, reactions between RHs and OH produce HO_2 and organic peroxy radicals (RO_2 with R denotes an organic group, e.g. CH_3) (*R4*) (Sillman, 1999).

(*R3*) $CO + OH \xrightarrow{O_2} CO_2 + HO_2$

(*R4*) $RHs + OH \xrightarrow{O_2} RO_2 + HO_2$

The fate of both RO_2 and HO_2 radicals in polluted environments (NO > 400 pptv) is mainly reaction with NO with formation of NO_2 (Seinfeld et al., 1996).

(*R5*) $RO_2 + NO \xrightarrow{O_2} R'CHO + HO_2 + NO_2$

(*R6*) $HO_2 + NO \xrightarrow{O_2} OH + NO_2$

In R5, R'CHO represents intermediate organic species, typically including aldehydes and ketones. The directly emitted RHs and intermediate organics are collectively referred to as VOCs (Sillman, 1999). The rate constants for reactions

of HO_2 and RO_2 radicals with NO are nearly 1000 times faster than the rate constant of NO with O_3. Although HO_2 and RO_2 concentrations are 1000 times lower than O_3 concentrations, these reactions thus provide a mechanism for the net production of O_3 (Thornton et al., 2002). *R3* to *R6* do not only increase the O_3^{surf} production due to NO_2 production, but they also reduce the O_3^{surf} titration (*R2*) by reducing the NO concentration (*R5* and *R6*).

The largest photochemical sink for O_3 in the troposphere is photolysis with the formation of $O(^1D)$ and O_2 (*R7*). After the photolysis, a small part of $O(^1D)$ reacts with H_2O forming OH radicals (*R8*) (see Figure 1.2):

(R7) $O_3 \xrightarrow{h\nu(\lambda<340nm)} O(^1D) + O_2$

(R8) $O(^1D) + H_2O \longrightarrow 2\ OH$

As this O_3 removal path depends on water vapor (H_2O) mixing ratio, it is most effective at the low altitudes, where the radiation is intense and humidity is high (Seinfeld et al., 1996).

Another important tropospheric O_3 sink is deposition onto solid or liquid surfaces. Dry deposition processes account for about 25% of the total O_3 removed from the troposphere (Lelieveld et al., 2000). Dry deposition of O_3 to the terrestrial Earth surface is highly dependent on land cover. Deposition to non-vegetated surfaces is generally slower than deposition to vegetated surfaces (Wesely et al., 2000), the latter process varies according to plant species, and seasonal changes in leaf area index (Hardacre et al., 2015). In rural areas, dry deposition to terrestrial surfaces drives the diurnal variation in O_3^{surf} (Simpson, 1992).

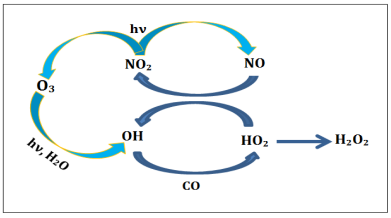

Figure 1.2. Mechanism for O_3-HO_x-NO_x-CO chemistry in the troposphere. (Adopted from D. Jacob, 1999)

1.2.2 Meteorology

Chemical O_3 production and deposition on the surface are known to be strongly affected by meteorological conditions. For example, O_3 levels tend to be higher under hot, sunny conditions favorable for photochemical O_3 production in the presence of precursors. At the same time, higher atmospheric temperatures (AT) cause convection, which can enhance vertical O_3 transport out of the region. Furthermore, wet, rainy weather with high relative humidity (RH) is typically associated with the low O_3 levels provided by lower intense photochemical production (Lelieveld et al., 1990; Tarasova et al., 2003).

Wind can transport air both horizontally and vertically. The influence of wind on the O_3 mixing ratio can have two effects. Transport can dilute O_3 either by transporting it far away from the site or by carrying clean air to the site. It can also increase O_3 by transporting polluted air to the site. For instance, lake (and sea) breeze circulations can increase (decrease) the O_3 mixing ratio over land by transporting polluted (clean) air from sea to the land and vice versa for land breeze (Wentworth et. al., 2015). If the boundary layer acts as a source of O_3 due to chemical production, the growth of the wind speed leads to the decrease of the O_3 concentration because of the vertical mixing.

Major episodes of high O_3^{surf} concentrations arise from stagnant high-pressure weather systems. These systems, which are characterized by a widespread sinking of air through most of the troposphere and a pronounced inversion of the AT profile, promote high O_3^{surf} concentrations. Moreover, winds associated with major high pressure systems are generally light that leads to a greater chance for O_3 and its precursors to accumulate in the planetary boundary layer (PBL) (National Research Council, 1991). The PBL extends from the earth surface to a capping inversion that typically locates at the height of 1–2 km by midafternoon over land surface. Below the capping inversion, i.e. 10-60% of PBL depth that is also called entrainment zone in the daytime, the PBL is divided into two sub-layers, i.e. mixed layer and surface layer. The mixed and surface layer construct 35-80% and 5-10% of PBL depth, respectively (Stull, 1988) (see Figure 1.3). At night, the radiative cooling leads to the formation of a stable surface layer and a uniform residual layer above it. Often, the O_3 concentration in the residual layer remains high due to the absences of the depositions and removal process such as emission sources aloft. Moreover, O_3 in the residual layer can transport to the long distance with the prevailing winds. In the morning, as the sun heats up the ground, the mixed layer starts developing. As the mixed layer grows, trapped pollutants in the residual layer can be entrained downward into the mixed layer. This effect leads to a rapid increase of the O_3^{surf} concentrations. In addition, the emitted emissions near the surface can be mixed into upper levels of the atmosphere (Zhang et al., 1999).

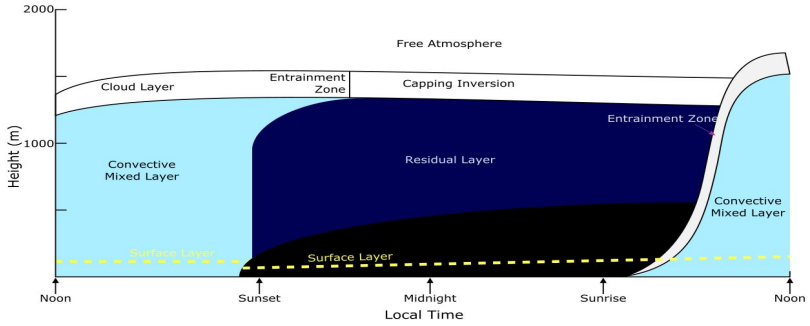

Figure 1.3. The structure of the planetary boundary layer height in a day. (Adopted from Stull, 1988)

1.2.3 Comparison of the temporal O_3 evolution at a rural versus an urban site

Anaylsis of the population distribution distinguish the station type of area between urban and or rural. Figure 1.4 shows the average hourly evolution of the measured O_3 concentration in summertime at a rural and an urban site near the city of Malaga (Spain). From Figure 1.4, it appears that the O_3 concentrations at both stations reach a minimum in the early morning and a maximum in the late afternoon. The daily cycle results from the effect of daytime photochemical production and downward transport of O_3 rich from upper levels, combined with O_3 loss by dry deposition, when photochemical production cease (Seinfeld et al., 1996). The average O_3 concentration at the rural site is higher than that at the urban site (see Figure 1.4). This can attribute to the lower number of sinks in the rural area and to the arrival of precursors transported from mid and long distances (Hov et al., 1992; Saitanis, 2003; Duenas et al., 2004). The diurnal variation, in particular in the early morning, at the urban site is more pronounced than that at the rural site. That is a clear indicator of lower sources of NO at the rural site (Duenas et al., 2004).

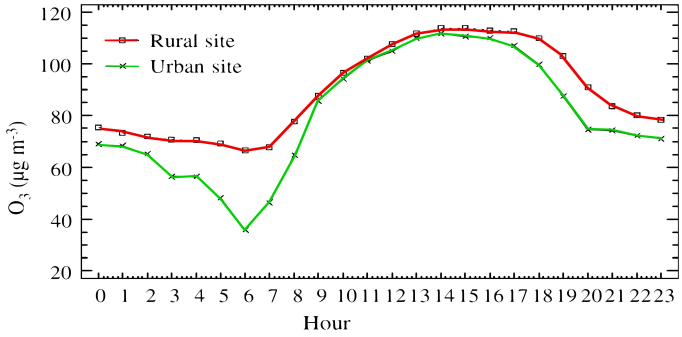

Figure 1.4. Average hourly evolution of O_3 concentrations in summer at a rural and an urban site near the city of Malaga (Spain). (Duenas et al., 2004)

In addition, the assessment of emission sourcses in the srounding area divides stations types into traffic, industrial, and background. For instance, at the traffic sites, where have a high antropogenic emission sources such as NO_x, and CO, the O_3^{surf} concentration is less than the background. That might be generatet from the larger contribution of the NO titration to the O_3^{surf} concentration in these areas.

1.3 Modeling of the O_3^{surf} and its uncertainties

ACMs simulate the time evolution of 3D fields of traces gases. The O_3 concentration in ACMs is calculated as the sum of three comprehensive processes as:

$$C_{O3}{}_t^i = chem_t^i + trans_{h,v}{}_t^i + dep_t^i \quad (1.1)$$

here, $C_{O3}{}_t^i$ is the O_3 level at a given time (t) and location (i), $chem_t^i$ represents the net effect of photochemical processes, i.e. both production and sink, on the O_3

level at a given t and i, $trans_{h,v}{}_t^i$ represents the net transport impact on O_3 level, including both horizontal (h) and vertical (v) transports, at a given t and i, and dep_t^i is the amount of O_3 removed by both dry and wet deposition processes (Jacob, 1999).

Chemistry in ACMs is computed by chemical mechanisms developed from laboratory and field measurements. They consist of chemical species, reactions, rate constants and photochemical data (Stockwell et. al., 2012). Uncertainties in the chemical mechanisms could introduce large uncertainties into the product yields and concentrations calculated in photochemical models (Gao et al., 1996). Chemistry as a source of uncertainty was analyzed by Gao et al. (1996) finding that known uncertainties in reaction rates and stoichiometric coefficients caused 20% uncertainty in simulated concentrations of O_3 and most other species (Sillman, 1999).

Meteorology especially wind speed is one of the sources of uncertainties in simulation of the O_3 concentrations. Uncertainties in wind speeds arise from imprecise measurements, the stochastic nature of wind and the need for interpolation based on measurements separated by 200 km or more. Moreover, ACMs show difficulties in resolving local flow patterns arising from the complex topography and land–sea circulation that can effect on the O_3 concentration (Pfister et al., 2013). In addition, uncertainties in estimating PBL height (PBLH) (Marsik et al., 1995; Tong et al., 2011) contribute to errors in simulation of the O_3 concentration (Sillman, 1999).

Despite the importance of dry deposition processes, they are some of the most uncertain and poorly constrained aspects of the tropospheric O_3 budget (Wild, 2007). This uncertainty arises from the complexity and heterogeneity in dry deposition, which depend on meteorological conditions and the characteristics of the surface, along with a lack of long-term observation dataset for many surface covers, including oceans, tropical forests and deserts (Hardacre et al., 2015). Many ACMs use a dry deposition scheme developed by Wesely (1989) with some modifications (Wang et al., 1998; Ganzeveld, 1995; Val Martin et al., 2014).

One the other hand, for simulating of the O_3 concentration, an ACM requires emission data for several species, such as NO, CO, CH_4, etc., that are given as input. These input data create from the emission inventories, by which NO_x and other species concentration estimated for each individual source sectors in a region according to limited number of measurements. Here, large uncertainties are introduced due to the uncertainties on all the input parameters in the calculation (Saikawa et al., 2017). The uncertainty in temporal variability in NO_x emissions becomes larger with decreasing spatial scale of emissions, especially where surface types are inhomogeneous (Ding et al., 2017). In addition, uncertainties can arise from representing local emissions and chemistry, transport, and mixing

processes up to uncertainties in boundary conditions (BCs) (Pfister et al., 2013). Global models commonly used for BCs have underestimated high pollution levels in transported plumes (Rastigejev et al., 2010), which potentially underestimates the degree to which long-range pollution contributes to high O_3^{surf} events. Quite often the chemical lateral BCs are introduced to the model as time and space invariant concentration fields (i.e. climatological concentration) having a negative impact on the quality of the O_3^{surf} simulation (Katragkou et al., 2010). The importance of temporally varying chemical BCs in ACMs has been mentioned in the literatures (Akritidis et al., 2013; Andersson et al., 2015).

1.4 Literature review

Regular enhancement of the O_3^{surf} concentration over the Med region in summer is shown by numerous modeling studies and confirmed by observation studies (Lelieveld et al., 2002; Kalabokas et al., 2004; Richards et al., 2013; Doche et al., 2014; Safieddine et al., 2014). The large photochemical activities, which are observed over the Med region under clear sky conditions and high solar intensity present in the summer, has been suggested as one of the major mechanism controlling the O_3^{surf} concentration over the region (Lelieveld et. al., 2002). In addition, depending on the meteorological conditions, the significance of the contribution of photochemical activities to the O_3^{surf} concentration differs. For instance, high-pressure systems, which prevail in the summertime over Central and Eastern Med sub-regions, create a low dispersion condition in the PBL increasing the O_3^{surf} concentration. On the contrary, periods with low O_3^{surf} concentration over Eastern part are associated with low atmospheric pressure system where high uplifting air movements, strong westerly airflow, that cause dispersion of pollutants, reduce the O_3^{surf} concentration (Kalabokas et al., 2008; 2016). Doche et al. (2014) showed that relative position and strength of the meteorological synoptic systems are the key factors to explain the O_3^{surf} variability over the Med basin. Several studies involving measurements as well as model simulations (Millan et al., 1997, 2000) have addressed the causes of episodes with high ozone-mixing ratios over Western Med. It was found that the typical synoptic meteorological conditions found during the summer in this part of the Mediterranean, with a lack of strong synoptic advection combined with the orographic characteristics and the sea and land breezes, favor episodes where high levels of O_3 over this region. Thus, it is evident that, not only photochemical reactions but also synoptic and meso-scale meteorological conditions do contribute to the O_3^{surf} variability over the Med region, and their contributions vary from site to site.

One aspect of concern is that aforementioned phenomena not only contain a specific impact on the O_3 concentration, but also occur in a various time scales. For instance, NO titration influence on O_3 concentration in rush hours during

morning and afternoon. Thus, its influence of the titration on the O_3 concentration occurs in a very short time scale. On the other hand, the photochemical reactions have a daily influence on O_3 concentration. As an example, the O_3 level builds up in the daytime and decreases at nighttime. The scale analysis has been used for the study of O_3 variability in different time scales (Rao et al., 1997; Hogrefe et al., 2000; Tarasova et al., 2003) finding that more than 50% of total O_3 variability arises from the diurnal variability. In addition, the spectral analysis has been used in several modeling studies. For instance, a series of AQMEII (Air Quality Model Evaluation International Initiative) models evaluated versus measurements over the European (EU) and North American (NA) regions (Hogrefe et al., 2013; Solazzo et al., 2016). Hogrefe et al. (2013) showed that all the modeling systems underestimated the observed variance in all scales and had a worse performance at locally influenced sites. Moreover, a poor correlation has been found with the observed O_3 fluctuations on the intraday time scale for all the modeling systems over both regions. On the other hand, Soalzzo et al. (2016) showed that the variance model error for the simulated O_3^{surf} is mainly associated with the diurnal variability of O_3^{surf} at a timescale of one day or two. In addition, they mentioned that the variance of the spectral components is not associated with the area type of monitoring stations.

1.5 Objectives and structure of the thesis

As seen in section 1.4, O_3^{surf} variability arises from various processes within different time scales, such as intraday (ID), diurnal (DU), and synoptic (SY), and the DU variability contributes to more than 50% of the total O_3^{surf} variability (Hogrefe et al., 2000). Here, a few questions arise about the observed O_3^{surf} variabilities over the Med region that will be answered in this thesis.

Q1) how much do the O_3^{surf} variabilities in each time scale, such as DU, ID, and SY, contribute to the total O_3^{surf} variability over the Med region? Considering that over the Med region, the SY meteorological conditions have a large influence on the O_3^{surf} variability.

Q2) how does the O_3^{surf} variability in each time scales change moving from a rural to an urban site? Considering that, the total O_3^{surf} variability at a rural site is different from an urban site (see section 1.2.3).

Q3) which variable(s), such as AT, RH, SP, U, V, and NO_x, have a larger influence on the O_3^{surf} variability at each time scales over the Med region?

To answer these questions, we used observations from the Tropospheric Ozone Assessment Report (TOAR) database at 76 rural and 109 urban Med stations for summers 2010, 2011, and 2012. At first, a scale analysis was applied to determine the O_3^{surf} variability at each time scale ID, DU, and SY. In the next step, differences of the O_3^{surf} variability at each time scale at the different station types were

investigated. Then, a multiple linear regression (MLR) analysis was used to quantify the relationship between O_3^{surf} and the variables such as AT, RH, SP, U, V, and NO_x at each time scales (see Table 1.1).

As mentioned in section 4.1, the results of the modeling study (Solazzo et al., 2016) showed that the DU variability of the O_3^{surf} has the largest contribution to the model variance error over NA and EU regions. Moreover, the variance model error of the DU and SY are peculiar to each model and have to assess individually. Thus, a few questions arise about the simulated O_3^{surf} variability by the WRF-Chem model over the Med region that will be addressed in the thesis.

Q4) to what extent does the WRF-Chem model reproduce the observed O_3^{surf} variability at each time scale?

Q5) is the model variance error for the O_3^{surf} simulations over the Med region majorly related to the DU variability similar to the NA and EU?

Q6) does the WRF-Chem model simulation show the observed relationship between O_3^{surf} and other variables? To what extend does the model variance error arise from the misrepresentation of the relationship between O_3^{surf} and other variables?

To answer these questions, we used the model output of the WRF-Chem model simulation for the similar variables and time-periods as the observations dataset. Similar to the observation, the spectral analysis was applied to the simuated datasets. The results of it compared with the observations. We identified the model performance at each time scales by using a model performance indicator i.e. modified mean square error. An MLR analysis was performed on the simulation dataset and compared with the observations (see Table1.1).

Accordingly, the outline of the thesis structure is as follows. In Chapter 2, we describe the datasets including observations and the WRF-Chem model simulation that used in this study. A description of procedures and methods, used for the O_3^{surf} analysis in this study, is given in Chapter 3. Results of the application of the method to the dataset are shown in Chapter 4. The discussion of the results and their accuracies are described in Chapter 5. Finally, the summary and outlook are given in Chapter 6.

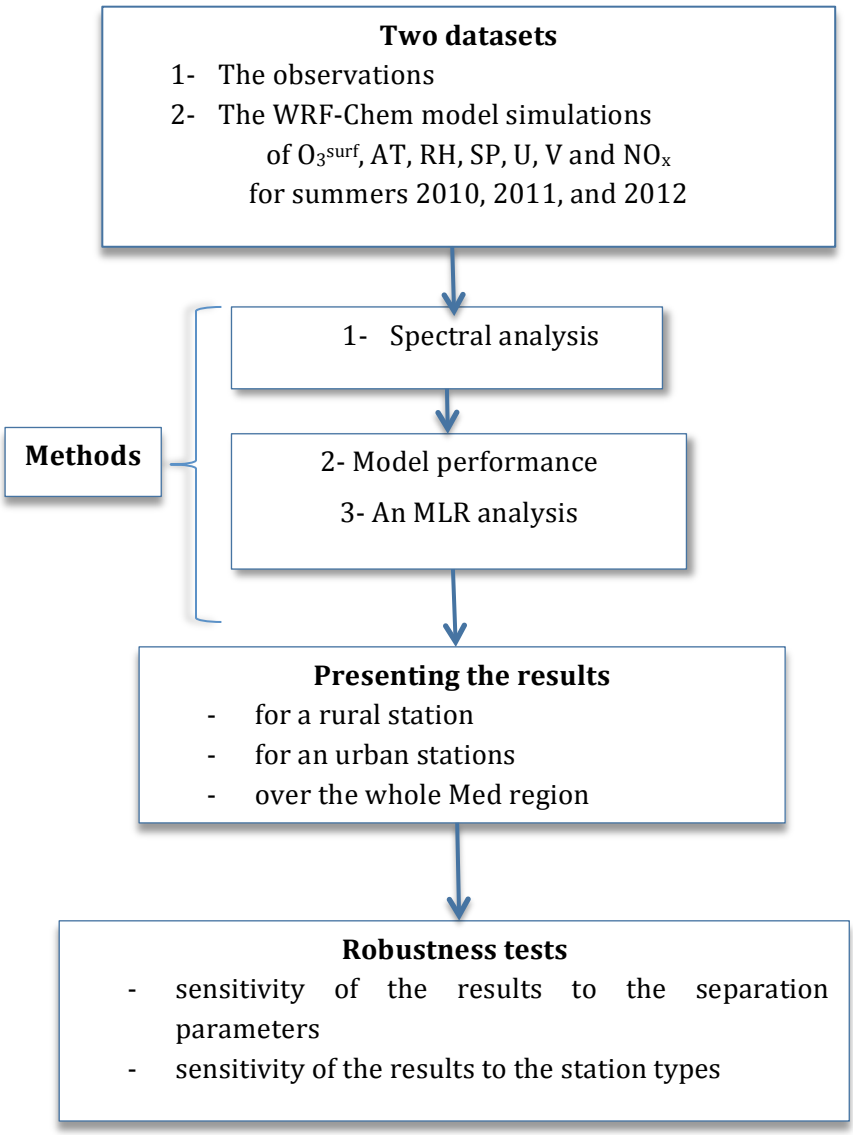

Table 1.1. A summary of the contents in the thesis.

2 Data

2.1 Observations

The Tropospheric Ozone Assessment Report (TOAR) has produced a relational database of global O_3^{surf} observations that are available through a web interface (https://join.fz-juelich.de) as described by Schultz et al. (2017). In addition to O_3^{surf}, the database contains measured data for several chemical trace gases such as NO, NO_2, and CO. However, the data for NO, NO_2, and CO confines to certain stations located mostly in EU, and for a limited period. Besides that, this database hosts data for several assimilated meteorological variables such as AT, RH, etc. resulting from the 6 km reanalysis with the COSMO model (COSMO-Rea6; Bollmeyer et al., 2014).

2.1.1 Chemical species

The measured data for O_3^{surf}, NO, and NO_2 were collected for summers (JJA-June, July, and August) 2010, 2011, and 2012. Since this study focuses on the O_3^{surf} over the Med region, the data were selected for longitude and latitude range of 10°W-35°E and 28°N-48°N, respectively, at low altitude sites, i.e. ''stations_alt'' < 1500 m according to the TOAR categorization. The selected stations cover two of the area types for stations, rural and urban. Rural sites tend to exhibit larger O_3^{surf} mixing ratio than urban sites and there is no tendency of increasing values with altitude in urban environments. This indicates that O_3^{surf} in the urban environment is primarily controlled by local chemical processes and to a lower degree influenced by large-scale advection (Schultz et al., 2017). The data were taken for nearly 600 stations. The number of the stations varies depending on the trace gases and the year (Figure 2.1).

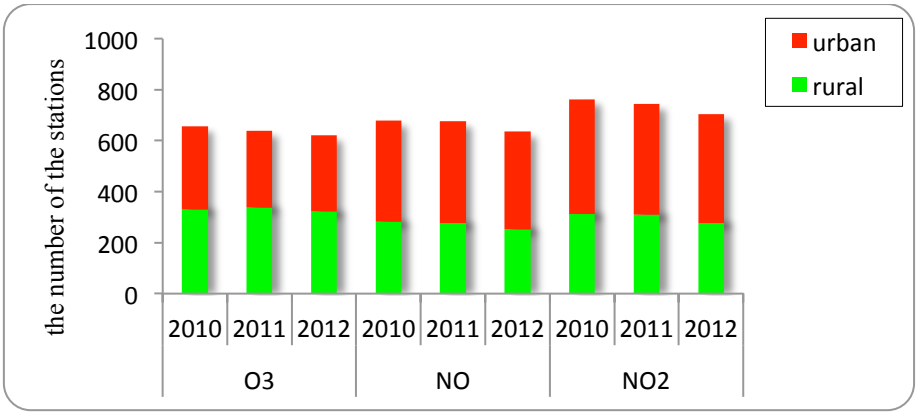

Figure 2.1. The number of the rural and urban stations containing the measured data for three chemical species, i.e. O_3, NO, and NO_2, for summers of 2010, 2011 and 2012 over the Med region.

As there is no warranty that the stations cover data for the chemical species over the whole period, stations having more than 10% missing data-points of the O_3^{surf}, NO, and NO_2 for summers 2010, 2011, or 2012 were not considered in this analysis, to obtain a regular and consistent dataset. As figure 2.2 shows, there were only 100 and 150 stations for the rural and urban sites, respectively.

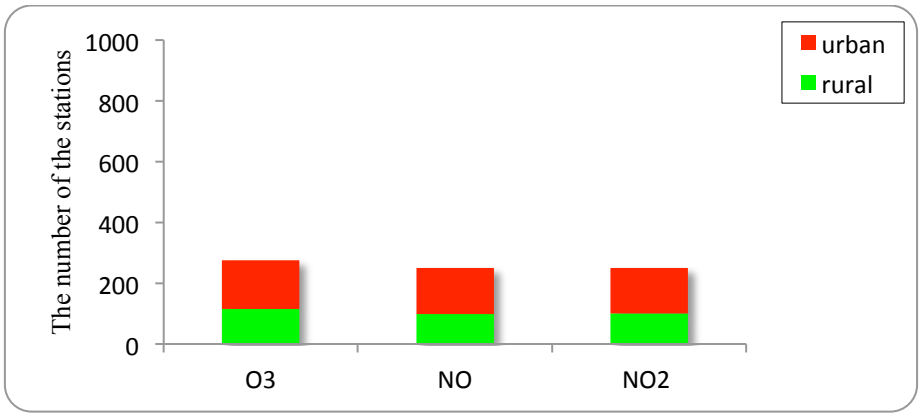

Figure 2.2. The number of the rural and urban stations containing more than 90% data-points of the O_3, NO, and NO_2, for summers of 2010, 2011 and 2012 over the Med region

In the next step, only the sites containing measurements for O_3^{surf}, NO and NO_2 were selected. This additionally dropped the number of stations to 87 for rural and 137 stations for urban area. Of these stations, 9 rural and 53 urban stations from southern France contained all required data except NO data for summer 2010. To retain these stations in the analysis, data for all the variables in summer 2010 at these sites were replaced by their value in summer 2012. The geographical locations of the selected sites are shown in figure 2.3.

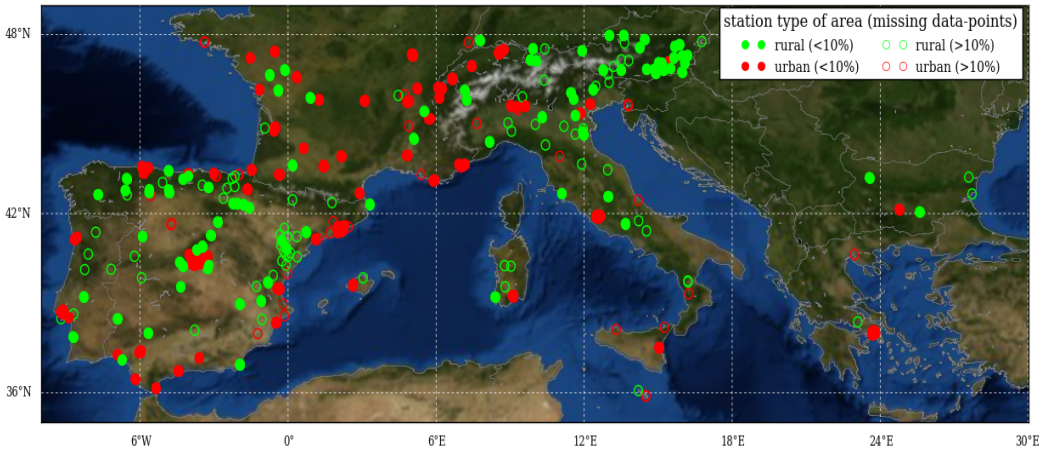

Figure 2.3. The geographical location of the stations having the measured data for the entire variables, i.e. O_3, NO, NO_2, AT, RH, SP, U, and V, for summers 2010, 2011 and 2012.

The TOAR data support metadata information including station type of area, such as rural, suburban, urban, and remote, and station-types, such as background, industrial, traffic, and unknown (Schultz et al., 2017). The number of the stations for each station types is given in Table 2.1. According to the results in Table 2.1, nearly 62 of the rural stations in this study belong to the background and the rest 14 belong to industrial types. In the urban area 84, 5, and 9 stations belongs to the background, industrial and traffic site, respectively (see table 2.1).

Table 2.1. The number of the stations for each station-type, such as background, industrial, traffic, and unknown, in this study.

stations type of area	stations type	numbers
rural	background	67
	industrial	17
	traffic	3
urban	background	94
	industrial	7
	traffic	35
	unknown	1

2.1.2 Meteorological data

Data of AT, RH, SP, U, and V for summers 2010, 2011, and 2012, were obtained from the TOAR database. The data were available for all stations, i.e. 100 rural and 150 urban. Since the data for these variables are the product of the COSMO-rea6 that is interpolated to each station location, they have no missing value.

2.2 The WRF-Chem model

The results of the Weather Research and Forecasting with Chemistry (WRF-Chem) model version 3.7.1 were analyzed in this study. The model simulations were performed by the research group of T. Butler at the Institute for Advanced Sustainability Studies (IASS), Potsdam, Germany. A description of the model setup is given in the next section.

2.2.1 Description of the model setup

A single domain, that covers the area between 25.4°N-62.1°N and 22°W-47°E (see Figure. 2.4), with 30 km grid spacing and 35 vertically stretched layers from the ground up to 50 hPa was used. Model simulations were conducted for the period of 1st May to 31st August of 2010, 2011 and 2012. The first month of output was treated as model spin up (i.e. the length of time for which the model has been run since initialization until it has approached its own equilibrium (Seferian et al., 2016)) and was discarded. The instantaneous model output, stored every hour, was used for the analysis. The physics options such as schemes for the radiation, land surface, PBL, and microphysics of the model used for this study include the Morrison double-moment microphysics scheme (Morrison et al., 2009), the Grell-Freitas cumulus parameterization (Grell and Freitas, 2014), the Rapid Radiative Transfer Model (Iacono et al., 2008) for the longwave and shortwave scheme, the Yonsei University boundary-layer parameterization (Hong et al., 2006), and the Monin-Obukhov scheme for the surface layer (Jimenez et al., 2012). The setting file (i.e. namelist) for this simulation is available in appendix F.

The initial and lateral boundary conditions for the meteorological fields were taken from the European Centre for Medium-Range Weather Forecasts (ECMWF) reanalysis products (https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era-interim). Initial and boundary conditions for chemical fields were obtained from the global Model for Ozone and Related chemical Tracers (MOZART) simulations (Emmons et al., 2010). The boundary condition for both meteorology and chemistry were updated every six hours, and then interpolated in time by the WRF-Chem model. Since a single simulation was performed over a four-month period, a nudging technique was applied for horizontal winds, temperature and water vapor above the PBL using the ECMWF analyses products to ensure that the simulated large-scale circulation does not drift from the measured synoptic conditions. Anthropogenic emissions were obtained from the TNO-MACC III emission inventory for Europe (Kuenen et. al., 2014). As the model domain used in this study extends beyond the edges of TNO-MACC III inventory, emissions at the domain edges of this study were filled using data from the HTAP V2 inventory (http://edgar.jrc.ec.europa.eu/htap_v2). Biomass burning emissions were based on the Fire INventory from NCAR (FINN) (Wiedinmyer et al., 2011). Biogenic emissions were computed on-line using the

Model of Emissions of Gases and Aerosols from Nature (MEGAN) (Guenther et al., 2006). In this setup, gas-phase chemistry was simulated by the MOZART chemical mechanism for NOx emissions. The photolysis rates were computed using the Fast Tropospheric Ultraviolet and Visible (FTUV) Radiation Model (Tie et al., 2003, Li et al., 2005). The dry deposition was calculated following Wesely (1989) resistance method, while the wet removal scheme was based on the study of Neu et al., (2012). The WRF-Chem model couples the online meteorology simulations with the chemistry, meaning they are solved together in a physically consistent manner (e.g. Zhang, 2008). The meteorology and chemistry components in the WRF-Chem model use the same horizontal and vertical grids and the same time-steps, eliminating the need for temporal interpolation (e.g. Grell et al., 2005).

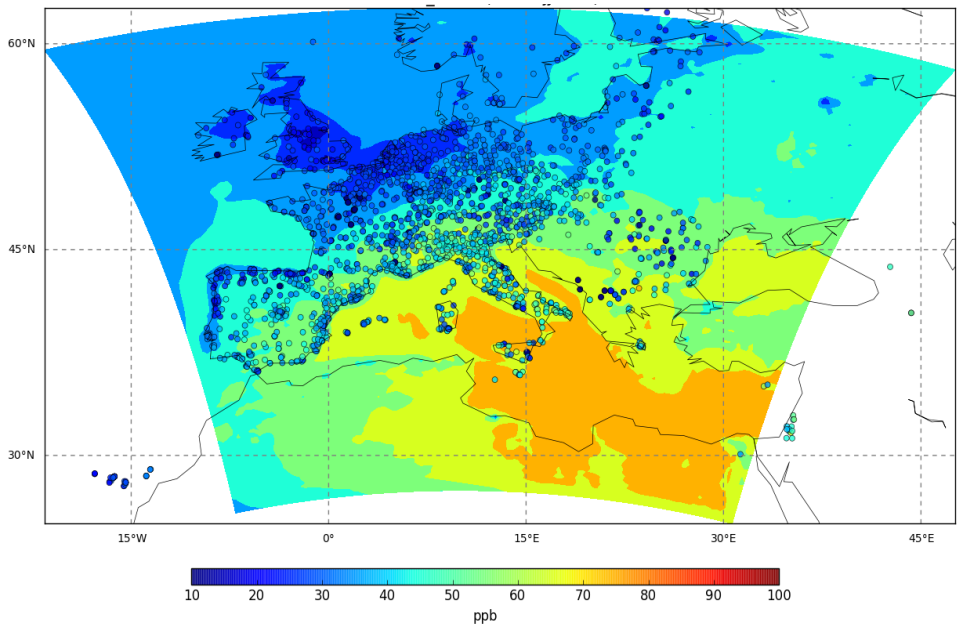

Figure 2.4. Simulated O_3^{surf} mixing ratio by the WRF-Chem model avaraged over summer (JJA) 2012. The circles show the measured O_3^{surf} mixing ratio

Chapter 3. Methodology

3 Methodology

3.1 Pre-screening of the stations according to NO_x bias

Although NO_x is recognized as one of the main precursors for the O_3 formation, an accurate estimate of NO_x in the models is difficult to achieve. As mentioned in section 1.2, the uncertainties in the estimated NO_x in a model majorly stem from the uncertainties in the emission inputs. A concern arise when in the models NO_x levels are majorly determined by the emissions, one can assume some representativeness of the corresponding station only if the simulated and measured NO_x at a station are in a reasonable agreement. To ensure the representativeness of the station in this study, only the stations showing a lower error (i.e. high accuracy and precision) of the simulated NO_x when compared to the measured NO_x were selected (pre-screening process). Two criteria were defined for the pre-screening process. The first considers the bias of estimated NO_x mixing ratio at the sites, while the second deals with the variability of the bias in estimated NO_x among sites. Accordingly, only the stations containing a low bias for estimated NO_x mixing ratio and for which the bias is comparable with the bias at other sites, were chosen for this study. This method includes four steps as follow:

1- Calculating the difference (C^i_{diff}) between simulated and measured NO_x mixing ratio:

$$C^i_{diff} = C^i_{NOx(WRF-Chem)} - C^i_{NOx(TOAR)} \quad (3.1)$$

here $C^i_{NOx(TOAR)}$ and $C^i_{NOx(WRF\text{-}Chem)}$ refer to the NO_x mixing ratio at the station i in the TOAR and WRF-Chem datasets, respectively.

2- Calculating the difference (σ^i_{diff}) between standard deviation of the simulated and measured NO_x mixing ratio:

$$\sigma^i_{diff} = \sigma^i_{NOx(WRF-Chem)} - \sigma^i_{NOx(TOAR)} \quad (3.2)$$

here $\sigma^i_{NOx(TOAR)}$ and $\sigma^i_{NOx(WRF\_Chem)}$ indicate the standard deviation of the estimated NO_x mixing ratio at the station i for the measurement and the WRF-Chem data, respectively.

3- Determining an upper ($U_{C_{NOx}}$) and a lower ($L_{C_{NOx}}$) limit for the estimated NO_x mixing ratio:

$$U_{C_{NOx}}, L_{C_{NOx}} = C_{diff}^{avg} \pm 1.5 \times \sigma_{C_{diff}} \quad (3.3)$$

here C_{diff}^{avg} and $\sigma_{C_{diff}}$ are the average and standard deviation of the C_{diff}^{i}, respectively.

4- Determining an upper ($U_{\sigma_{NOx}}$) and a lower ($L_{\sigma_{NOx}}$) limit for standard deviation of NO_x mixing ratio:

$$U_{\sigma_{NOx}}, L_{\sigma_{NOx}} = \sigma_{diff}^{avg} \pm 1.5 \times \sigma_{\sigma_{diff}} \quad (3.4)$$

here σ_{diff}^{avg} and $\sigma_{\sigma_{diff}}$ are the average and standard deviation of the σ_{diff}^{i}, respectively. This method was applied for the simulated and measured NO_x mixing ratio at both rural and urban sites. Figure 3.1 shows the stations chosen for this study (inside of the dashed green square) together with the outlier stations (outside of the dashed green square). The sides of the dashed green square in Figure 3.1 indicate $U_{C_{NOx}}$, $L_{C_{NOx}}$, $U_{\sigma_{NOx}}$, and $L_{\sigma_{NOx}}$. The circles in the dashed green squares in Figure 3.1 show the station where there is no high discrepancy between the simulated and measured NO_x mixing ratio (*ppb*) at that site and in comparison to other sites. This is on the contrary for the circles outside of the dashed green square, which indicates the stations having a large discrepancy between the simulated and the measures NO_x. From the rural and urban sites, 11 and 28 stations were identified as outlier stations, respectively. The number of the remaining stations is given in Table 3.1. These outlier stations constitute 13% and 26% of the rural and urbans sites. Moreover, they mainly belong to the traffic and industrial sites, i.e. red and yellow circles in Figure. 3.1.

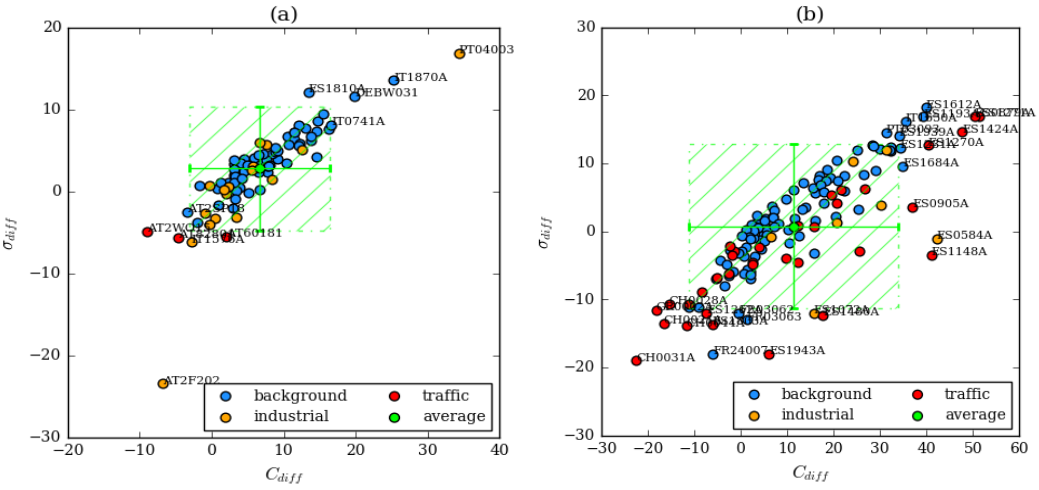

Figure 3.1. Pre-screening of the stations, (showing with circles) at the **(a)** rural and **(b)** urban sites. The circles inside (outside) the dashed green squares show the selected (outlier) stations in this study. One urban station, i.e. GR0032, has been not shown in this figure, because of the extremely high discrepancy between the simulated and measures NO_x mixing ratio at this site.

Table 3.1. The number of the stations after pre-screening of the stations in this study.

stations type of area	stations type	numbers
rural	background	62
	industrial	14
	traffic	0
urban	background	84
	industrial	5
	traffic	19
	unknown	1

3.2 Spectral decomposition

3.2.1 Definition of scale (spectral) analysis

The presence of various scales of motion in time series of meteorological and air quality variables can complicate analysis and interpretation of data (Rao et al., 1997). Separation of time series of O_3 and meteorological data into synoptic, seasonal, and long-term components is necessary since the processes occurring at different frequencies are caused by different physical phenomena: the synoptic-scale component is attributable to weather and short-term fluctuations in precursor emissions, seasonal scale to variation in the solar angle, and long-term scale to changes in climate, policy, and/or economics (Rao et al., 1994, 1995; Porter et al., 1996). The scale (spectral) analysis is a method to separate the O_3 time series into different temporal components (Hogrefe et al., 2000). The concept of scale analysis is widely used in physical science, including meteorology and climatology and air pollution (Goody et al., 1998; Salcedo et al., 1999; Hogrefe et al., 2003). In recent years, several studies have applied this concept for the analysis of measured O_3 and related meteorological variables (Tarasova et al., 2003; Seo et al, 2014) for a better understanding of the processes that control the relationships (statistical) between O_3^{surf} and meteorological parameters. Furthermore, this concept has been used for the evaluation of CTMs to discern which of the component(s) or scale(s) of forcing, meaning that different atmospheric processes influencing the predicted and observed pollutant concentrations, are simulated well by the model and the component(s) or scale(s) of forcing needing further improvement in the model (Kang et al., 2013; Solazzo et al., 2013, 2016).

3.2.2 Spectral decomposition technique

Several decomposition techniques such as Kolmogorov-Zurbenko (KZ) (Rao and Zurbenko, 1998), wavelet-transform (Lau and Weng, 1995), monthly anomaly technique (Wilks, 1995), adaptive window Fourier transform (Zurbenko and Porter, 1998) have been developed for the separation of time scales in a time series. In this study, the KZ filter was chosen due to the simplicity of its operation and dealing with. In addition, the KZ filter is able to work efficiently also in

presence of missing data which is one of the reasons for the widespread application of this technique (Rao et al., 1997; Hogrefe et al., 2000; Kang et al., 2013; Galmarini et al., 2013; Seo et al., 2014; Solazzo et al., 2016). Therefore, using this filter in this study is providing the opportunity to compare the results from this work with the outcomes from previous investigations.

3.2.2.1 Kolmogorov-Zurbenko filter

The KZ filter is a technique that can be used to calculate the contribution of various scale of motion to the total variance of the variables (Eskridge et al., 1997). This filter was first introduced by Kolmogorov and later formalized by Zurbenko (Zurbenko, 1989). Theoretically, the filter is based on the iterative moving average (MA) in which the simple MA of m points is computed by:

$$S(t) = \frac{1}{m}\sum_{j=-(m-1)/2}^{(m-1)/2} ORG(t)_j \quad (3.5)$$

Here ORG and t represent the original time series and its time step (or sampling resolution, e.g. one hour), respectively, and S is the input for the each iteration. Therefore, the filter can be expressed as:

$$KZ_{m,k} = R_{i=1}^{k}\left\{ J_{p=1}^{w_i} \left[S(t_i)_p\right] \right\} \quad (3.6)$$

Here m and k are window length and number of iteration, respectively. R and J represent iteration and running window, respectively, and w_i is defined as:

$$w_i = L_i - m + 1 \quad (3.7)$$

Here L_i is the length of $S(t_i)$.

The square transfer function of the $KZ_{m,k}$ is given by:

$$\left|H_{m,k}(\omega)\right|^2 = \left[\frac{1}{m}\,\frac{sin\,(m\pi\omega)}{sin\,(\pi\omega)}\right]^{2k} \quad (3.8)$$

Here ω is the frequency (unit of days per period). $KZ_{m,k}$ is a low pass family filter, in which high frequency (short time period) variations are removed from the time series (Eskridge et al., 1997). The band of frequency and the level of suppression in this filter are controlled by m and k, respectively. $KZ_{103,5}$ removes all frequencies higher than 0.002 per hour (≈ lower than 20 days), while $KZ_{15,3}$ eliminates all frequencies higher than 0.05 per hour (≈ lower than one day) (Figure 3.2). Accordingly, $KZ_{103,5}$ contains oscillations occurring in a long time period, whereas $KZ_{15,3}$ contain oscillations within a short time as well as a long time range. In figure 3.2, it is possible to observe how lower frequencies and more suppression appear in the time series by increasing m and k, respectively. Thus, the desired frequencies in this filter can be obtained by modifying m and k. As the lowest and highest resolvable frequencies in a time series depend on the length of data records and sampling resolution, respectively, the shortest and largest resolvable time periods in this study are 2 hours and 30-40 days, respectively.

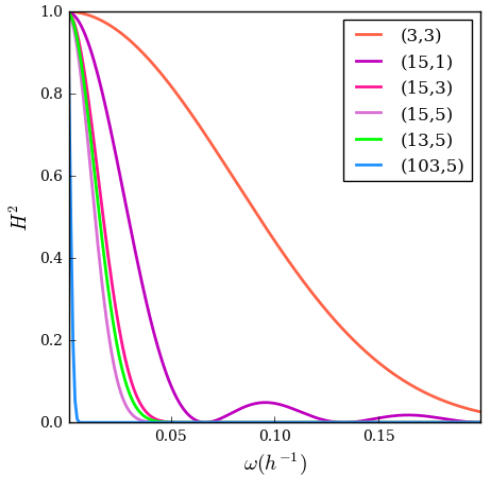

Figure 3.2. The dependency of the KZ transform function (*H in* Eq. 3.8) to various sets of (m_i, k_i). In this figure ω is a frequency (per hour). Lower frequency and more suppression appear by increasing m and k, respectively.

A simple criterion to filter all oscillations smaller than a certain time step d is (Eskridge et al., 1997):

$$m \times \sqrt{k} \leq d \tag{3.9}$$

The high frequency which are eliminated can be reconstruct by taking the difference between *ORG* and $KZ_{m,k}$. A band-pass filter is obtained by making the difference between two KZ filters with different m and k. In this study, several spectral components for the O_3^{surf} mixing ratio and for other meteorological variables were estimated as:

$$ID(t_{<12h}) = ORG(t) - KZ_{m1,k1} \tag{3.10}$$

$$DU(t_{12h\text{-}2.5d}) = KZ_{m1,k1} - KZ_{m2,k2} \tag{3.11}$$

$$SY(t_{2.5d\text{-}21d}) = KZ_{m2,k2} - KZ_{m3,k3} \tag{3.12}$$

$$BL\ (t_{>21d}) = KZ_{m3,k3} \tag{3.13}$$

here m_1, m_2, m_3 and k_1, k_2, k_3 are set to 3, 13, 103 and 3, 5, 5, respectively. The sum of all spectral components recreates the ORG:

$$ORG\ (t) = ID(t) + DU(t) + SY(t) + BL(t) \tag{3.14}$$

3.2.2.2 Error propagation in spectral components

A clear-cut separation of the components of Eq. 3.14 cannot be achieved, as the separation is a non-linear function of the parameters m and k (Rao et al., 1997; Solazzo et al., 2016). The perfect separation of spectral components is only possible if a step function is a transfer function in a spectral decomposition technique. This happens only when passing the frequencies below the cutoff frequency, i.e. H=1, and attenuating all frequencies above that, i.e. H=0 (Eskridge et al., 1997).

The dependency of the spectral components can be checked with the definition of the total variance of *ORG(t)*:

$$\sigma^2(ORG) = \sigma^2(ID) + \sigma^2(DU) + \sigma^2(SY) + \sigma^2(BL) + 2cov(ID,DU) + 2cov(ID,SY) + 2cov(ID,BL) + 2cov(DU,SY) + 2cov(DU,BL) + 2cov(SY,BL) \quad (3.15)$$

Here σ and cov are the variance of each spectral components and the covariance between them, respectively (Rao et al., 1997). In this study, the covariance terms between adjacent spectral components, i.e. *cov(ID, DU), cov(DU, SY),* and *cov(SY, BL)*, were summed up in cov_{adj}. In a similar way, the covariance of non-adjacent spectral components, i.e. *cov(ID, SY), cov(ID,BL),* and *cov(DU,BL)*, were summed up to $cov_{non\text{-}adj}$. Thus, the Eq. 3.15 was rewritten:

$$\sigma^2(O) = \sigma^2(ID) + \sigma^2(DU) + \sigma^2(SY) + \sigma^2(BL) + cov_{adj} + cov_{non-adj} \quad (3.16)$$

In this equation, higher σ indicates higher contribution of the variance of a spectral component to the total variance. A small covariance between components attests a better separation. Therefore, although in theory, the optimum value for m and k can be obtained by minimizing the sum of the covariance among spectral components ($cov_{adj} + cov_{non-adj}$), the influence the optimized *m* and *k* carry on this analysis is not known.

3.3 Model performance indicator (modified mean square error)

In this study, the mean square error (MSE) was used as a statistical metric to evaluate the model simulation performance. The MSE is defined as the squared mean of the difference between a simulated (X_{mod}) and observed variable (X_{obs}):

$$MSE = \frac{\sum_{i=1}^{n}(X_{mod\,i} - X_{obs\,i})^2}{n} \quad (3.17)$$

Here *n* is the length of the time series.

The *MSE* is equal to the sum of the variance and squared bias ($\bar{X}_{mod} - \bar{X}_{obs}$):

$$MSE = (\bar{X}_{mod} - \bar{X}_{obs})^2 + Var\,(X_{mod} - X_{obs}) \quad (3.18)$$

The *Var* in Eq.3.18 is the variance operator as:

$$Var\,(X_{mod} - X_{obs}) = var\,(X_{mod}) + var\,(X_{obs}) - 2\,cov\,(X_{mod}, X_{obs}) \quad (3.19)$$

By inserting Eq.3.19 into Eq.3.18, MSE can be expressed as:

$$MSE = (\bar{X}_{mod} - \bar{X}_{obs})^2 + var\,(X_{mod}) + var\,(X_{obs}) - 2\,cov\,(X_{mod}, X_{obs}) \quad (3.20)$$

Expanding on Eq.3.20, Theil (1961) derived the following equation for the MSE:

$$MSE = (\bar{X}_{mod} - \bar{X}_{obs})^2 + (\sigma_{Xmod} - \sigma_{Xobs})^2 + 2\,(1 - r)\,\sigma_{Xmod}\,\sigma_{Xobs} \quad (3.21)$$

Here σ_{Xmod} and σ_{Xobs} are the standard deviations of the simulated and observed variable, respectively, r shows the correlation between the simulated and

observed variable. Pindick and Rubinfeld (1998) defined Theil's coefficients from the ratio of three terms on the right hand side of Eq. 3.21. Murphy (1988) provided examples of scores obtainable from the terms in Eq. 4.21. Solazzo et al. (2016) modified the MSE as:

$$MSE = (\bar{X}_{mod} - \bar{X}_{obs})^2 + (\sigma_{Xmod} - r\,\sigma_{Xobs})^2 + \sigma_{Xobs}^2(1 - r^2) \quad (3.22)$$

Although the three last equations (Eq.3.20, Eq.3.21, and Eq.3.22) consist of the relevant model evaluation metrics (bias, variance, and correlation), only Eq.3.22 was used in this study and the terms in this equation were represented following Solazzo et al. (2016):

MSE = bias error + explained error + unexplained error (3.23)

Here the bias error shows the degree of trueness, i.e. how much the differences between the averages observed and simulated values is. Both explained error and unexplained error contain the variance error, i.e. the discrepancy in amplitude or phase between the variability of the simulated and observed values determining the precision of the model. In more details, the explained error reveals the variance error arising from the variability of the simulated variables not observed in measurements. The unexplained error contains the lack of observed variability in the model. This error consists of different sources arising from non-systematic errors such as linearization of nonlinear processes, turbulence closure, and representativeness.

3.4 Multiple linear regressions analysis

One of the most widely used methods to quantify the relationship between O_3 concentration and meteorological or chemical variables are the multiple linear regressions (MLR) (Tarasova et al., 2003; Abdul-Wahab et al., 2005; Kovac-Andric et al., 2009; Rasmussen et al., 2012). The MLR is used to estimate a linear relationship between one dependent (response, predictant or regressant) variable and multiple independent (explanatory, predictor, and regressor) variables. One of the main advantages of this method is the simple interpretation of the relationship.

3.4.1 Multiple linear regressions model

The relationship between the dependent variables with the independent ones in MLR is estimated by fitting a linear equation to the observed data. Equation 3.24 is a MLR model with two predictors x_1 and x_2, that describes a plane in 3D space y-x_1-x_2 (Figure 3.3):

$$Y_i = a_0 + a_{1,i}\,x_{1,i} + a_{2,i}\,x_{2,i} + \varepsilon_i \qquad i = 1, \ldots, n \quad (3.24)$$

Here n is the number of sample, dependent variable, a_0 is the intercept of the plane, a_1 and a_2 are the slopes of the (hyper-)plane in the direction of the x_1 and x_2, respectively, which is estimated via least square error, the minimum distance between fitted line (plane) and the response variable. The a_1 and a_2 are the partial

regression coefficients. In theory, a_1 represents the average change in y corresponding to one unit change in x1 while x2 is constant. In a similar way, a2 shows the average change in y corresponding to the unit change in x2 when holing x1 constant. Thus, the higher the ai, the larger would be the relationship between Y and xi. ε is the residual which contains the corresponding random error, i.e. distance between true value and estimated response.

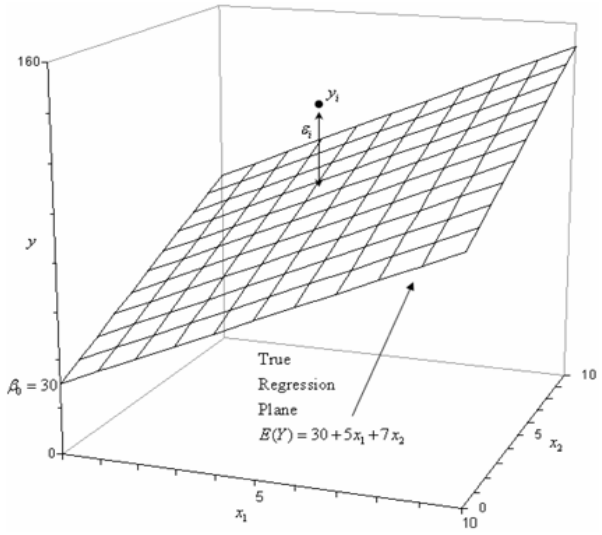

Figure 3.3. A MLR with two predictors x_1 and x_2, (http://reliawiki.org/index.php/Multiple_Linear_Regression_Analysis, last access 24.04.2018).

The MLR for k predictors is extended as:

$$y_i = a_0 + a_{1,i}\, x_{1,I} + a_{2,i}\, x_{2,I} + \ldots + a_{k,i}\, x_{k,i} + \varepsilon_i \qquad (3.25)$$

To quantify the relationship between O_3^{surf} mixing ratio and other variables such as AT, RH, SP, U, V, and NO_x (i.e. the usual proxies for the regression model of the O_3^{surf} (Bloomnfield et al., 1996; Roemer et al., 2001; Tarasova et al., 2003)), a MLR model (Eq. 3.26) was defined as:

$$O_3^{surf}{}_j\,(t_i) = a_{0,i} + a_{1,j}\,AT(t_i) + a_{2,j}\,RH(t_i) + a_{3,j}\,SP(t_i) + a_{4,j}\,U(t_i) + a_{5,j}\,V(t_i) + a_{6,j}\,NO_x(t_i) \qquad i = 1, \ldots, n \qquad j = 1, \ldots, p \qquad (3.26)$$

Here n and p show length of the time series and the stations' number, respectively. In Eq. 3.26, O_3^{surf} is considered as a dependent variable predicting from AT, RH, AP, U, V and NO_x at each station. Accordingly, $a_{1,j}$ shows the regression coefficient indicating the O_3^{surf} concentration changes resulting from a unit AT change at a specific sites, holding other variables constant. Likewise, $a_{2,j}$, $a_{3,j}$, $a_{4,j}$, $a_{5,j}$, and $a_{6,j}$ show the regression coefficient indicating the O_3^{surf} concentration changes resulting from a unit change of RH, SP, U, V, and NO_x at a specific sites, respectively, holding other variables constant. Therefore, the high regression coefficient ($a_{k,j}$) indicates a significant effect of the corresponding variable on the O_3^{surf} level.

3.4.2 Data standardization

A MLR model is estimated based on few assumptions such as linear relationship between independent variables and dependent variables, absence of multicollinearity between predictors, normal distribution of residuals (i.e. predicted minus observed values) , choice of the number of predictors, etc. (StatSoft Inc., 2013). These assumptions need to be taken into account to obtain an appropriate MLR model. However the linearity assumption can virtually never be confirmed, the MLR procedures are not greatly affected by minor deviations from this assumption (StatSoft Inc., 2013). One approach to reduce multicollinearity between predictors is standardization. In this study, Z-standardization (a linear transformation) applied to the entire variables as:

$$V_i = \frac{v_i - \bar{v}}{\sigma} \qquad i= 1, \ldots, n \qquad (3.27)$$

Here v, $\bar{v}$ and σ are the variable, its average and standard deviation, respectively. The Z-standardization changes variables so that their average and standard deviation become zero and one, respectively.

As there is a correlation among the variables in this study, in particular among meteorological variables, the spectral components of each variable were first standardized and then used as input in the MLR model. The standardization also helped the determination of the relative importance of the standardized regression coefficients**.**

3.4.3 Coefficient of multiple determinations (R^2)

There are several statistical tools to quantify the accuracy of a MLR model. One of the most efficient statistical parameter is the coefficient of multiple determinations (R^2) which indicates the amount of explained variability by the MLR model:

$$R^2 = \frac{SS_R}{SS_T} = 1 - \frac{SS_E}{SS_T} \qquad (3.28)$$

Here SS_R is the sum of the square deviation between predicted value (y_i) (predicted by MLR model) and averaged actual value of the response variable ($\bar{Y}$):

$$SS_R = \sum_{i=1}^{n} (y_i - \bar{Y})^2 \qquad (3.29)$$

SS_T is the total sum of the square of the difference between the actual value (Y_i) and the average actual value of the response variable:

$$SS_T = \sum_{i=1}^{n} (Y_i - \bar{Y})^2 \qquad (3.30)$$

SS_E is the sum of the error (residuals) of the square of the difference between predicted and actual value of the response variable:

$$SS_E = \sum_{i=1}^{n} (y_i - Y_i)^2 \qquad (3.31)$$

The relationship between SS_R, SS_E, and SS_T is defined as:

$$SS_T = SS_R + SS_E \quad (3.32)$$

As adding more predictors increases the explained variance (SS_R), using only R^2 might lead to a misinterpretation of the MLR model. In a specific model, two predictors might be enough to explain a large part of the variability of the dependent variable. Adding more predictors to the MLR model might lead to an over-fitting model. This is literally means that the regression model becomes tailored to fit the quirks and random noise in your specific sample rather than reflecting the overall population. That can produce misleadingly high R^2 values (Minitab Inc., 2010) because R^2 assumes that every single predictor explains the variation in the dependent variable. To overcome this issue, a statistical quantity named adjusted R^2 (R^2_{adj}, it only increases if the new predictor actually affects the dependent variable) was defined:

$$R^2_{adj} = 1 - \frac{MS_E}{MS_T} \quad (3.33)$$

MS_E and MS_T are the average squared errors and the total squared average respectively:

$$MS_E = \frac{SS_E}{dof(SS_E)} \qquad dof(SS_E) = n - (k+1) \quad (3.34)$$

$$MS_T = \frac{SS_T}{dof(SS_T)} \qquad dof(SS_T) = n - 1 \quad (3.35)$$

Here *dof* shows the degrees of freedom. To quantify the amount of explained O_3^{surf} variance in the MLR model, R^2_{adj}(Eq. 3.33) was used in this study. Following, R^2_{adj} is simplified as R^2. Higher R^2 indicates that the MLR model explains more O_3^{surf} variability.

4 Results

In this chapter, first, results of the analysis are presented for two selected stations, one representative for the rural and one for the urban sites. These examples illustrate how the spectral decomposition and correlation analysis perform, and how different station characteristics influence the results. Thereafter, the results are discussed for all measurement sites in the Med region.

4.1 O_3^{surf} analysis at a rural site

As an example of a typical rural site, ASIAGO CIMA EKAR station (ACE) in Italy was chosen. This station is located at latitude 45.50°N and longitude 11.34°E, in the altitude 1366 *m*. Thus, it counts as a representative of the Central Med and Alpine region. The ACE station belongs to the background station type. First, the time series of the spectral components of the measured O_3^{surf} at the ACE station is presented. In the next step, the time series of the spectral components of the simulated O_3^{surf} are compared against the measurements. Then, the results of the evaluation of the model performance are shown for each spectral component. After that, the results of the MLR model are shown for each spectral component.

4.1.1 Time series of the spectral components

4.1.1.1 Measurement

As described in section 3.2., the measured O_3^{surf} time series were decomposed into four spectral components, namely ID, DU, SY, and BL, by using the KZ filter technique. Figure 4.1 shows hourly time series of O_3^{surf} mixing ratio and its spectral components at the ACE station, averaging over the summers of 2010, 2011, and 2012. The appendix contains additional figures for the individual years (Figures A1-A4). Figure 4.1.a shows the time series of O_3^{surf} mixing ratio before decomposition. The time series of the ID component is shown in panel (b). This component is made of very fast oscillations varying between -2 and 2 *ppb*, and a few high episodic events in which O_3^{surf} varies by 10 *ppb* within a very short time period, i.e. one hour or two. That might show the impact of irregular mesoscale phenomena such as convection on the O_3^{surf} variability. Panel (c) shows the time series of DU O_3^{surf} component varying by 15 *ppb*. Although the time series of the DU O_3^{surf} component oscillates fast, it is slower and more regular than the ID O_3^{surf}

oscillation in panel (b). The DU O_3^{surf} variability is associated with the DU variation of the solar flux and the resulting differences between daytime photochemical production and nigh time removal of O_3^{surf} (Hogrefe et al., 2000). On the other hand, the time series of the DU component in Figure 4.1.c includes a few episodic events with low O_3^{surf} sustaining for a few days. This might indicate either the influence of SY meteorological systems on DU O_3^{surf} variability or the presence of the SY O_3^{surf} variability within the DU component due to the imperfect separation of the O_3^{surf} spectral components. Panel (d) shows features of the SY O_3^{surf} variability, typically ranges from -10 to 10 *ppb*, lasting for several days. The BL part of O_3^{surf} at this station is given in panel (e). The behavior of this component is made of very slow oscillations changing gradually by 15 *ppb*. This is likely the average background O_3^{surf} gently increase in summer, i.e. seasonal cycle, at the ACE station, as one expects from typical summertime O_3^{surf} level in the Med region (see Chapter 1).

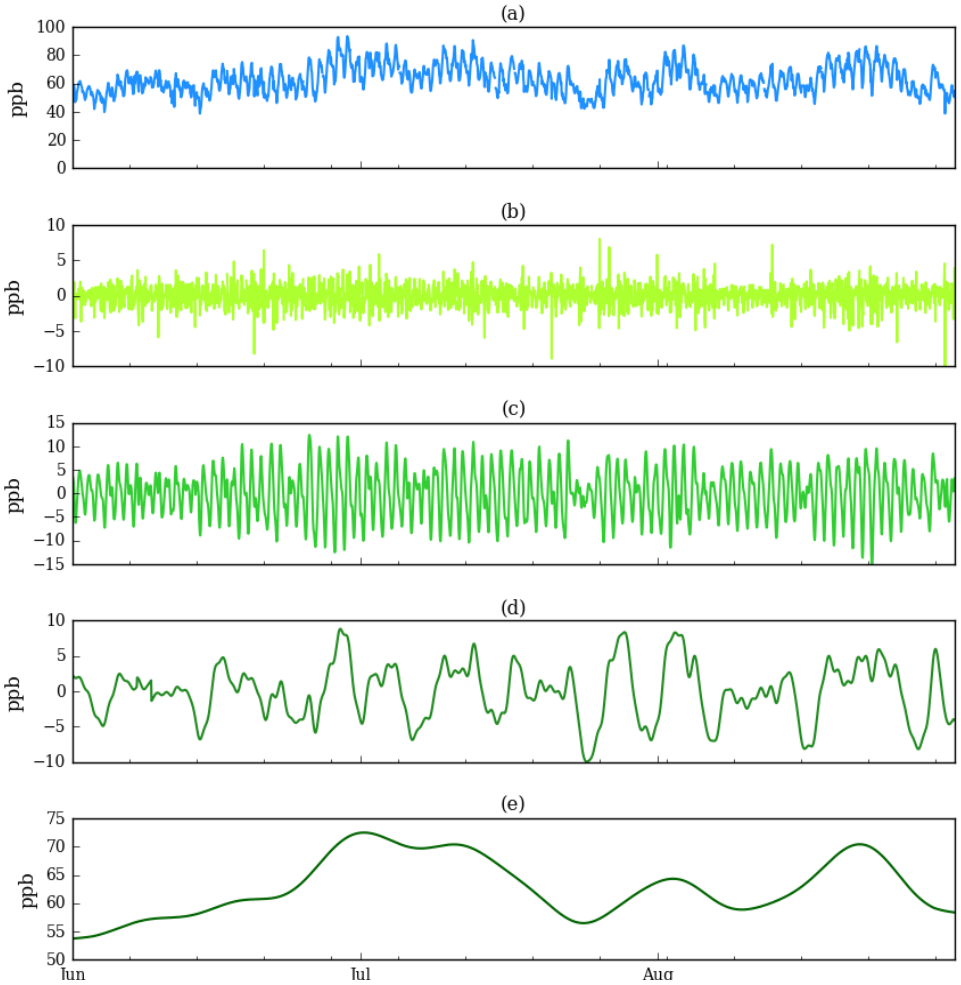

Figure 4.1. Hourly time series of the spectral components of the measured O_3^{surf} at the ACE station. The figure shows the time series of the **(a)** ORG, **(b)** ID, **(c)** DU, **(d)** SY, and **(e)** BL components. The values are averaged over the summers of 2010, 2011, and 2012.

4.1.1.2 The WRF-Chem model simulation

Time series of the simulated O_3^{surf} mixing ratio were decomposed to four spectral components by using the KZ filter technique. In the following, the results of the spectral components of the simulated O_3^{surf} versus those of the measurements at the ACE station are compared.

-The ID spectral component

Figure 4.2 shows the time series of the ID component of the simulated and measured O_3^{surf} mixing ratio at the ACE station. Since the ID component oscillates very fast, the first part of the time series for three days, i.e. 1st-3rd June 2010, is presented in this figure, to clearly see the difference between ID components of the model and observation. The time series of the ID component for the individual year are given in the Appendix (panel (a) in Figures A2-A4). Figure 4.2 shows that the ID component of the simulated O_3^{surf} oscillates fast and irregular like those of the observation, albeit it matches with the observed features only at some limited periods. In Figure 4.2, the mismatch between ID component of the simulated and measured O_3^{surf} appears not only in their amplitude, but also in their phase. For instance, in this figure, at early hours on 1st June, ID components of the simulated and measured O^{3surf} oscillates in a same way (in-phase, but with both similar and dissimilar amplitudes), while later it the ID component starts to rise in the model and to decline in the observation (anti-phase). That induces a phase differences in addition to the amplitude differences in the time series of the model and of the observation that both refer to the ''variance error''. Since ID covers the short-term fluctuations, the large variance error of the ID would indicate that the model simulation does not capture short-term events, which may be due to local influences such as land-sea breeze, vertical mixing height, etc. rather than regional-scale phenomena. Nevertheless, it would be an interesting analysis to look at regional consistency in the ID signals, but it is going beyond the scope of this thesis.

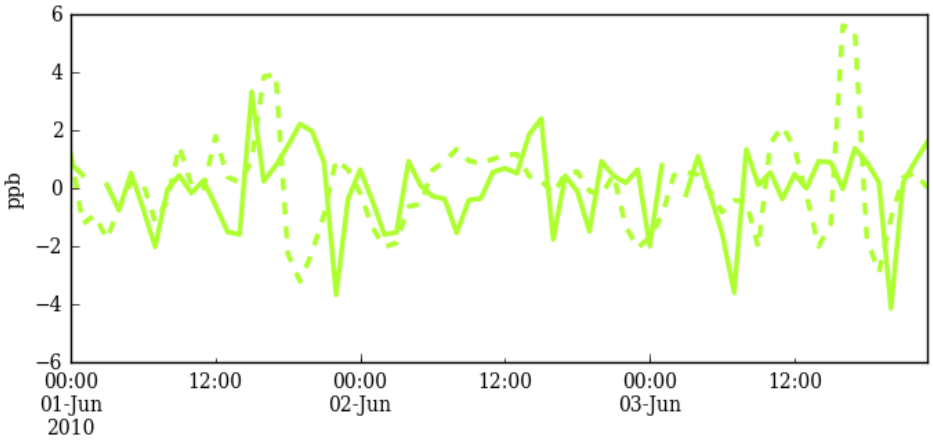

Figure 4.2. Hourly time series of the ID component of O_3^{surf} at the ACE station for 1st - 3rd June 2010. The solid and dashed lines show the measurement and the WRF-Chem model simulation, respectively.

-The DU spectral component

Time series of the DU variability of the measured and simulated O_3^{surf} mixing ratio at the ACE station are given in Figure 4.3. Similar to the ID component, for the DU, first part of the time series are shown in this figure and the time series of individual years are available in the Appendix (panel (b) in Figures A2-A4). From Figure 4.3, it appears that the simulated O_3^{surf} majorly captures the DU variability of the observed ozone. However, similar to the ID component, the amplitude and phase of the DU variability in the simulated O_3^{surf} are different from those in the measured O_3^{surf} indicating the model variance error in simulating the DU O_3^{surf} variability. A possible reason for this error might be related to the model's deficiency in the parameterizations of several phenomena such as convection, PBLH, relationship between PBLH and O_3^{surf} concentration.

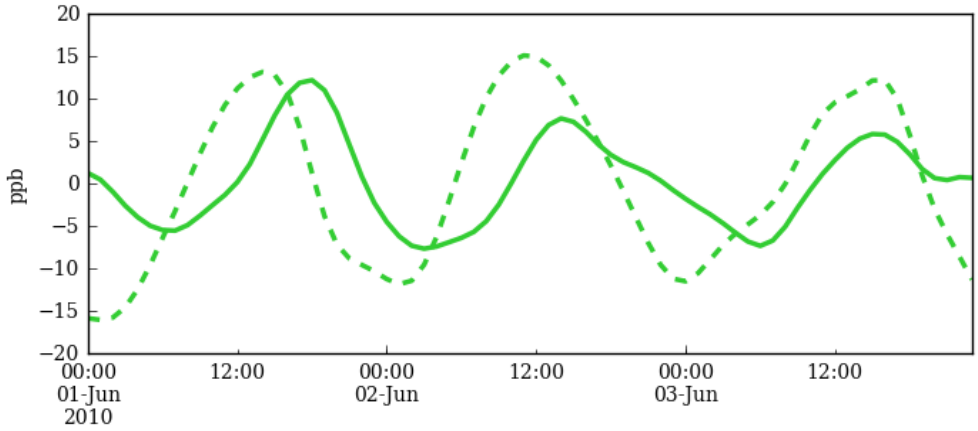

Figure 4.3. As in Figure 4.2, but for the DU component.

-The SY spectral component

Figure 4.4 shows the SY component of the observed and the simulated O_3^{surf} mixing ratio at the ACE station for summer 2010. Time series of this component for the individual years, i.e. summer 2011 and 2012, are given in the Appendix (panel (c) in Figures A3-A4). Figure 4.4 shows that, in similarity with the measured ozone, the simulated O_3^{surf} consists of SY variability. The SY component of the O_3^{surf} mixing ratio shows the medium-term O_3^{surf} fluctuations. In Figure 4.4, the variability of the SY component of the simulated O_3^{surf} appears in a different amplitudes and phases from those of the measurement, i.e. variance error. Similar to the DU component, one of the possible reasons for the variance error of the SY component of the simulated O_3^{surf} is weakness in the model parameterizations of some phenomena such as convection, turbulent exchange between PBL and free troposphere. Although, these processes and likely their influence on O_3^{surf} occur in the DU scale, the variation in the intensity of these phenomena triggers by the weather system, i.e. SY scale.

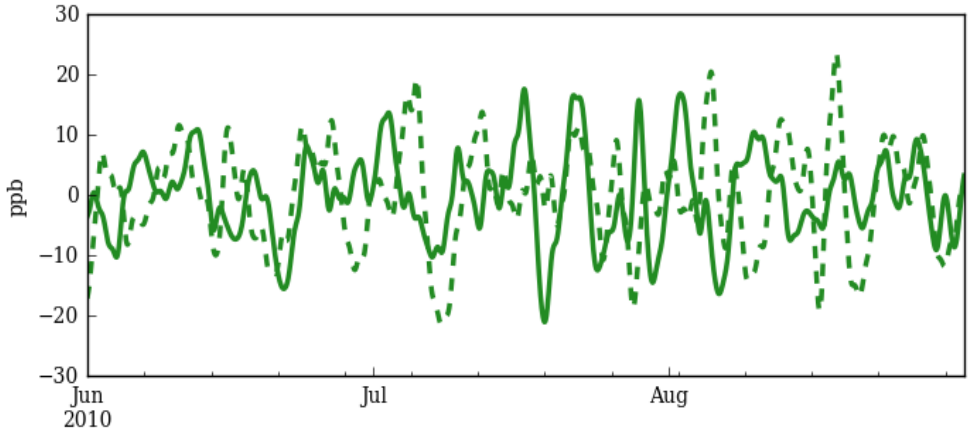

Figure 4.4. Hourly time series of the SY component of O_3^{surf} mixing ratio at the ACE station for summer 2010. The solid and dashed lines show the measurement and the WRF-Chem model simulation, respectively.

-The BL spectral component

Figure 4.5 shows the time series of the BL component of the simulated and measured O_3^{surf} mixing ratio in summer 2010 at the ACE station. The time series for 2011 and 2012 are given in the Appendix (panel (d) in Figures A3-A4). Figure 4.5 shows that at some periods the simulated O_3^{surf} matches with the measured O_3^{surf}, whereas at some periods it does not match. In Figure 4.4, the mismatch between the model simulation and observation appears as both variance and bias error. Possible reasons for this variance error might arise from the modeling deficiency in simulating long-range transports that is likely due to the error in model lateral boundary conditions, the lack of the emission variability in the model (see section 1.3). The bias error in Figure 4.5 appears as an underestimation of the BL component of the simulated O_3^{surf} in comparison to the measurement. A similar bias error, i.e. underestimation of the BL component of the simulated ozone, has been seen in the BL time series for summers 2011 and 2012 (panel (d) in Figures 3A-4A). Therefore, this might indicate that the bias model error arises from systematic errors in the input data such as O_3^{surf} precursors' emission that is likely due to the low model resolution, error in emission's estimation (see section 1.3).

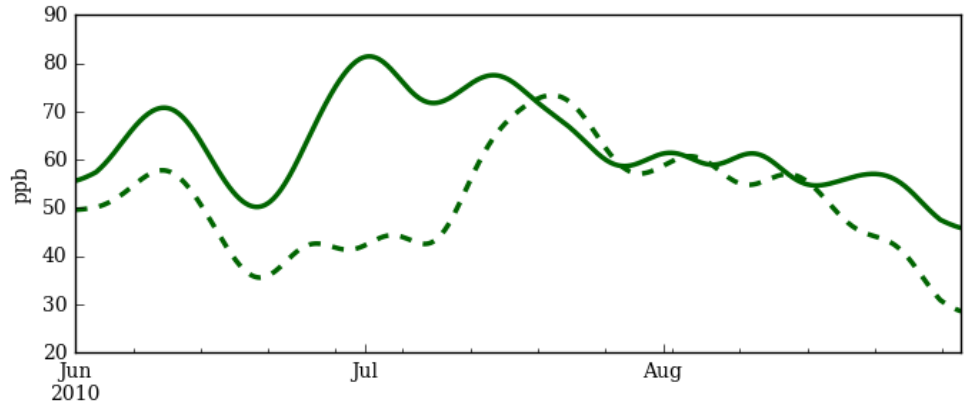

Figure 4.5. As in Figure 4.4, but for the BL component.

4.1.2 Model performance of the spectral components

First, the MSE of the simulated O_3^{surf} and its spectral components at the ACE station were calculated by using Eq. 3.17. The MSE of the individual spectral components, i.e. ID, DU, SY, and BL, of the simulated O_3^{surf} and their sum (TOT) at the ACE station are listed in Table 4.1 (see MSE_c row). The second row of Table 4.1, i.e. MSE_{cc}, shows the MSE of cross components such as ID-DU, ID-SY, ID-BL, DU-SY, DU-BL, and SY-BL. In Table 4.1, the MSE_{cc} for the ID component is 15.27 resulting from the sum of the MSEs of the ID-DU, ID-SY, and ID-B. Likewise, the MSE_{cc} for the DU component is 34.57 showing the sum of the DU-ID, DU-SY and DU-BL. The last row in Table 4.1, i.e. MSE, shows the sum of the MSE_c and MSE_{cc} for each component. Accordingly, the MSE of the original simulated O_3^{surf} time series, before decomposition, is 531.39 (see Table 4.1). The number within parenthesis in Table 4.1 shows the percentage contribution of that spectral component to the total model error. Accordingly, the ID, DU, SY and BL components accounts for 4.98%, 30.23%, 22.51%, and 42.43% of the model's MSE, respectively, (see Table 4.1). This implies that the significant part of the errors of the simulated O_3^{surf} at the ACE station is caused by the error of the BL, DU and SY components and only a negligible part of that is produced by the error of the ID component. Nevertheless, the reason for the small error of the ID term is likely because of the small magnitude of the ID component in comparison to other components, such as DU, SY, and BL.

Table 4.1. MSEs of the spectral components for the simulated O_3^{surf} at the ACE station. The values were averaged over the summers of 2010, 2011, and 2012. The values within the parenthesis show the percentage of the errors with respect to the MSE_{ORG} (= 531.39) at this site. The sum of the values in this table may not be matched, mathematically. That is because of using a dot product instead of a cross product among the spectral components.

spectral components	ID	DU	SY	BL	TOT
MSE_c	11.19 (2.1%)	126.09 (23.72%)	85.38 (16.06%)	211.8 (39.85%)	434.47 (81.76%)
MSE_{cc}	15.27 (2.87%)	34.57 (6.5%)	34.23 (6.44%)	13.68 (2.57%)	96.92 (18.24%)
MSE	26.46 (4.98%)	160.67 (30.23%)	119.62 (22.51%)	225.48 (42.43%)	531.39 (100 %)

In the next step, the portions of the bias and variance errors in the MSE_c were calculated by using Eq. 3.18. The bias error, as discussed earlier, shows the systematic differences between simulated and measured ozone, whereas the variance error reflects the discrepancy between the variability of simulated and measured O_3^{surf}, e.g. incorrect amplitude or phase of the oscillation. The bias and variance errors of the spectral components of the simulated O_3^{surf} at the ACE station are given in Table 4.2. Table 4.2 shows, the largest bias error, 130.95, belongs to the BL component making 24.64 % of the model MSE. Unlike the BL

with a very large bias error, other components such as ID, DU, and SY contain a negligible bias error. The reason for that is in the separation technique, these three components are zero mean fluctuations around BL. Thus, the ID, DU, and SY components are unbiased and only contain the variance error. The largest variance error, 126.065, belongs to the DU. The variance error of the SY (= 85.145) and BL (= 80.859) components are moderately high. The ID holds a small variance error, 11.09, comparing to the variance error of other components. Having already mentioned the insignificant variance error of the ID is mainly due to the small magnitude of the ID in comparison with other spectral components. In total, at the ACE station, nearly 57.05% and 24.65% of the model MSE has been produced by the variance and bias errors of the spectral components, respectively. In the next step, the variance error, which is the larger contributor to the MSE of the model, was only assessed.

Table 4.2. The MSE_c, bias and variance error of the spectral components for the simulated O_3^{surf} for the ACE stations. The values in the parenthesis show the percentage of the errors with respect to the MSE_{ORG} (= 531.39) at this site.

spectral components	ID	DU	SY	BL	SUM
MSE_c	11.19 (2.1%)	126.09 (23.72%)	85.38 (16.06%)	211.8 (39.85%)	434.47 (81.76%)
bias	0.00 (≈ 0%)	0.001 (≈ 0%)	0.028 (≈ 0%)	130.95 (24.64%)	130.98 (24.65%)
variance error	11.09 (2.08%)	126.06 (23.72%)	85.14 (16.02%)	80.86 (15.21%)	303.15 (57.05%)

The apportionments of the variance error, i.e. explained and unexplained errors, at the ACE station were calculated by using Eq. 3.22 (or Eq. 3.23) and shown in Table 4.3. Since the time period of our data may not be long enough to have a robust BL variability, the analyses in this thesis mainly focus on the short term components, i.e. ID, DU, and SY. The explained error shows the portion of the spurious variability of the model, while the unexplained error indicates the variability of the observation that is not explained in the model. According to the result of the analysis, the explained error of the DU component is 90.30 holding the largest contribution to the model variance error (= 222.15) in comparison to other components (see Table 4.3). This implies that the model produces some DU O_3^{surf} variabilities that do not exist in the observation. The unexplained error of the DU component is 35.76 contributing to 16.1% of the model variance that is lower than its explained error (see Table 5.3) showing the major features of the observed DU O_3^{surf} variability are captured by the model. On the other hand, the unexplained errors of the ID and SY in Table 4.3 are larger than their explained errors and contribute to 2.8% and 23.66% of the model variance error, respectively, reflecting the observed ID and SY variability which are not captured

by the model. As mentioned in the previous section, this can be explained by the model deficiency in simulating the meso-scale and synoptic phenomena.

Table 4.3. The apportionments of the model variance error at the ACE station. The values in the parenthesis show the percentage of the errors with respect to the sum variance error (= 222.15) at this site.

spectral components	ID	DU	SY	SUM
variance error	11.09 (5%)	126.06 (56.74%)	85.14 (38.32%)	222.15 (100%)
explained error	4.94 (2.2%)	90.30 (40.5%)	32.56 (14.72%)	127.66 (57.47%)
unexplained error	6.15 (2.8%)	35.76 (16.1%)	52.58 (23.66%)	94.48 (42.53%)

4.1.3 Statistical modeling of the O_3^{surf} variability

4.1.3.1 Regression analysis for the spectral components

The variabilities of the spectral components, i.e. ID, DU, and SY, of the measured and the simulated O_3^{surf} were estimated (modeled) by using a MLR model (Eq. 3.26). Even though the distribution of NO_x mixing ratios is a log-normal, using logarithm of NO_x and NO_x as a predictor in the MLR model did not generate much differences in the regression analysis results. Table 4.4 lists the results of the MLR model for the spectral components at the ACE station. From the results in Table 4.4 it appears that R^2 for the ID spectral component of the measured O_3^{surf} is 0.01, while that is 0.61 for the WRF-Chem model simulation. This shows that the MLR model explains more ID O_3^{surf} variability in the model simulation than in the measurements. Results of the regression coefficients in Table 4.4 indicate a low difference between the observation and the model simulation for most of the regression coefficients, i.e. a_1, a_2, a_3, a_4, and a_5, for the ID component; however, there is a large difference between a_6, i.e. NO_x as a predictor, in the model simulation and that in the observation. In the model simulation a_6 is -0.72 showing that one unit change in the ID spectral component of NO_x (= 1.82 *ppb*, see Table A1) leads to a reduction of 0.72 of a unit ID spectral component of O_3^{surf} (=2.25 *ppb*, see Table A1), holding other variable constant. However, one unit change in the ID spectral component of measured NO_x (=0.4 *ppb*, see Table A1) leads to an increase of 0.03 of a unit ID spectral component of the measured O_3^{surf} (=2.48 *ppb*, see Table A1). In other word, increasing 1 *ppb* NO_x in the measurements and the WRF-Chem model leads to the O_3^{surf} increase by 0.22 *ppb* in the observation and O_3^{surf} reduction by 0.89 *ppb* in the model simulation, respectively (see Table A4.5). This indicates a stronger relationship between ID variability of NO_x and O_3^{surf} in the WRF-Chem model than the observation.

Similar to the R^2 for the ID spectral component, the R^2 for the DU spectral component of the simulated data is larger than that of the measurements at the

ACE station (Table 4.4). This indicates that in the WRF-Chem model simulation, the ID and DU O_3^{surf} variability control by the variability of the meteorological variables and in particular, NO_x, whereas in the measurements approximately one third of O_3^{surf} variability explains via the variability of these variables and the remaining two third generates by other variables rather than meteorological variables and NO_x. As seen in Figure 4.3, the differences between DU variability of the simulated and measured O_3^{surf} ending up to conclude that, a few variables apart from AT, RH, SP, U, V, and NO_x, controls the DU variability of the O_3^{surf}.

Unlike R^2 for the ID and DU spectral components, R^2 for the SY spectral component for both datasets is nearly identical, ≈ 0.85 (see Table 4.4). It means that variability of the predictors is able to explain 85% of the SY O_3^{surf} variability in the model simulation and in the measurement dataset. The regression coefficients for the SY component in Table 4.4 indicates that the SY O_3^{surf} variability in both datasets is majorly explained by the variability of AT and NO_x. Nevertheless, there are some discrepancies between regression coefficients in the simulated and observed datasets. For instance, in the model simulation, there is a correlation between SY variability of O_3^{surf} and RH (a_2=0.22 in Table 4.4), while that is not observed in the measurements (a_2=-0.02 in Table 4.4). That might be possible sources of the explained model errors of the SY O_3^{surf} components (see Table 4.3).

Table 4.4. The regression coefficients at the ACE station. The a_1, a_2, a_3, a_4, a_5, and a_6 shows when AT, RH, SP, U, V, and NO_x used as predictors in the MLR, respectively, for the standardized spectral components of the observations and the model simulation datasets.

datasets	the observations			the WRF-Chem model simulation		
spectral components	**ID**	**DU**	**SY**	**ID**	**DU**	**SY**
R^2	0.01	0.31	0.86	0.61	0.86	0.88
a_1	-0.03	-0.27	0.52	-0.12	0.22	0.66
a_2	-0.04	-0.08	-0.02	-0.12	0.07	0.22
a_3	-0.07	-0.11	-0.04	-0.09	-0.16	-0.16
a_4	0.04	0.17	-0.01	0.13	0.04	0.05
a_5	-0.01	0.11	0.16	0.07	0.20	0.23
a_6	0.03	0.46	0.36	-0.72	-0.54	-0.51

- Summary of the section 4.1

What do we learn from the analysis of the observed and simulated O_3^{surf} for the ACE station?

Analysis of the spectral components of the measured O_3^{surf} at the ACE station revealed that the O_3^{surf} time series consists of very fast and irregular ID, regular DU cycles, gradual SY variability, and very slow BL oscillations. Moreover, the ID

component of O_3^{surf} alters by ±4 *ppb* in a day, while the DU varies by ≈ ±15 *ppb*. Depending on a weather system, the O_3^{surf} level varies by ≈ ±10 *ppb* in a SY scale, ≈ ±2 *ppb* per day. Thus, the major variability of O_3^{surf} mixing ratio occurs in a daily time-scale that results from the solar flux changes influencing photochemical reactions. The meso-scale phenomena slightly perturb the DU cycle in a very short time scale, in both positive (increase) and negative (decrease) directions of the DU variability. The variation of the weather SY system modifies the intensity of these phenomena and probably the DU O_3^{surf} variability.

The analysis of the spectral components of the simulated O_3^{surf} mixing ratio showed that the DU O_3^{surf} variability in the model simulation possess a similar DU shape (although in different phases) as that in the measurement, whereas the simulated ID and SY O_3^{surf} variability only at some time periods holds the same pattern as the measurements.

The results of the model performance showed that the model simulation of O_3^{surf} consists of 75% and 25% variance and bias error, respectively. Moreover, the whole model bias error holds by the BL components, whereas the variance error generates by the entire spectral components of ozone. The largest model variance error in this analysis was associated to the DU component because of the differences in the phase of the DU O_3^{surf} variability in the simulation and observation. The results of the variance error apportionment revealed that the DU error majorly has been arisen from the explained model error indicating the DU O_3^{surf} variability in the model simulation that does not match with the observation. That might generate from the inherent model error, due to the model deficiencies in simulating some phenomena and their influence on the O_3^{surf} concentration such as PBL and the exchange between free troposphere and PBL, or from the input model data such as emissions, and meteorological data. Moreover, this analysis showed that the ID and SY model variance error majorly are associated to the unexplained model error reflecting that the model is not able to simulate all the observed ID and SY O_3^{surf} variability at this staion.

The result of the regression analysis of both standardized datasets showed that the MLR model is able to explain more O_3^{surf} variability in the model simulation than the observations dataset. This demonstrates that the O_3^{surf} variability in the model simulation majorly produces by the variability of the predictors, such as AT, RH, PS, U, V, and NO_x; however in reality other variables or phenomena might play a role. That might be a possible source for the unexplained model error. The regression coefficients of both datasets showed that in the model simulation, there is strong anti-relationship between O_3^{surf} and NO_x for three spectral components, while that is not observed in reality. That might be possible sources for the explained model error.

4.2 Differences of the O_3^{surf} analysis at one urban site

As an example, one typical urban site, ALCORCÓN 2 station (ALC) in Spain was chosen. This station is located at latitude 40.20°N and longitude 3.50°W at the altitude of 693 *m*. The ALC station accounts as a representative site of the Western Med region and belongs to the background urban station type. In this section, first, the time series of the spectral component of the simulated and observed O_3^{surf} at this station are compared with those at the ACE station. Then, the variability of the spectral components and model performance of the simulated O_3^{surf} for the ALC site are presented.

Figure 4.6 shows the time series of the ID component of the observed and simulated O_3^{surf} at this station and those at the ACE. This figure shows that both model simulation and observation contain a larger ID fluctuation at this station in comparison with the ID components at the ACE station. In Figure 4.6, the measured O_3^{surf} mixing ratio at the ALC station drops in early morning and late afternoon. This likely indicates that the NO titration within the traffic hours at this station. A similar oscillation in the model appears a few hours later than the time of those oscillations in the measurements. That might be because of the NO titration at the urban sites in the model too. However, the NO titration is less pronounced in the ID time series of the simulated and measured O_3^{surf} at the ACE station (see Figure 4.6).

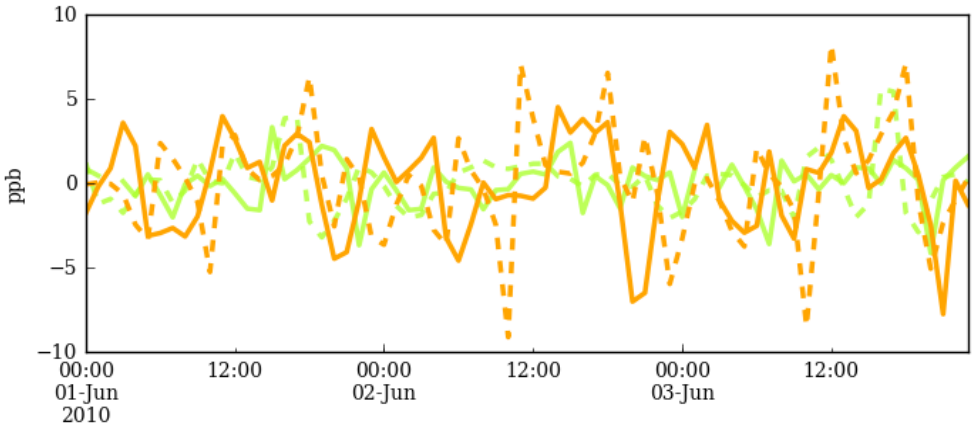

Figure 4.6. Hourly time series of the ID spectral component of O_3^{surf}at the ALC and ACE stations. The orange and green-yellow colors represent the data for ACL and ACE stations, respectively. The solid and dashed show the measurement and the WRF-Chem model simulation dataset, respectively.

Figure 4.7 shows the DU component of the observed and simulated O_3^{surf} at the ALC and ACE stations. Here, it appears that the amplitude of the DU component of both simulated and measured O_3^{surf} mixing ratio at the ALC site is larger than those at the ACE site. Figure 4.7 reveals a bi-modal DU component of the measured O_3^{surf} at the ALC station, which does not appear for the ACE station. This stems from the NO titration at the ALC station, as the bi-modal feature does not appear in the O_x (O_3^{surf} + NO_2) time series (see Figure E3). In addition, Figure

4.7 shows that the model simulation does not capture DU bi-modality of the O_3^{surf} at the ALC station. That might be related to the inadequate mixing of the boundary layer air with the residual layer aloft in the model simulation.

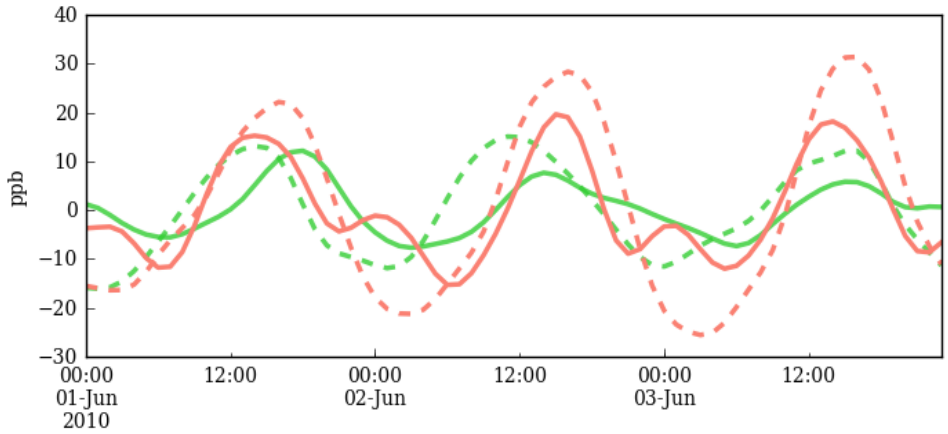

Figure 4.7. As in Figure 4.6, but for the DU component. The salmon and lime-green colors represent the data for ACL and ACE stations, respectively.

Likewise, Figure 4.8 shows the SY component of the simulated and measured O_3^{surf} mixing ratio at both ALC and ACE stations containing irregular variation, in which the oscillations at the ACE station is larger than those of the ALC station.

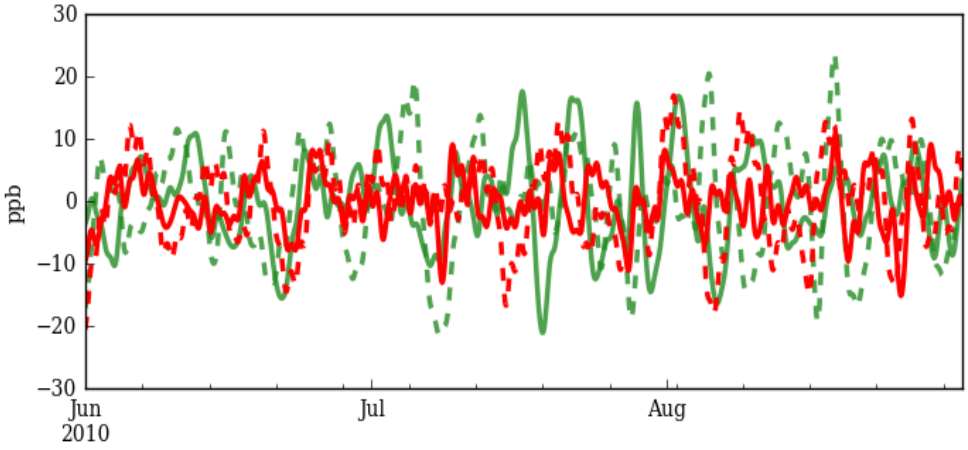

Figure 4.8. As in Figure 4.6, but for the SY component. The red and forest-green colors represent the data for ACL and ACE stations, respectively.

Figure 4.9 shows the lower mixing ratio for the BL component of the measured O_3^{surf} at the ALC in comparison to that at the ACE site reflecting the O_3^{surf} sink due to the NO_x titration process in urban sites. Moreover, a similar feature appears in the model simulation data (see Figure 4.9), although it is lower significant than that in the observation. Apparently, the BL component of the simulated and measured O_3^{surf} mixing ratio at the ALC station are closer in comparison to those at the ACE station, showing lower model bias in simulating O_3^{surf} at the ALC site in comparison to the ACE site.

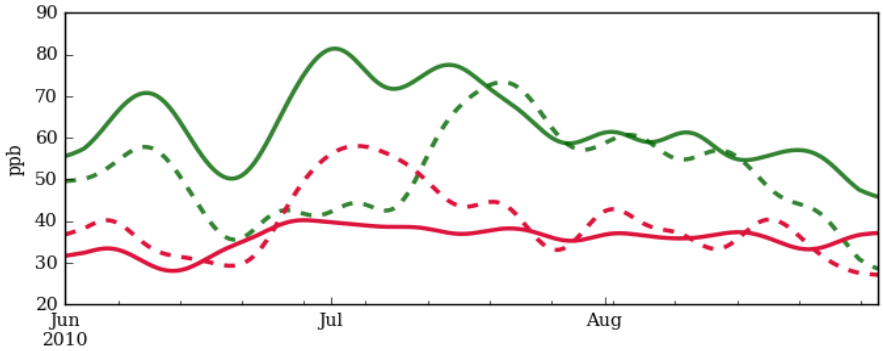

Figure 4.9. As in Figure 4.6, but for the BL component. The crimson and dark-green colors represent the data for ACL and ACE stations, respectively.

The MSE, MSE_c and MSE_{cc} of the spectral components of the simulated O_3^{surf} for the ALC station are listed in Table 4.6. Similar to the MSEs for the ACE station, nearly 75% and 25% of the MSE for the ALC station are covered by the sum of the MSEc (=211.28) and MSE_{cc} (= 80.36), respectively. The lower MSE of the model at the ALC station in comparison to the MSE of the simulated O_3^{surf} for the ACE station (=531.39) reflect a better model performance at the ALC than the ACE site.

Table 4.6. As in Table 4.1, but for the ALC station. The MSE_{ORG} at this station is 291.64.

spectral components	ID	DU	SY	BL	TOT
MSE_c	16.36 (5.62%)	101.44 (34.85%)	49.85 (17.12%)	43.63 (14.99%)	211.28 (72.58%)
MSE_{cc}	17.61 (6.05%)	28.98 (9.95%)	23.10 (7.94%)	10.65 (3.66%)	80.36 (27.55%)
MSE	33.97 (11.67%)	130.42 (44.8%)	72.96 (25.06%)	54.29 (18.65%)	291.64 (100%)

The MSE_c, variance and bias errors of the spectral components of the simulated O_3^{surf} for both ALC and ACE stations are shown in Table 4.7. According to the results of the analysis, unlike the lower MSE_c for the ALC station in comparison to that for the ACE site, the MSE for the ID component at the ALC station (=16.33) is larger than that for the ACE site (=11.19). The reason for that appears in the larger unexplained error of the ID O_3^{surf} at the ALC station (=12) in comparison to that at the ACE station (=6.15) (given in Table 4.8).

On the other hand, the lower MSE_c for other spectral components, i.e. DU, SY, and BL, at the ALC station with respect to those at the ACE station in Table 4.7 might explain the better model performance in simulating the O_3^{surf}, in particular the BL, at the ALC in comparison to the ACE staion. Unlike the largest contribution of the BL into the MSE_c at the ACE station (39.85%), this component only contributes to 14.97% of the MSE_c at the ALC station. As seen the closeness of the BL component

of the observed and simulated O_3^{surf} at the ALC station in the Figure 4.9, and considering the fact that the BL component holds bias and variance errors, the model at the ALC site contains lower bias (=0.054) and variance error (=43.49) for the BL in comparison to those for the ACE station (= 130.95 and 80.86 for the bias and variance errors, respectively). Thus, the model contains lower bias error in simulating the O_3^{surf} for the ALC (urban) station in comparison to that for the ACE (rural) site, explaining the better model performance at the ALC station.

Table 4.7 shows unlike the lower variance error of the DU component for the ALC station (=101.48) in comparison to the ACE station (=126.06), this component (i.e. DU) generates the significant portion of the variance error at both stations (34.86% and 23.72% for the ALC and ACE stations, respectively). This shows that the major source of the model variance error is generated by the DU component at both sites. Moreover, the larger contribution of the DU component to the variance error of the simulated O_3^{surf} for the ALC station (34.86%) in comparison to that for the ACE station (23.72%) reflects the significance (importance) of the DU variance error at the urban site. In comparison to the ACE station, the simulated O_3^{surf} at the ALC station contains larger unexplained and smaller explained errors. This demonstrates that the model simulation is not able to capture all DU variabilities of the measured O_3^{surf} at the ALC site. As seen in Figure 4.7, the presence of bi-modal DU variability in the measured O_3^{surf} that was not appeared in the simulated O_3^{surf}.

Table 4.7. As in Table 4.2, but for the ALC and ACE stations. The values in the parenthesis show the percentage of the errors with respect to the MSE_{ORG} (=291.64 and 531.39) at these sites, respectively.

spectral components		ID	DU	SY	BL
ALC	**MSE_c**	16.36 (5.62%)	101.44 (34.85%)	49.85 (17.12%)	43.63 (14.99%)
	bias	0.00 (≈ 0%)	0.00 (≈ 0%)	0.00 (≈ 0%)	0.05 (≈ 0%)
	variance error	16.36 (5.62%)	101.48 (34.86%)	49.85 (17.12%)	43.59 (14.97%)
ACE	**MSE_c**	11.19 (2.1%)	126.09 (23.72%)	85.38 (16.06%)	211.80 (39.85%)
	bias	0.00 (≈ 0%)	0.00 (≈ 0%)	0.03 (≈ 0%)	130.95 (24.64%)
	variance error	11.09 (2.08%)	126.06 (23.72%)	85.14 (16.02%)	80.86 (15.21%)

Table 4.8. As in Table 4.3, but for the ALC and ACE stations. The values within the parenthesis show the percentage of the errors with respect to the sum variance error, i.e. 167.70 and 222.15, for the ALC and ACE stations, respectively.

spectral components		ID	DU	SY	SUM
ALC	variance error	16.36 (9.76%)	101.48 (60.51%)	49.85 (29.73%)	167.70 (100%)
	explained error	4.35 (2.6%)	58.57 (34.93)	31.85 (18.99%)	94.78 (56.52%)
	unexplained error	12 (7.16%)	42.91 (25.59%)	18.00 (10.74%)	72.92 (43.48%)
ACE	variance error	11.09 (5%)	126.06 (56.74%)	85.14 (38.32%)	222.15 (100%)
	explained error	4.94 (2.2%)	90.30 (40.5%)	32.56 (14.72%)	127.66 (57.47%)
	unexplained error	6.15 (2.8%)	35.76 (16.1%)	52.58 (23.66%)	94.48 (42.53%)

Table 4.9. As Table 4.4, but for the ALC station.

datasets	the observations			the WRF-Chem model simulation		
spectral components	ID	DU	SY	ID	DU	SY
R^2	0.67	0.88	0.22	0.61	0.93	0.72
a_1	0.15	0.55	0.06	0.07	0.31	0.65
a_2	-0.03	-0.07	-0.21	-0.004	-0.009	0.23
a_3	-0.001	-0.009	0.06	-0.13	-0.04	0.04
a_4	0.03	0.008	-0.17	0.006	-0.01	0.2
a_5	0.02	0.06	0.14	-0.05	0.02	0.1
a_6	-0.74	-0.42	-0.59	-0.75	-0.63	-0.31

- Summary of section 4.2

What do we learn from the comparison of analysis between the ACE (rural) and the ALC (urban) stations?

Analysis of the spectral components of the measured O_3^{surf} mixing ratio at the ALC station revealed larger amplitude of the ID and DU components in comparison to

those at the ACE station. That might reflect a stronger photochemical activity at the ALC station in comparison to that at the ACE. The SY component of measured O_3^{surf} at the ALC site showed lower variability in comparison to that at the ACE station. The BL component of O_3^{surf} contained a lower mixing ratio at the ALC station in comparison to that at the ACE station, reflecting the lower average O_3^{surf} background at urban sites than rural sites (from the O_3^{surf} sink due to the NO_x titration process in urban sites (see section 1.2.3).

The analysis of the spectral components of the simulated O_3^{surf} mixing ratio in this section showed that the model is able to capture the major observed feature in the O_3^{surf} variability at both sites. For instance, it shows the larger amplitude of the ID and DU components and the lower mixing ratio of the BL component of O_3^{surf} at the ALC in comparison to the ACE site that appeared in the analysis of the spectral components of the measured ozone. Nevertheless, this analysis showed some observed features of the measured O_3^{surf} at the ALC station (such as bi-modal DU component) that the model does not represent.

The MSE showed the better model performance (lower MSE) of simulated O_3^{surf} for the ALC site in comparison to that for the ACE site. That was majorly associated to the lower model bias in simulating O_3^{surf} for the ALC site in comparison to that for the ACE. Since model bias is introduced by both internal sources, such as model error such as linearity of non-linear process, omitted and unresolved variables and process, and external sources, such as error in input files for meteorology and emission data, boundary conditions, (Solazzo et al., 2016), the low model bias at the ALC station might arise from a lower uncertainty in the model and inputs for the urban site in comparison to the rural site. Moreover, the better model performance at the ALC station can be associated to the lower model variance error of the DU, SY and BL components at the ALC in comparison to those at the ACE station showing the lower model error at the urban sites.

Nevertheless, the variance error of the ID component at the ALC station was larger than that for the ACE site showing more model difficulties in capturing the meso-scale phenomena (due to the missing resolution) at the urban site. A larger unexplained error (from the variance error apportionment) of the ID component at the ALC sites in comparison to that at the ACE station demonstrated some intraday observed O_3^{surf} variability that are missed in the model for simulating O_3^{surf} at the ALC (urban) station.

4.3 The O_3^{surf} analysis over the whole Med region

4.3.1 Variability of the spectral components

Table 4.11 lists variances of the measured and simulated O_3^{surf} fmixing ratio and their spectral components at the rural and urban sites over the Med region. From this table it appears that the variance of both measured O_3^{surf} mixing ratio at the urban sites (=230.92) is larger than that at the rural sites (=193.42). Moreover, the variance of the ID and DU components of the measured O_3^{surf} at the urban sites is larger than that at the rural sites, whereas the SY and BL variance at urban sites are lower than that at the rural sites (table 4.11). The WRF-Chem model shows similar features (e.g. lager ID and DU variability and lower SY and BL variability at the urban sites in comparison to those at the rural sites) as those in the measurement (table 4.11). Although the WRF-Chem model is in agreement with the measurements in showing some O_3^{surf} variability's features at the rural and the urban sites, it does not show a similar magnitude of simulated O_3^{surf} variability as those in measurements (table 4.11). For instance, the variances of the simulated O_3^{surf} are 322.82 and 337.67 at the rural and urban sites, respectively, while those in the observation are 193.42 and 230.92. Thus, O_3^{surf} variability is larger in the model than the measurements. The larger variability is not only appeared in the original simulated O_3^{surf} mixing ratio, but also in its spectral components (DU, SY, and BL). The only exception to it is the variances of the ID component, which are 4.35 and 5.66 in the model and 5.81 and 9.15 in the measurements at the rural and urban sites, respectively. As seen in the previous sections, an underestimating of the ID O_3^{surf} variability in the model arises from the model's difficulties in capturing meso-scale phenomena.

Table 4.11. Variance of the measured and the simulated O_3^{surf} mixing ratio and their spectral components averaged over the entire rural and urban sites over the Med region.

datasets	the observations			the WRF-Chem model simulation		
spectral components	ID	DU	SY	ID	DU	SY
rural	5.81	83.33	27.10	4.35	93.16	68.26
urban	9.15	113.6	25.91	5.66	132.78	56.67

4.3.2 The model performance of the spectral components

Table 4.12 lists the MSEs of the spectral components of the simulated O_3^{surf} averaging over the entire rural and urban stations of the Med region. Unlike the lower MSE at the ALC (an urban station) in comparison to the ACE (a rural station) in the previous section, table 4.12 shows the MSE at both rural and urban sites are nearly the same, i.e. 445.84 and 438.37 for the rural and urban sites, receptively. Moreover, at both

rural and urban sites, nearly 55%, 20%, 20% and 5% of the MSE of the simulated O_3^{surf} mixing ratio consist of the error in the BL, SY, DU and ID component, respectively. However, from table 4.13, the error of the BL component only contributes to 14.3% and 10.75% of the model variance error at rural and urban sites, respectively, and the rest (≈ 40%) contribute to the model bias error. Thus, the contribution of the BL component to the model variance error is lower than the DU and SY component at both sites.

On the other hand, from table 4.13, the variance error of the ID component only contributes to 2.2% and 2.88% of the total model error at the rural and urban sites, receptively. Likewise, the variance error of the DU components makes 13.17% and 15.23% of the total model error at rural and urban sites, respectively. Table 4.13 shows that the contribution of the SY variance model error to the total model error is 14.68% and 13.92% for the rural and urban sites, respectively. Therefore, first the variance error of the ID and DU components contain more contribution to the total model error at the urban sites in comparison to that at the rural sites. That might be as a result of the stronger O_3^{surf} photochemistry (ID and DU variability) at urban sites in comparison to rural sites. Secondly, the contribution of the variance error of the SY component at the rural sites is more than that at the urban sites.

Table 4.12. As in Table 4.1, but for all rural and urban sites The values within the parenthesis show the percentage of the errors with respect to the MSE_{ORG} (= 445.84 and 438.37) for the rural and urban sites, respectively.

spectral components		ID	DU	SY	BL	TOT
rural	**MSE_c**	9.02 (2.02%)	58.79 (13.18%)	65.49 (14.68%)	238.13 (53.38%)	371.45 (83.26%)
	MS_{cc}	9.82 (2.20%)	22.39 (5.01%)	27.64 (6.19%)	14.54 (3.26%)	74.39 (16.68%)
	MSE	18.84 (4.22%)	81.19 (18.2%)	93.13 (20.87%)	252.67 (56.64%)	445.84 (99.93%)
urban	**MSE_c**	12.62 (2.88%)	66.73 (15.23%)	61.00 (13.92%)	224.36 (51.22%)	364.72 (83.26%)
	MSEcc	12.36 (2.82%)	24.71 (5.64%)	24.63 (5.62%)	11.94 (2.72%)	73.64 (16.81%)
	MSE	24.98 (5.70%)	91.44 (20.87%)	85.63 (19.55%)	236.31 (53.94%)	438.37 (100%)

Table 4.13. As in Table 4.2, but for all rural and urban sites. The values within the parenthesis show the percentage of the errors with respect to the MSE_{ORG} (= 438.37 and 445.84) for the rural and urban sites, respectively.

spectral components		ID	DU	SY	BL
rural	MSE_c	9.02 (2.02%)	58.79 (13.18%)	65.49 (14.68%)	238.13 (53.38%)
	bias	0.00 (≈ 0%)	0.01 (≈ 0%)	0.01 (≈ 0%)	175.54 (39.35%)
	variance error	9.01 (2.02%)	58.74 (13.17%)	65.49 (14.68%)	62.60 (14.03)
urban	MSE_c	12.62 (2.88%)	66.73 (15.23%)	61 (13.92%)	224.36 (51.22%)
	bias	0.00 (≈ 0%)	0.01 (≈ 0%)	0.01 (≈ 0%)	177.25 (40.46%)
	variance error	12.61 (2.88%)	66.71 (15.23%)	61 (13.92%)	47.11 (10.75%)

The results in Table 4.14 shows that a larger portion of the ID variance model error consists of the unexplained model error at both sites, reflecting the model deficiency in capturing meso-scale phenomena due to the missing model resolution. Moreover, in table 4.14, contribution of the unexplained model error to the total model error at the urban sites (2.03%) is larger than that at rural sites (1.28%), showing the model weaknesses in simulating photochemical reactions in a very short time scale. In similarity with the ID component, the contribution of the variance error of the DU component to the total model error in both sites majorly consists of the unexplained error, which is larger at the urban (11.04%) than at the rural (8.92%) sites (given in Table 4.14). A larger portion of the model variance error of the SY component at both urban and rural sites arises from the explained error (table 4.14), which might arise from the strong model boundary forcing.

Table 4.14. As in Table 4.3, but for all rural and urban sites. The values within the parenthesis show the percentage of the errors with respect to the sum model variance error (= 133.27 and 140.32) for the rural and urban sites, respectively.

spectral components		ID	DU	SY	SUM
rural	variance error	9.01 (6.76%)	58.74 (44.09%)	65.49 (49.14%)	133.27 (100%)
	explained error	3.32 (2.49%)	18.97 (14.23%)	42.54 (31.92%)	64.83 (48.65%)
	unexplained error	5.7 (4.27%)	39.79 (29.86%)	22.94 (17.22%)	68.44 (51.35%)
urban	variance error	12.61 (8.99%)	66.71 (47.54%)	61 (43.47%)	140.32 (100%)
	explained error	3.71 (2.65%)	18.35 (13.07%)	38.12 (27.17%)	60.18 (42.89%)
	unexplained error	8.89 (6.34%)	48.36 (34.46%)	22.88 (16.3%)	80.14 (57.11%)

4.3.3 The coefficient of determination (R^2)

In this section, first, the results of the R^2 are given for three spectral components, ID, DU and SY, of both datasets at the rural stations over the Med region. Next, R^2 of the spectral components of at the urban sites are shown. Due to the diversity of R^2 at each station, the results are presented by using box plots. For the sake of simplicity, descriptions of the results are mostly based on the mean and median R^2. Nevertheless, the R^2 at each station is shown as a scatter point, next to the box plots.

Figure 4.10 shows the R^2 from the MLR model for the spectral components of both datasets at the rural sites. The comparison of the R^2 in Figure 4.10 showed that the largest and lowest R^2 attribute to the DU and ID components, respectively, for both datasets. This indicates that the MLR model explains more of the DU O_3^{surf} variability and less of the ID O_3^{surf} variability. The R^2 of the SY component is between the R^2 for the ID and DU for both datasets (Figure 4.10) reflecting that only a part of the SY O_3^{surf} variability is explained by the predictors, AT, RH, P, U, V, and NO_x. However, the lower R^2 for the entire spectral components of the measured dataset than the simulated data in Figure 4.10 shows a less correlation between O_3^{surf} and other variables, such as AT, RH, etc., in reality than the WRF-Chem model. Besides that, Figure 4.10 illustrates the presence of several outliers, with low R^2, in the DU component of both datasets. It shows a low correlation between DU O_3^{surf} variability and other variables at these stations. A possible reason for that might be a peculiar characteristic in the DU O_3^{surf} cycle at these sites or their geographical location.

In similarity with the rural sites, the largest R^2 belongs to the DU component of both datasets at the urban sites (Figure 4.11) indicating a high correlation between O_3^{surf} and the predictors at the urban sites. The R^2 of the ID component of the measured data at the urban sites is larger than the rural sites (see Figure 4.10.a and 4.11.a). It reflects a larger correlation between ID variability of the O_3^{surf} and the predictors at the urban sites than the rural sites. The similar feature appeared in the simulated datasets as well (see Figure 4.10.b and 4.11.b).

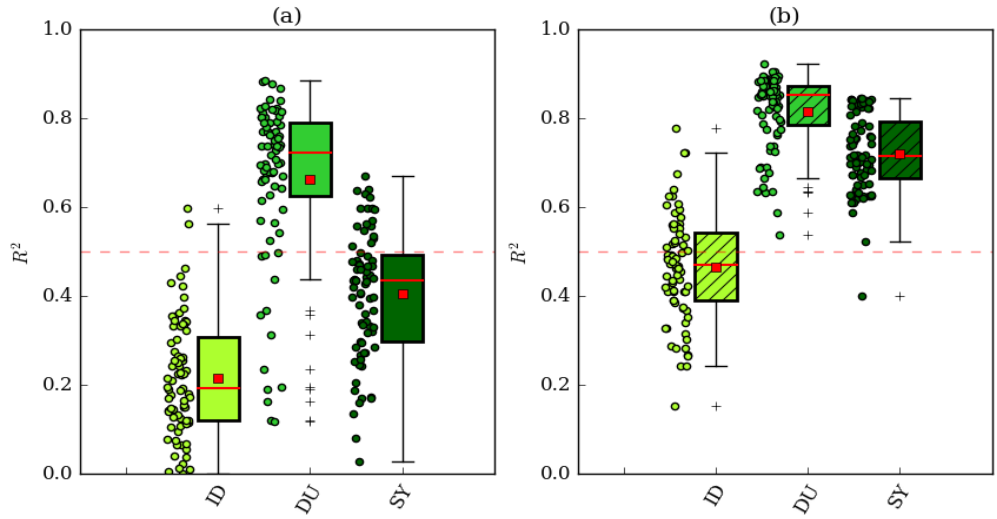

Figure 4.10. The R^2 from MLR model for ID, DU and SY components of **(a)** the measurements and **(b)** the simulated datasets, at the rural sites. The scatter point adjacent to the box plot presents the R^2 at each station.

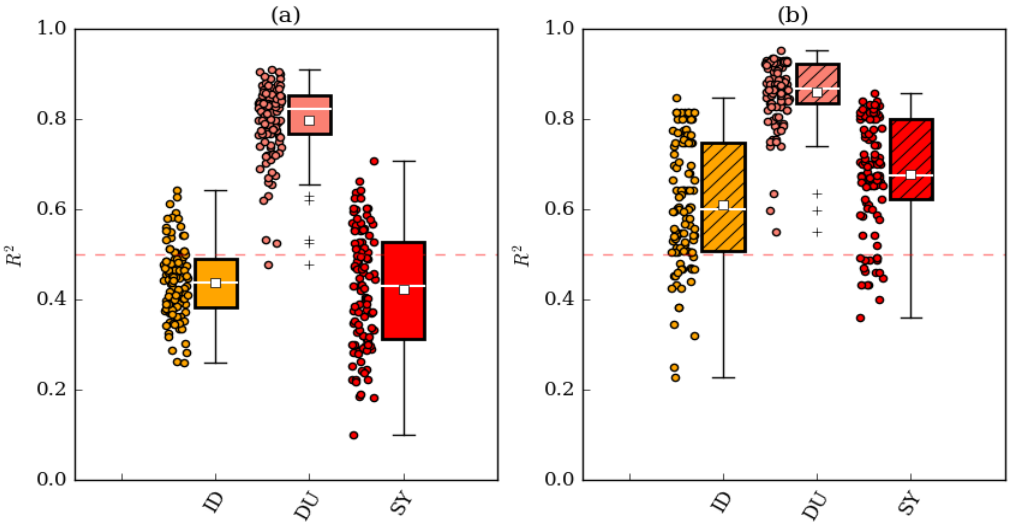

Figure 4.11. Similar to Figure 4.10, but at the urban sites.

4.3.4 Regression coefficients

In this section, the regression coefficients of the entire spectral components for both datasets are shown for the entire rural and urban stations. Similar to the section 4.3.3, in this section the results are given by using box plot and the

descriptions of the results are based on the mean and median values of the plots. Moreover, in this section the results are shown for the rural sites and the corresponding results for the urban sites are available in Appendix D.

The regression coefficients from the MLR model for the standardized ID component of two datasets at the rural sites are given in Figure 4.12. The regression coefficients in Figure 4.12.a show a large a_6 reflecting a strong relationship between O_3^{surf} and NO_x. Although the strong relationship between O_3^{surf} and NO_x is appeared in the simulated datasets (Figure 4.12.b), it is more pronounced in the WRF-Chem model than that the reality. In similarity with the rural sites, a strong anti-correlation between ID component of the O_3^{surf} and NO_x were shown for the urban stations (Figure 1D). The relationship between measured O_3^{surf} and NO_x at the urban sites is stronger than the rural sites (Figure 4.12.a and 1D.a) reflecting the stronger contribution of the photochemistry to the ID O_3^{surf} variability at urban sites than the rural sites. However, it is not well represented in the a_6 of the simulated datasets (Figure 4.12.b and 1D.b). Besides that, Figures 4.12 and 1D show a low correlation between ID variability of the measured O_3^{surf} and AT that are not well represented in the simulated data. The regression coefficients of the ID component of other variables, RH, P, U and V, in both datasets at both rural and urban sites are negligible.

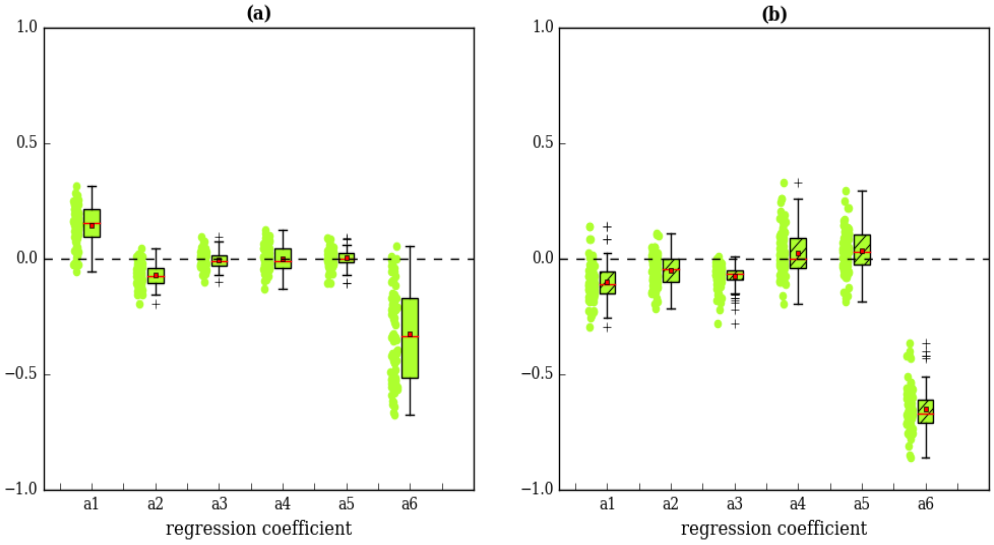

Figure 4.12. Regression coefficients for the ID spectral components of **(a)** the observed and **(b)** the simulated datasets at the rural sites. The a_1, a_2, a_3, a_4, a_5, and a_6 show the regression coefficient in the MLR model when AT, RH, SP, U, V, and NO_x used as predictors, respectively.

The regression coefficients of the DU component at the rural sites for the WRF-Chem model and observation datasets are illustrated in Figure 4.13. The results in Figure 4.13.a shows that the largest coefficient is attributed to the a_1, i.e. predictor AT. However, the largest regression coefficient in the simulated datasets is a_6, i.e. predictor NO_x (Figure 4.13.b). It demonstrates that in reality the DU O_3^{surf} variabilities are mainly produced by the DU AT variabilities, while in the WRF-Chem model simulation is mainly controlled by the DU NO_x

variability. Nevertheless, there is a relationship between O_3^{surf} and AT in the simulated datasets, but not as significant as that in the observed datasets. Besides, a_2, i.e. RH predictor, is nearly the same as a_6 in the observation datasets (Figure 4.13.a). This might show that both RH and NO_x variables have a similar influences on the DU measured O_3^{surf} variability at the rural sites. On the other hand, a low and positive value of the a_2 in Figure 4.13.b shows not only an insignificant relationship between the O_3^{surf} and RH, but also an opposite relationship between them in the WRF-Chem model simulation. The results of both model and the observation in Figure 4.13 show an insignificant a_3, a_4 and a_5, reflecting a poor relationship between O_3^{surf} and the corresponding variables, U, V and NO_x. Therefore, at the rural sites of the Med region, the DU O_3^{surf} variability in reality mainly generates by the DU variability of the AT, RH and NO_x, while in the WRF-Chem model simulation that produces by the DU variability of the NO_x and AT. In addition, the WRF-Chem model simulation shows a stronger relationship between O_3^{surf} and NO_x and a less intense relationship between O_3^{surf} and AT and between O_3^{surf} and RH, in comparison with those in reality. A similar behavior as rural sites appear in the regression coefficients at the urban sites (Figure 2D).

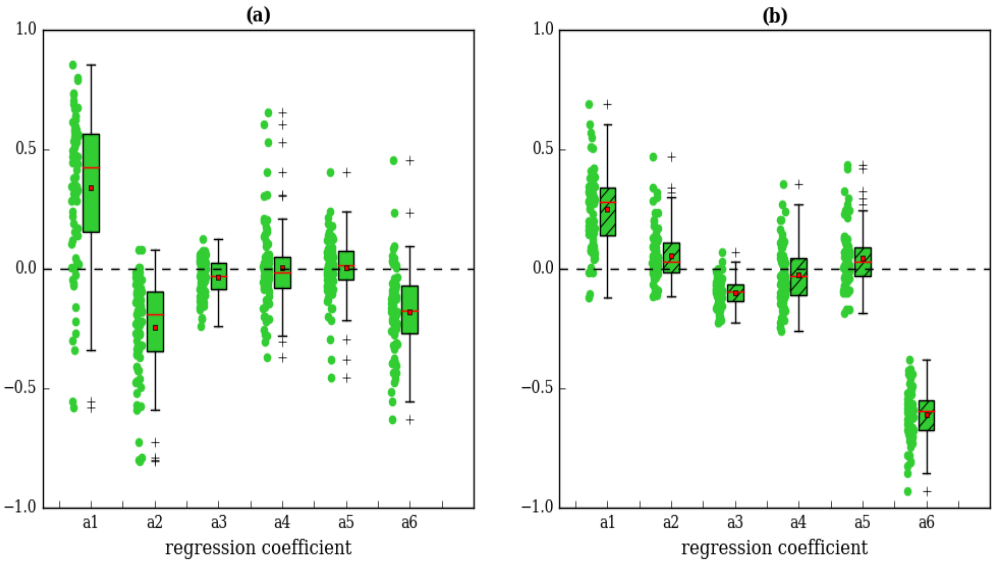

Figure 4.13. As Figure 4.12, but for the DU component.

The regression coefficient of the SY components of both datasets at the rural sites is presented in Figure 4.14. Figure 4.14.a shows the largest coefficient in the observation belongs to the a_2, RH predictor. However, in the simulated datasets the largest coefficient belongs to the a_1, AT predictor (see Figure 4.1.4.b). This reflects that the major part of the SY O_3^{surf} variability in reality generates by the SY RH variability, while in the WRF-Chem model simulation that produces by the SY AT variability. Nevertheless, there is a relationship between the SY variability of O_3^{surf} and that of AT in the observation, but not as significant as that in the model (Figure 4.14). In addition, Figure 4.14 shows a

strong anti-correlation between O_3^{surf} and NO_x in the simulated datasets that is not presented in the measurements. In both datasets, other regression coefficients such as a_3, a_4, and a_5 are not strong as a_1, a_2 and a_6. Almost similar features as the rural sites appear in the regression coefficients of the SY component at the urban sites (see Figure 3D).

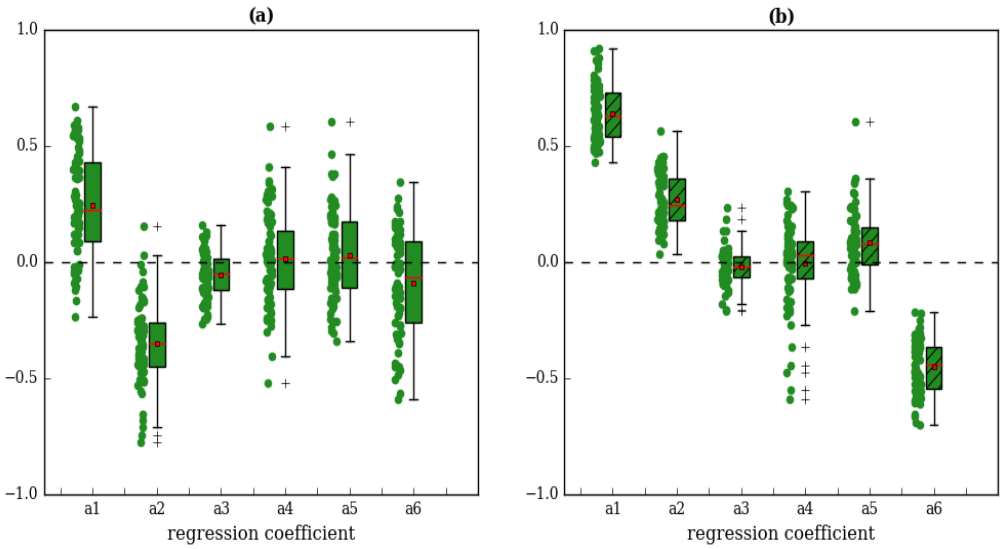

Figure 4.14. As Figure 4.12, but for the SY component.

- Summary of section 4.3

What do we learn from analysis of the measured and simulated O_3^{surf} over the Med region?

Analysis of the variance of the measured O_3^{surf} at the rural and urban sites showed more variability at the urban sites than at the rural sites. In addition, the variance of the ID and DU O_3^{surf} components are larger at the urban sites than the rural sites, whereas the variance of the SY and BL components is lower at the urban sites than the rural sites.

The variance analysis for the simulated O_3^{surf} showed that the WRF-Chem model is able to capture similar features of the O_3^{surf} variability at both rural and urban sites as those in the observation. Nevertheless, there was a disagreement between the model and the monuments in showing the amount of the variance. Overall, the analysis showed that the WRF-Chem model contains more variability for the simulated O_3^{surf} and its spectral components (except ID) in comparison to those in the measurements.

The model performance showed that the model contains almost the same MSE for simulating O_3^{surf} at both rural and urban sites. Nearly 40% MSE of the model at both sites generates from the bias error and the rest (60%) from the variance error. The variance error of the DU (ID) component contributes to major (minor) portion of the variance error at both sites. Moreover, at both rural and urban

sites, the contribution of the SY and BL components to the model variance error is lower (larger) than the DU (ID) component at those sites. Nevertheless, in terms of the order of the magnitude, the variance model error is mainly associated with the DU and next to the SY and BL components as expected from the variance of these components. On the other hand, the results of the MSE apportionment revealed that the variance model error and unexplained error of the ID and DU components at the urban sites are larger than the rural sites. Considering the known feature of the O_3^{surf} variability at urban versus rural sites, i.e. strong O_3^{surf} photochemistry at urban, a possible sources of the model variance error of the ID and DU components can be the model deficiency in resolving the meso-scale phenomena such as PBLH, NO titration, and their influence on O_3^{surf} concentration. The variance model error and explained error of the SY component at the urban sites is lower than the rural sites, as expected from the lower SY O_3^{surf} variance at the urban sites than the rural sites.

5 Discussion

5.1 Robustness analysis of the method

5.1.1 Sensitivity to the separation parameters

According to the results in the previous chapter, more than half of the model error in simulating O_3^{surf} consists of the variance error. Of these variance errors the DU and SY components contributes most to the total model variance error, whereas the variance error of the ID has the least contribution. In addition, the variance of the DU and SY components create most of the short-term O_3^{surf} variability, while the variance of the ID term only makes a small portion of that. Although theoretically the perfect separation refres to the zero covariance between components (cov_{cc}), in reality that is impossible. The reason behind is the presence of correlations between components, particularly those which are adjacent to each other such as ID-DU, DU-SY and SY-BL (Kang et al., 2013). The parameters m and k in the KZ filter tune the windows length and sharpness of the filter, respectively. Thus, the variances of the spectral components depend on the choice of m and K. One can find an optimum value for m_i and k_i for which cov_{cc} is minimized. Hereon, several questions might arise such as:

- How much will the variances of spectral terms change, if we use the optimum value for m_i and k_i?

- Will the dominant spectral components at a station remain the same as before, if the optimum value for m_i and k_i are used for the decomposition?

- Do different variables yield the same optimum m_i and k_i values? How much do they differ?

- How much will the variance error change, if the optimum spectral components are used for the model evaluation?

- How do the regression coefficients and R^2 change in the MLR analysis of the optimum spectral components?

In the following, these questions will be addressed. The optimum m_i and k_i, those mentioned in the methodology chapter, have been shown here with $m_{i\text{-}opt}$ and $k_{i\text{-}}$

$_{opt}$, and their references, with $m_{i\text{-}ref}$ and $k_{i\text{-}ref}$. We used the term "optimum spectral components" for the spectral components from the KZ($m_{i\text{-}opt}$, $k_{i\text{-}opt}$), and "reference spectral components" for those from the KZ($m_{i\text{-}ref}$, $k_{i\text{-}ref}$).

5.1.1.1 Variance of the spectral components

To figure out the sensitivity of the variance of the spectral components to m_i and k_i, i.e. length of the window and iteration times in the separation technique (Chapter 3), an iterated experiment was set up in which for each iteration one of the parameters, either m_i or k_i, was modified, holding the others constant. Since in this study only the short-term spectral components, i.e. ID, DU and SY, had been analyzed, the sensitivity of their variance to m_1, m_2, k_1, and k_2, (i.e. tuning parameters in the separation technique for the ID, DU and SY) was assessed holding k_3 and m_3 constant.

Two possibilities, 3 and 5, for k_1 and two possibilities, 3 and 5, for m_1 were assumed. The combination of these values generates four different states such as (3,3), (3,5), (5,3) and (5,5) for the (m_1,k_1). According to Eq.3.9, in the KZ(m_1,k_1) filter, these states refer to the fluctuation occurring in the time period lower than 5, 6.7, 8.6 and 11.2 hour, respectively, which recognizes as an ID($t_{<12h}$). Likewise, different possibilities for m_2 and k_2 were assumed (Table 5.1). Different combinations of these values generate fifteen states for the KZ(m_2, k_2) filter that are referring to the DU($t_{12h\text{-}2.5d}$) fluctuation. Therefore, in this experiment, sixteen states are eligible to use as m_i and k_i in the KZ filter.

Table 5.1. Eligible values for the k_i, iteration times, and m_i , length of the window, in the separation technique. The values were selected based on Eq. 3.9 to obtain the ID($t_{<12h}$) and DU($t_{12h\text{-}2.5d}$).

parameters	possibilities
k_1	3, 5
k_2	3, 5, 7
m_1	3,5
m_2	9, 11, 13, 15, 17

The time series of O_3^{surf} was decomposed by applying sixteen states in the KZ(m_i, k_i) filter. Subsequently, the cov_{cc} for each state at every station were calculated. Figure 5.1 shows the variability of cov_{cc} as a function of m_i and k_i at the ACE station. The sensitivity of cov_{cc} to the iteration time, k_1 and k_2, assuming (3,13) as (m_1, m_2) is given in Figure 5.1.a. In this figure, cov_{cc} is nearly 27 when (k_1, k_2) set to (3,7), while it increases to 30 by setting (k_1, k_2) to (5,3). Because of the need for minimizing cov_{cc}, (3,7) represents a more appropriate option than (5,3) for the (k_1, k_2). In a similar way, Figure 5.1.b shows the variability of cov_{cc} as a function of the window size, m_1 and m_2, assuming (3,5) as (k_1, k_2). In this figure, the cov_{cc} is nearly 27 when (3,17) is applied for the (m_1,m_2) in the KZ filter, whereas it is raised to 32 when (5,9) is used as (m_1, m_2). Hence, (3,17) is a more

appropriate selection than (5,9) for (m_1, m_2) in the KZ filter. Besides that, these results imply the higher sensitivity of the KZ filter to the window size (m_i) than to the iteration time (k_i), a fact that is also approvable from the root square of k in Eq. 3.9.

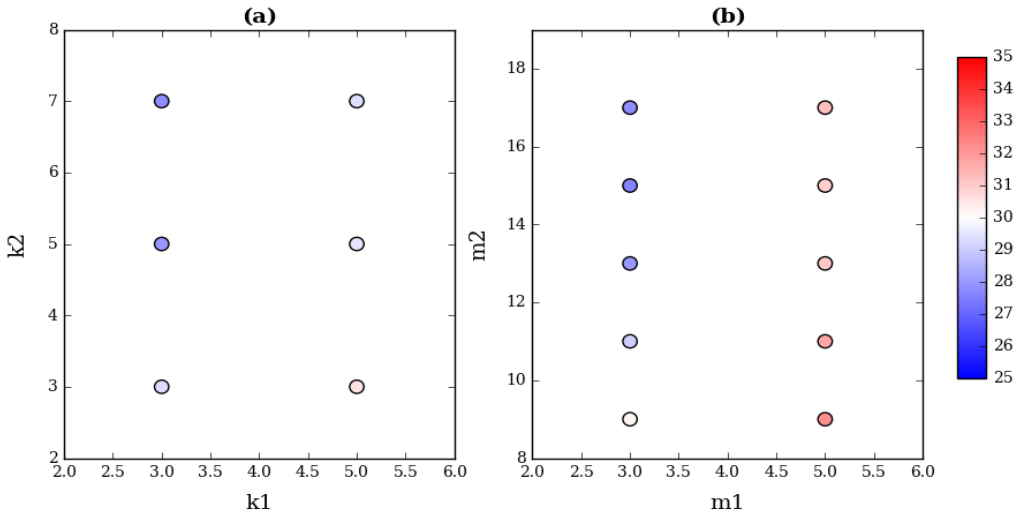

Figure 5.1 Sensitivity of cov_{cc} to the m_i and k_i at the ACE station. Figure shows the sum of the cov_{cc} among the spectral components of the measured O_3^{surf} for several **(a)** iteration times (k), assuming (3, 13) as (m_1,m_2), and **(b)** window sizes (m), assuming (3,5) as (k_1,k_2).

In the next step, the ($m_{i\text{-}opt}$, $k_{i\text{-}opt}$) among sixteen possibilities of the (m_i, k_i) was found for each station. As an example, at the ACE station, (3, 3) and (17, 3) were selected as the optimum sets for (m_1, k_1) and (m_2, k_2) in the KZ filter. At this station, ($m_{1\text{-}opt}$, $k_{1\text{-}opt}$) is the same as ($m_{1\text{-}ref}$, $k_{1\text{-}ref}$), while ($m_{2\text{-}opt}$, $k_{2\text{-}opt}$) is different from ($m_{2\text{-}ref}$, $k_{2\text{-}ref}$), i.e. (13, 5).

To determine the impact of the optimum m_i and k_i sets on the analysis, the O_3^{surf} time series was decomposed by using the KZ($m_{i\text{-}opt}$, $k_{i\text{-}opt}$) filter. Figure 5.2 shows the contribution of the optimum spectral components to the O_3^{surf} variation for both datasets at the ACE station. The contribution of the ID_{opt} to the O_3^{surf} variability remains the same as ID_{ref} at this site, i.e. 4.8% and 2.2% for the measurements and the WRF-Chem model, respectively (panels (a) and (c) in Figure 5.2). Figures 5.2.a and 5.2.c show that the contribution of the DU_{ref} to the measured O_3^{surf} variability is 29% while it is 34.9% for the DU_{opt}. Moreover, the contributions of the SY_{ref} and SY_{opt} to the measured O_3^{surf} are 43.5 and 38%, respectively. Thus, using the KZ($m_{i\text{-}opt}$, $k_{i\text{-}opt}$) filter for the decomposition of the measured O_3^{surf} increases the contribution of the DU to the total O_3^{surf} variability by 5.9% and decrease the SY by 5.5%. Therefore, contribution of the spectral components to the measured O_3^{surf} variability at the ACE station slightly changes (≈ 5%) by using the optimum $m_{i\text{-}opt}$, $k_{i\text{-}opt}$ for the KZ filter. In similarity with the measurements, applying KZ($m_{i\text{-}opt}$, $k_{i\text{-}opt}$) for the decomposition of the simulated O_3^{surf} increase the DU contribution by 8.1%, quantitively meaning 49.7 and 57.8 for the DU_{ref} and DU_{opt}, respectively, in Figures 5.2.b and 5.2.d. In addition, the

contribution of the SY decreases by 3.3%, which are 24.3 and 21 for the SY_{ref} and SY_{opt}, respectively, in Figures 5.2.b and 5.2.d. Therefore, the contribution of the spectral components to the total O_3^{surf} variability is not robust against the optimum $m_{i\text{-}opt}$, $k_{i\text{-}opt}$ at the ACE station. Nevertheless, the dominant spectral component, i.e. SY for the measurement and DU for the WRF-Chem model simulation, remains the same as those in the reference decomposition at this site.

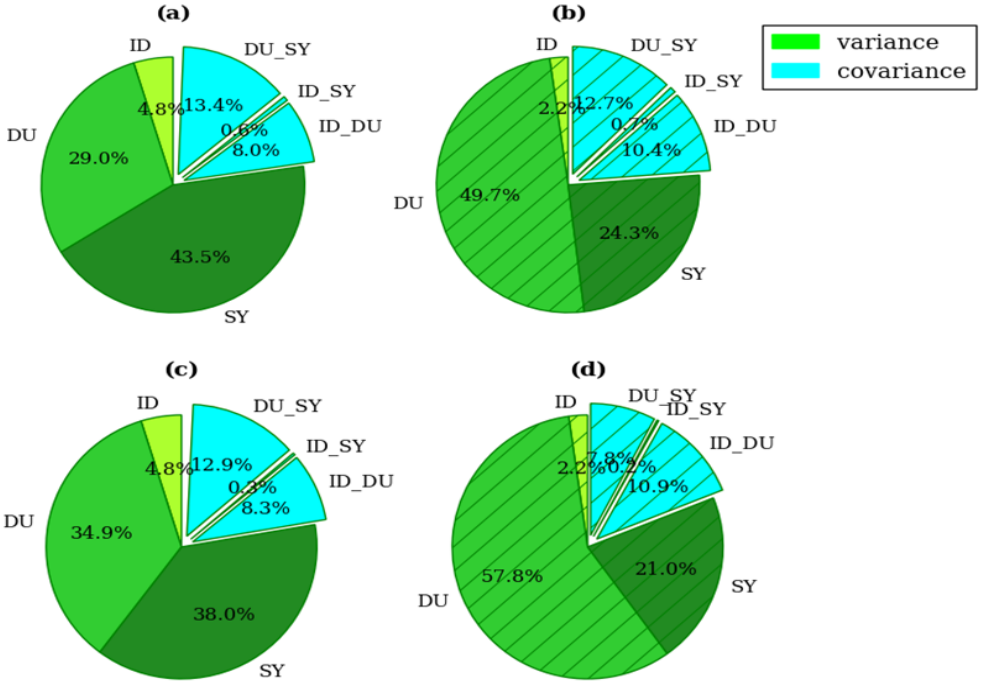

Figure 5.2. Relative contribution of the spectral components variability to the short-term O_3^{surf} variability at the ACE station. The pie chart shows when **(a)-(b)** (13, 5), and **(c)-(d)** (17, 3) use as (m_2, k_2) in the KZ filter. The solid and dashed pie charts represent the results for the measurements and the WRF-Chem model simulation datasets, respectively.

Furthermore, the O_3^{surf} time series was decomposed by using the KZ($m_{i\text{-}opt}$, $k_{i\text{-}opt}$) filter for the entire rural and urban stations. The contributions of the optimum spectral components to the total O_3^{surf} variability were calculated. The differences between the contribution of the optimum and references spectral components to the total O_3^{surf} variability for both datasets are given in Table 5.2. From the results in Table 5.2 it appears that at the rural sites the contribution of the ID_{opt} only increase by 0.3% and 0.2% in comparison to the ID_{ref} of the measurements and the WRF-Chem model, respectively. Moreover, at these sites, the contribution of the DU_{opt} to the total measured and simulated O_3^{surf} variability increase by 8.9% and 7.2%, respectively. However, the contribution of the SY_{opt} in both measurement and the model datasets decrease by 3.6% and 4.9%, respectively (Table 5.2). The results in Table 5.2 show that the contribution of the ID_{opt} to the total O_3^{surf} variability at the urban sites for both datasets remains the same as that of the ID_{ref}. Similar to the rural sites, at the

urban sites the contribution of the DU_{opt} increase by 10.8%, 9.8% and the contribution of the SY_{opt} declines by 4.2%, 5.6%, for both measurement and the model datasets, respectively (Table 5.2). Therefore, using KZ($m_{i\text{-}opt}$, $k_{i\text{-}opt}$) instead of KZ($m_{i\text{-}ref}$, $k_{i\text{-}ref}$) changes the amount of the contribution of the spectral components to the total O_3^{surf} variability changes, e.g. nearly 10% and 5% for the DU and SY components, respectively. Nevertheless, similar to the references spectral components, the DU_{opt} and ID_{opt} contribute to the largest and smallest portion of the total O_3^{surf} variability at both rural and urban sites.

Table 5.2. Contribution of the spectral components (percentages) to the total O_3^{surf} variability using ($m_{i\text{-}ref}$, $k_{i\text{-}ref}$) and ($m_{i\text{-}opt}$, $k_{i\text{-}opt}$) in the KZ filer for the rural and urban stations.

datasets		the observations			the WRF-Chem model simulation		
spectral components		ID	DU	SY	ID	DU	SY
rural	($m_{i\text{-}ref}$, $k_{i\text{-}ref}$)	3.8	49.2	21.5	2.1	42.5	32.3
	($m_{i\text{-}opt}$, $k_{i\text{-}opt}$)	4.1	58.1	17.9	2.3	49.7	27.4
	differences	0.3	8.9	-3.6	0.2	7.2	-4.9
urban	($m_{i\text{-}ref}$, $k_{i\text{-}ref}$)	4.4	53.6	15.5	2.1	48	26.1
	($m_{i\text{-}opt}$, $k_{i\text{-}opt}$)	4.4	64.4	11.3	2.1	57.8	20.5
	differences	0	10.8	-4.2	0	9.8	-5.6

From above discussion, the improvement of separation that can be obtained from optimizing k and m is not small. However, it gets more complicated to interpret the results, in particular if one also optimizes k and m for other variables (Table 5.3). Therefore, we assumed the regularities of the spectral components are robust and insensitive to the separation technique.

Table 5.3. The ($m_{i\text{-}opt}$, $k_{i\text{-}opt}$) for the spectral decomposition of all variables at the ACE station.

datasets	the observations		the WRF-Chem model simulation	
(m_i, k_i)	(m_1,k_1)	(m_2,k_2)	(m_1,k_1)	(m_2,k_2)
O_3^{surf}	(3,3)	(17,3)	(3,3)	(17,5)
AT	(3,3)	(17,5)	(3,3)	(17,7)
RH	(3,3)	(17,5)	(3,3)	(17,7)
SP	(3,3)	(9,3)	(3,3)	(9,3)
U	(5,5)	(9,3)	(3,3)	(17,5)
V	(3,3)	(17,7)	(3,3)	(17,7)
NO_x	(3,3)	(17,3)	(3,3)	(17,7)

5.1.1.2 The Model variance error

To assess the sensitivity of the analysis of the model performance (i.e. model variance error) to the separation parameters, m_i and k_i, first the optimum spectral components, i.e. ID_{opt}, DU_{opt}, and SY_{opt}, were obtained by using ($m_{i\text{-}opt}$, $k_{i\text{-}opt}$) instead ($m_{i\text{-}ref}$, $k_{i\text{-}ref}$) in the KZ filter. Second, the variance model errors for the optimum spectral components were calculated. The results of this analysis for the ACE station are given in Table 5.4.

From the results in Table 5.4 it appears that the model variance error of the ID_{opt} remain the same as IDr_{ef}, = 11.09. That is because at this station, ($m_{1\text{-}opt}$, $k_{1\text{-}opt}$) is the same as ($m_{1\text{-}ref}$, $k_{1\text{-}ref}$), i.e. (3,3), (see section 5.1.1.1). However, the variance error of the DU_{opt} increases by 6.04 compared to that of DU_{ref} (Table 5.4), which is likely due to the increasing of the DU_{opt} variability with respect to the DU_{ref} at this site, as seen in the previous section. Likewise, the variance error of the SY_{opt} shows a reduction of 1.35 comparing to the SY_{ref} (see Table 5.4). As seen in section 5.1.1.1 the SY_{opt} variability is lower than the SY_{ref} at this station. Overall, the variance error of the entire short-term spectral components (ID+DU+SY) at the ACE station increases by 4.82 ppb^2, by using (m_{iopt}, k_{iopt}) instead of ($m_{i\text{-}ref}$, $k_{i\text{-}ref}$) in the KZ filter. This alteration counts as a negligible uncertainty in the variance error of the O_3^{surf} simulation at this station. In similarity with the variance error, the robustness of the explained and unexplained errors was investigated. The results of this anaylsis for the ACE station are given in Table A4. Results in Table A4 shows that the explained and unexplained model error for simutaing of the O_3^{surf} at the ACE station are 131.32 and 95.65 when ($m_{i\text{-}opt}$, $k_{i\text{-}opt}$) used in the KZ filter. As seen in Table 4.14, the explained and unexplained model error at the ACE station are 127.66 and 94.48, respectively, using ($m_{i\text{-}ref}$, $k_{i\text{-}ref}$) in the KZ filter. This reflects that varainance error apportionements are less sensitivite to the separataion paramaters, m_i and k_i, in the KZ technique.

Table 5.4. The model variance error at the ACE station when using ($m_{i\text{-}opt}$, $k_{i\text{-}opt}$) and ($m_{i\text{-}ref}$, $k_{i\text{-}ref}$) in the KZ filter.

spectral components		ID	DU	SY	SUM
ACE	**($m_{i\text{-}ref}$, $k_{i\text{-}ref}$)**	11.09	126.06	85.14	222.15
	($m_{i\text{-}opt}$, $k_{i\text{-}opt}$)	11.09	132.1	83.79	226.97
	differences	0	6.04	-1.35	4.82

A similar robustness analysis was performed on the variance model error for simulating the O_3^{surf} at the entire rural and urban sites. Results of this analysis are given in Table 5.5. Results in Table 5.5 show that the variance model error for simulating the ID O_3^{surf} component of at the rural sites are 9.9 and 9.01 by using ($m_{i\text{-}opt}$, $k_{i\text{-}opt}$) and ($m_{i\text{-}ref}$, $k_{i\text{-}ref}$) in the KZ filter, respectively. That indicates a very small change, 0.98, in the model variance error of the ID at the rural sites. However, there is no change in the model variance error of the ID at the urban

sites (Table 5.5). The results in Table 5.5 shows that the variance model error of the DU at both rural and urban sites are 70.37 and 83.74, respectively, if ($m_{i\text{-}opt}$, $k_{i\text{-}opt}$) are used in the KZ filter. However, those were 58.74 and 66.71 if ($m_{i\text{-}ref}$, $k_{i\text{-}ref}$) was used in the KZ filter (Table 5.5). This reflects an enhancement, 11.63 and 17.2, in the variance error of the DU at both rural and urban sites, respectively, when using ($m_{i\text{-}opt}$, $k_{i\text{-}opt}$) instead of ($m_{i\text{-}ref}$, $k_{i\text{-}ref}$) in the KZ filter. On the other hand, the variance model error of the SY component are reduced by 11.94 and 15.08 at the rural and urban stations, respectively, when using ($m_{i\text{-}opt}$, $k_{i\text{-}opt}$) insted of ($m_{i\text{-}ref}$, $k_{i\text{-}ref}$) in the KZ filter (Table 5.5). Nevertheless, the sum of the model variance error shows a small increase, 0.54 and 1.95, at the rural and urban sites, respectively. Moreover, in similarity to the DU_{ref}, at both rural and urban sites the variance error of the DU_{opt} has the main contribution to the model variance error. The results of the robustness analysis of the explained and unexplained error at the rural and urban sites are given in Table C4 and Table D4, respectively.

From the above discussion, one can conclude that, although the variance model errors and its apportionmnets at both rural and urban sites shows an uncertainty (≈ 10%) to the separation parameters, i.e. m_i and k_i, in the KZ technique, the main conclusions remain unchanged. For instance, the variance error of the DU (ID) has the largest (smallest) contribution to the model error. The unexplained model errors of the ID and DU are larger than their unexplained model error, whereas for the SY the explained error is larger than the unexplained error.

Table 5.5. As in Table 5.4, but for all rural and urban sites. A similar table when (mi-ref, ki-ref) use in the KZ filter are given in Table 4.14.

spectral components		ID	DU	SY	SUM
rural	**($m_{i\text{-}ref}$, $k_{i\text{-}ref}$)**	9.01	58.74	65.49	133.27
	($m_{i\text{-}opt}$, $k_{i\text{-}opt}$)	9.9	70.37	53.55	133.81
	differences	0.89	11.63	-11.94	0.54
urban	**($m_{i\text{-}ref}$, $k_{i\text{-}ref}$)**	12.61	66.71	61	140.32
	($m_{i\text{-}opt}$, $k_{i\text{-}opt}$)	12.61	83.74	45.92	142.27
	differences	0	17.03	-15.08	1.95

5.1.1.3 Regression coefficients and R^2

To investiagte the sensitivity of the regression coefficients and R^2 to the separation parametrs, m_i, k_i, in the KZ technique, first, the time series of all variables were decomposed by using the KZ($m_{i\text{-}opt}$,$k_{i\text{-}opt}$). Then, a MLR model was estimated for each of the optimum spectral components. The results of this anaylsis for the R^2 for the ACE station are given in Table 5.6. The results in Table 5.6 show R^2 of the ID component for the observation and the WRF-Chem model simaulation datasets are 0.01 and 0.61, respectively, by using either ($m_{i\text{-}opt}$, $k_{i\text{-}opt}$)

or ($m_{i\text{-}ref}$, $k_{i\text{-}ref}$) in the KZ filter. That is related to the same ($m_{1\text{-}opt}$, $k_{1\text{-}opt}$) and ($m_{1\text{-}ref}$, $k_{1\text{-}ref}$) at the ACE station (see section 5.1.1.1). In Table 5.6 the R^2 for the DU component of the observed data only increase by 0.01 when using ($m_{i\text{-}opt}$, $k_{i\text{-}opt}$) instead of ($m_{i\text{-}ref}$, $k_{i\text{-}ref}$) in the KZ filter. However, this remains unchanged for the WRF-Chem model simulation data. This might be related to more similarity of the (m_i, k_i) among simulated variables than the observed data (see Table 5.3). On the other hand, results in Table 5.6 show that R^2 for the SY component decreases by 0.18 and 0.15 in the observed and simulated datasets, respectively. Thus, the R^2 of the SY component changes more than other componenets, i.e. ID and DU, at the ACE station. A possible reason for it might be that the SY variability only depends on (m_2, k_2), however the DU varaiability controls by both (m_1, k_1) and (m_2, k_2). Having seen in section 5.1.1.1, at the ACE station, ($m_{1\text{-}opt}$, $k_{1\text{-}opt}$) and ($m_{1\text{-}ref}$, $k_{1\text{-}ref}$) are the same, while ($m_{2\text{-}opt}$, $k_{2\text{-}opt}$) and ($m_{2\text{-}ref}$, $k_{2\text{-}ref}$) are different. From these, one can conclude that at the ACE station the SY component of the varaibles is more sensitive than the DU component. The results of the regression coefficients of the MLR model for the spectral components when using ($m_{i\text{-}opt}$, $k_{i\text{-}opt}$) for the KZ filter are given in Table A5. The results in Tabel A5 show only small changes comparing to Table 4.4, i.e. the regression coefficients of when using ($m_{i\text{-}ref}$, $k_{i\text{-}ref}$) in the KZ filter. This indicates that the regression coefficients show a low sensitivity to the separation parameters, m_i and k_i, in the KZ technique at the ACE station.

Table 5.6. As in Table 5.4, but for R^2.

datasets		**the observations**			**the WRF-Chem model simulation**		
spectral components		**ID**	**DU**	**SY**	**ID**	**DU**	**SY**
ACE	**($m_{i\text{-}ref}$, $k_{i\text{-}ref}$)**	0.01	0.31	0.86	0.61	0.86	0.88
	($m_{i\text{-}opt}$, $k_{i\text{-}opt}$)	0.01	0.32	0.68	0.61	0.86	0.73
	differences	0	0.01	-0.18	0	0	-0.15

Figure 5.3 shows the R^2 for the entire rural station when ($m_{i\text{-}opt}$, $k_{i\text{-}opt}$) was used in the KZ filter. The results in Figure 5.3 show that R^2 values for the DU component are the largest in both observed and simulated datasets. However, the smallest R^2 in both datasets belong to the ID component in Figure 5.3. Moreover, the R2 for the SY componenet is between the R^2 for the ID and DU for both datasets (Figure 5.3). Similar features were seen in Figure 4.10, i.e. when ($m_{i\text{-}ref}$, $k_{i\text{-}ref}$) was used in the KZ filter, that indicates robustness of the R^2 to the separation parameters, m_i and k_i, in the KZ technique at the rural stations. The sensitivity of the R^2 to the m_i and k_i are shown in Figure D4.

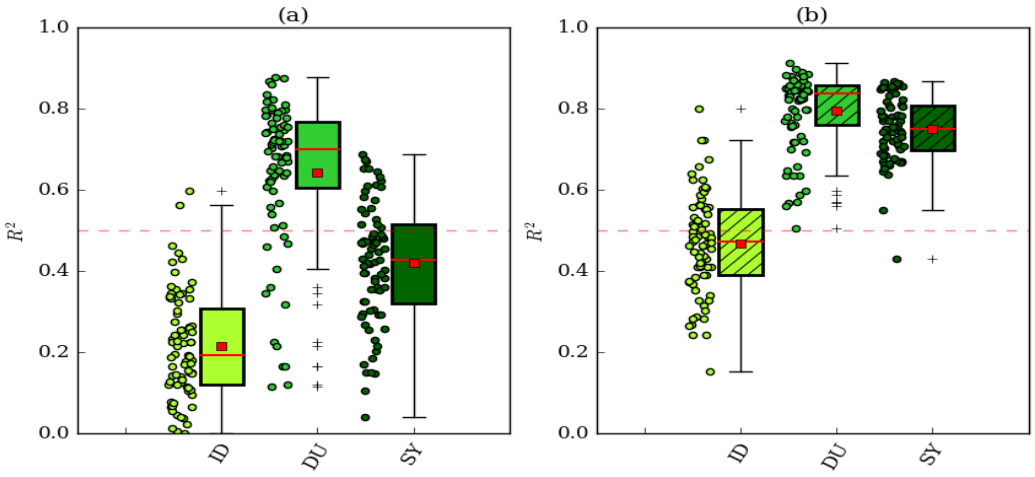

Figure 5.3. The R^2 from MLR model for the ID, DU and SY components when using ($m_{i\text{-}opt}$, $K_{i\text{-}opt}$) instead of ($m_{i\text{-}ref}$, $k_{i\text{-}ref}$) in the KZ filter for **(a)** the observations and **(b)** the WRF-Chem model simulation datasets, at the rural sites. The scatter point adjacent to the box plot presents the R^2 at each station. A similar plot are shown in Figure 4.10 when ($m_{i\text{-}ref}$, $k_{i\text{-}ref}$) was used in the KZ filter.

5.1.2 Sensitivity to the station types (background, industrial, and traffic)

As seen in Chapter 4, the WRF-Chem model simulation showed the major features of the differences between O_3^{surf} variability at the rural and urban sites. For instance, in similarity with the observations, the WRF-Chem model simulation shows a larger DU O_3^{surf} variability at urban sites than at the rural sites. Moreover, the regression analysis showed that R^2 of the ID and DU components increase moving from rural to urban sites (see section 4.3.3). In this section, the capability of the WRF-Chem model simulation in capturing the O_3^{surf} variability at different station types, such as background, industrial and traffic, of the rural and urban sites are assessed.

5.1.2.1 Variance of the spectral components

The variability of the spectral components of the measured and simulated O_3^{surf} for the different station types is given in Table 5.7. Results on Table 5.7 show that for all station types, the DU and ID components constructs the most and least O_3^{surf} variability in both datasets, respectively. However, in the rural area, the variability of the spectral components of the observed O_3^{surf} at the background sites is larger than that at the industrial stations (see Table 5.7). That might arise from the stronger influence of the meteorology, such as PBLH and transports, on the O_3^{surf} variability at background sites than at industrial stations. On the other hand, the ID and DU O_3^{surf} variability at the industrial-urban sites are larger than the background-urban station (see Table 5.7). A possible reason for that might be stronger photochemical activities at industrial-urban sites than at background-urban stations. In addition, the results in Table 5.7 show that the SY O_3^{surf} variability for the observations at the background sites for both rural and urban area is larger than industrial and traffic sites. From the results in Table 5.7, it appears that the WRF-Chem model simulation is able

to show the change of the O_3^{surf} variability moving from a staion type to another; however, that is not the same as the observation.

Table 5.7. The variance of the spectral components for each stations type.

datasets		the observations			the WRF-Chem model simulations		
spectral components		ID	DU	SY	ID	DU	SY
rural	background	5.85	86.87	28.04	4.46	98.92	69.8
	industrial	5.67	67.66	22.92	3.82	67.62	61.47
urban	background	8.71	113.02	27.84	5.33	127.69	58.14
	industrial	11.08	130.59	19.47	7.32	196.51	41.82
	traffic	10.25	112.31	19.63	6.65	140.87	54.65

5.1.2.2 The R^2

The R^2 for the ID component of both datasets at different station types are shown in Figure 5.4. Figures of the R2 for the DU and SY components are illustrated in Appendix E. Figure 5.4 shows that R^2 for the observations dataset at the rural-background and rural-industrial stations are nearly the same and lower than the urban sites. That arises from the stronger relationship between O_3^{surf} and NO_x, and likely stronger photochemistry, at the urban sites than the rural sites (see section 4.3.4). Although the WRF-Chem model is able to capture that, it shows a larger R^2 at the rural-industrial sites than the rural-background (see Figure 5.4). The same feature appears in R^2 of the DU component of the observation dataset; however, the WRF-Chem model is not able to capture it (see Figure E1). Thus, one can conclude that the WRF-chem model simulation is not able to capture the observed relationship between O_3^{surf} and the predictors at different station types.

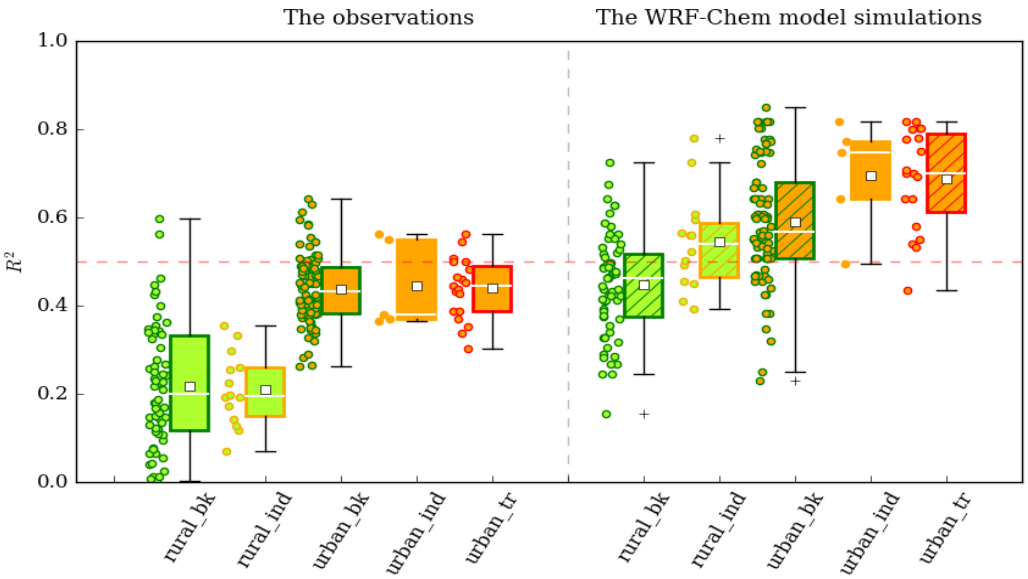

Figure 5.4. R^2 for the ID spectral component at different station types, background (bk), industrial (ind), and traffic (tr) of the rural and urban sites.

5.2 What do we learn from the analysis of the observations dataset about O_3^{surf} variability over the Med region?

The spectral analysis of the measured O_3^{surf} variability showed that the major portion, ≈ 50%, of the short-term O_3^{surf} variability at the Med stations consists of the DU variability. This is similar to previous findings (Solazzo et al., 2016; Galmarini et al., 2013) in which it has been expressed that the DU component derives the O_3^{surf} variability at the stations over EU and NA regions and accounts for more more than half of the total O_3^{surf} variance. In this study, we could confirm this for most of the Med stations. A possible reason for the high DU O_3^{surf} variability is the large influence of the daytime photochemistry production and nighttime O_3 removal on the O_3^{surf} variability (Rao et al., 1997; Hogrefe et al., 2000). Moreover, the contribution of the SY O_3^{surf} variability to the short-term O_3^{surf} variation in the measurements dataset is smaller than DU. A similar feature founds in the literature (Hogrefe et al., 2013; Solazzo et al., 2016), in which the variance of the SY component is followed after the DU variability for the O_3^{surf} over EU and NA. Unlike the DU and SY components, the ID produces a small variation into the O_3^{surf} variability. This finding is in agreement with the literature (Hogrefe et al., 2000, 2013; Galmarini et al., 2013; Solazzo et al., 2016), which all show that the ID component contains the smallest contribution to the total O_3^{surf} variance due to the small magnitude of its fluctuation.

In addition, similar to the finding of Galmarini et al. (2013) the DU O_3^{surf} variability is getting weaker when moving from urban to rural sites. A possible reason for it is that the reduction of the photochemical activity moving from urban to rural sites (Galmarini et al., 2013). The analysis of the measured O_3^{surf} at different station types, such as background, industrial, and traffic, showed that the variance of the spectral components varies by changing the station types. However, Solazzo et al. (2016) found that the variance of the spectral components is not associated with the area type of the monitoring stations, i.e. rural, urban, or suburban. The results of the spectral analysis over different Med sub-regions (See Figure E4) shows that the SY O_3^{surf} variability is more predominant over Western and Central Med than Eastern Med. That is likely due to the strong influence of the synoptic meteorological systems on the O_3^{surf} over the Med region (Doche et al., 2014; Kalabokas et al., 2008). However, the results showed that over Eastern Med, the DU O_3^{surf} variability constructs the major portion of the O_3^{surf} variability (See Figure E4). That might arise from the strong photochemistry over this sub-region.

The regression analysis of the standardized spectral components for the observations dataset showed that the MLR model, with AT, RH, SP, U, V, and NO_x as predictors, is able to explain more (less) than half of the DU (ID and SY) O_3^{surf} variability. That is might be because of the strong linear relationship between DU variability of the AT and O_3^{surf} as shown by the regression coefficients.

Moreover, the results showed that the ID O_3^{surf} variability is strongly influenced by the ID NO_x variability. As this influence is getting stronger at the urban sites than the rural sites, it might indicate a larger contribution of the chemical processes such as NO titration to the ID O_3^{surf} variability at the urban sites than the rural. In addition, from the results of the regression analysis, it appeared that the SY O_3^{surf} variability at the rural stations is majorly influenced by the SY variability of the RH and AT, while at the urban sites it is controlled by the SY variability of the RH, AT, and NO_x. That might reflect a larger influence of the SY meteorological systems on the O_3^{surf} at the rural sites than the urban. The MLR analysis did not indicate a strong relationship between O_3^{surf} and some predictors such as U, V, and SP. A possible reason for this might be the non-linearity of the relationship between O_3^{surf} and these variables.

5.3 What do we learn from the analysis about the simulated O_3^{surf} variability by the WRF-Chem model?

Spectral analysis of the model simulation showed that although similar to the observations, the model shows the larger contribution of the DU variability to the total O_3^{surf} variability in comparison to ID and SY, it (i.e. DU O_3^{surf} variability in the simulated data) is larger than the measurements. However, Hogrefe et al. (2013) stated that the regional models underestimate the observed variance of the spectral components. Moreover, from the results it appeared that altought in similarity with the measurements, the model is able to show a smaller contribution of the SY to the total O_3^{surf} variability in comparison to the DU, it (i.e. SY O_3^{surf} variability in the simulated data) is larger than the measurements. That might arise from the strong influence of the lateral boundary conditions on O_3^{surf} variability (Katragkou et al., 2010; Akritidis et al., 2013). In agreement with the previous finding (Hogrefe et al., 2000, 2013; Galmarini et al., 2013; Solazzo et al., 2016), the ID variability of the simulated O_3^{surf} is smaller than the measurements. This can arise from the model weaknesses to capture the mesoscale phenomena such as sea breeze due to the low model resolution. Moreover, in a close similarity with the measurements, moving from the rural to the urban sites, the DU O_3^{surf} variability of the simulated data is getting stronger. This might indicate that the WRF-Chem model simulation is able to show the stronger photochemistry in the urban sites than the rural. However, it is not able to capture the differences between O_3^{surf} variability at different station types, i.e. background, industrial, and traffic. This is likely because of the poor model resolution, in which the model loses its accuracy for the NO_x simulation and likely the O_3^{surf}.

The results of the model performance showed that the SY component of the simulated O_3^{surf} contains a larger explained model error than the unexplained error. This illustrates that the WRF-Chem model simulation contains some SY O_3^{surf} variabilities, which do not match the observations. However, for the ID and DU components, the larger unexplained model error in comparison to the

explained error indicates some missing short-term variabilities in the model. This might arise from the processes that do not resolve in the model simulation. Another possible reason for it might be the measurements random errors, which generate spurious variability in the observation dataset.

The regression analysis showed that, in similarity with the observations, in the WRF-Chem model simulation the largest and lowest R^2 belongs to the DU and ID components, respectively. However, R^2 for all spectral components of the model simulation was larger than the observations. This indicates a stronger relationship between O_3^{surf} and predictors in the model simulation than the measurements. A possible reason for that might arise from misrepresenting of the non-linear relationships, among variables such as O_3^{surf}, AT, and NO_x, in the WRF-Chem model. As seen from the regression coefficients, there is a stronger relationship between O_3^{surf} and NO_x in the simulated dataset than the observation reflecting the linearity of the relationship between O_3^{surf} and NO_x in the model simulation (see Figure A8). In addition, from the results of the regression analysis appeared that the model contains a weakness in showing the relationship between O_3^{surf} and RH. This can be caused by the RH error due to cloud covers uncertainties and its influences on O_3^{surf} (Kim et al, 2015; Wałaszek et al., 2017). Moreover, it might be arisen from the error in the model chemical mechanism due to the uncertainties of reaction rates or missing or misrepresenting of reactions in the OH chemistry. Nevertheless, in similarity with the observations, the MLR analysis of the simulated data did not show a strong relationship between the O_3^{surf} and some predictors, such as U, V, and SP. This might be related to the non-linearity of the relationship between O_3^{surf} and the variables.

6 Summary and conclusions

Regular enhancement of the O_3^{surf} concentration over the Med region in summer has been remarked in several observation and modeling studies (Lelieveld et al., 2002; Gerasopoulos et al., 2005; Kalabokas et al., 2008; Richards et al., 2013; Doche et al., 2014; Safieddine et al., 2014). Although various hypotheses have been introduced as a factor controlling the O_3^{surf} variability over the Med region, most of them are valid only for a specific Med sub-region and for a particular episodic event. Thus, the first aim of this study was to identify the O_3^{surf} variability over the entire Med region. A dataset of the surface based measurements including data at 76 rural and 109 urban Med stations was used for summers of 2010, 2011, and 2012. As many processes in various time scales, such as ID, DU, and SY, influence on the O_3^{surf} variability, it is more informative to study its variability in each time scale. The KZ filter spectral analysis is a method to separate the O_3 time series into different temporal components and has been used in several studies over different regions (Hogrefe et al., 2000; Tarasova et al., 2003; Seo et al., 2014; Kang et al., 2013; Solazzo et al., 2013, 2016). We used this technique to analyze the observed O_3^{surf} variability over the Med region. Before this study, there was no study in which O_3^{surf} variability in the entire Med region had been analyzed using the KZ filter. The results of this analysis showed that over the Med region the DU (ID) variability contribute to the largest (smallest) portion of the total O_3^{surf} variability. Moreover, the relative contribution of the SY variability to the total O_3^{surf} variability follows the DU variability. This confirms previous findings of (Hogrefe et al., 2013; Galmarini et al., 2013; Solazzo et al., 2016). In addition, in similarity to the finding of Galmarini et al. (2013), the results of the spectral analysis of the measured O_3^{surf} over the Med region showed that the DU O_3^{surf} variability decreases when moving from the urban to the rural sites. However, Solazzo et al. (2016) found that the variance of the spectral components is not associated with the area type of the monitoring stations, i.e. rural, urban, or suburban. Nevertheless, analyzing of the measured O_3^{surf} variability for different station types, such as background, industrial, and traffic, showed that the variance of the spectral components varies by changing the station types. To determine which variable(s), such as AT, RH, SP, U, V, and NO_x, has a larger influence on the O_3^{surf} variability at each time scale over the Med region, we used an MLR model. Before this study, there was no study in which MLR analysis has been used to study the spectral variability of the measured O_3^{surf} over the Med region. The results of the MLR analysis showed that the ID, DU, and SY variability of the measured O_3^{surf} are primarily driven from the corresponding component of the NO_x, AT, and RH, respectively.

Although many ACMs are able to simulate a regional O_3^{surf} maximum over the Med region, they show inconsistencies with respect to the observed O_3^{surf} variability in the individual spectral components. While the general cause of these discrepancies can be explained by the differences between models and measurements characteristics, such as point measurements versus model grid-boxes (Barnes et al., 2016), a more detailed investigation of a single model is needed to understand why the differences arise. Accordingly, the second aim of this study was to identify how the model simulates the O_3^{surf} variability over the Med region. In the study by Solazzo et al. (2016), a set of AQMEII models evaluated against the observation over EU and NA by using the spectral analysis. They found that the variance error of the models is mainly associated to the DU O_3^{surf} variability over both regions. Nevertheless, they mentioned that the variance error of the models is peculiar to each model and has to be investigated individually. In this study, we focused on the output of one simulation from the WRF-Chem model, which is one of the AQMEII models, and analyzed the results of three consecutive summer periods. At first, similar to the observations dataset, the spectral analysis was used to determine the amount of the simulated O_3^{surf} variability in each time scale. The results of this analysis showed that the WRF-Chem model is able to capture the relative contributions of the spectral components to the total O_3^{surf} variability, meaning that the DU (ID) has the largest (least) contribution to the total O_3^{surf} variability. However, the DU (ID) variability of the simulated O_3^{surf} was overestimated (underestimated) compared to the observations. Moreover, the analysis revealed that the model simulation contains some SY O_3^{surf} variability that does not match with the measurements. In addition, consistent with the measurements, the DU O_3^{surf} variability of the simulated data increases when moving from rural to urban sites. This might indicate that the model simulation is able to show the stronger photochemistry in the urban sites than the rural. However, the model is not able to capture the differences between O_3^{surf} variability at different station types, i.e. background, industrial, and traffic. Second, we used modified MSE, a similar indicator as Solazzo et al. (2016), to determine the model performance. The results of this analysis showed that, in similarity with the NA and EU regions, which were investigated in AQMEII, the model variance error for the O_3^{surf} simulations over the Med region is primarily attributed to the DU variability. In addition, we found that the larger part of ID and DU model variance errors arise from the unexplained model error showing the lack of the observed ID and DU variability in the model. However, the variance model error for the simulation of the SY O_3^{surf} variability arises from the explained model error, which is caused by the simulated SY variability that is not matched with the observations. Third, we applied the MLR model analysis to the simulated dataset. The results showed that the MLR model is able to explain more O_3^{surf} variability in the simulation dataset than the observations reflecting a stronger relationship between O_3^{surf} and the predictors in the model simulation than that in reality. Moreover, it

appeared that the relationship between simulated O_3^{surf} and NO_x is much stronger than in the measurements for all spectral components. Therefore, further investigation is needed to identify the source of the misrepresentation, very strong linear anti-correlation, of the O_3^{surf} and NO_x in the WRF-Chem model. In addition, the regression results revealed some weaknesses in presenting the relationship between O_3^{surf} and RH in the model. Thus, we recommend a detail investigation of the RH simulation and its relationship with the O_3^{surf} variability in the WRF-Chem model. The MLR results showed a stronger relationship between SY variability of the AT and O_3^{surf} in the model simulation than that in the observation. That might be a possible reason for the explained model error of the SY variability.

The statistical methods applied in this study were able to show the O_3^{surf} variability and its relationship with other variables on different time scales. This analysis can be applied to other meteorological variables or chemical species to study their variability. Furthermore, additional parameters, such as planetary boundary layer height, can easily be added to the MLR model. Extending the analyses presented here to ensemble simulations or sensitivity studies might allow elucidation of possible errors in the process parameterizations of atmospheric chemistry transport models.

Bibliography

Bibliography

Abdul-Wahab S. A., Bakheit C. S., and Al-Alawi S. M.: Principal Component and Multiple Regression Analysis in Modelling of Ground-Level Ozone and Factors Affecting its Concentrations, *Environmental Modelling and Software*, 20(10), 1263-1271, doi: 10.1016/j.envsoft.2004.09.001, 2005.

Akritidis D., Zanis P, Katragkou E., Schultz M.G., Tegoulias I., Poupkou A., Markakis K., Pytharoulis I., and Karacostas Th.: Evaluating the Impact of Chemical Boundary Conditions on Near Surface Ozone in Regional Climate–Air Quality Simulations over Europe, *Atmos. Res.*, 134, 116–130, doi: 10.1016/j.atmosres.2013.07.021, 2013.

Baklanov A., Schlünzen K., Suppan P., Baldasano J., Brunner D., Aksoyoglu S., Carmichael G., Douros J., Flemming J., Forkel R., Galmarini S., Gauss M., Grell G., Hirtl M., Joffre S., Jorba O., Kaas E., Kaasik M., Kallos G., Kong X., Korsholm U., Kurganskiy A., Kushta J., Lohmann U., Mahura A., Manders-Groot A., Maurizi A., Moussiopoulos N., Rao S. T., Savage N., Seigneur C., Sokhi R. S., Solazzo E., Solomos S., Sørensen B., Tsegas G., Vignati E., Vogel B., and Zhang Y.: Online Coupled Regional Meteorology Chemistry Models in Europe: Current Status and Prospects, *Atmos. Chem. Phys.*, 14, 317-398,doi:10.5194/acp-14-317-2014,2014.

Barnes E. A., Fiore A. M., and Horowitz L. W.: Detection of Trends in Surface Ozone in the Presence of Climate Variability, *J. Geophys. Res. Atmos.*, 121, 6112–6129, doi: 10.1002/2015JD024397, 2016.

Bocquet M., Elbern H., Eskes H., Hirtl M., Žabkar R., Carmichael G. R., Flemming J., Inness A., Pagowski M., Pérez Camaño J. L., Saide P. E., San Jose R., Sofiev M., Vira J., Baklanov A., Carnevale C., Grell G., and Seigneur C.: Data Assimilation in Atmospheric Chemistry Models: Current Status and Future Prospects for Coupled Chemistry Meteorology Models, *Atmos. Chem. Phys.*, 15, 5325–5358, doi: 10.5194/acp-15-5325-2015, 2015.

Bollmeyer C., Keller J. D., Ohlwein C., Wahl S., Crewell S., Friederichs P., Hense A., Keune J., Kneifel S., Pscheidt I., Redl S., and Steinkea S.: Towards a High-Resolution Regional Reanalysis for the European CORDEX Domain, *Q. J. R. Meteorol. Soc.*, doi: 10.1002/qj.2486, 2014.

Chou M. D., and Suarez, M. J.: An Efficient Thermal Infrared Radiation Parametrization for Use in General Circulation Models, *NASA Tech. Memo.*, 104606, 85, doi: 10.1.1.26.4850, 1994.

Committee on Tropospheric Ozone Formation and Measurement, Board on Environmental Studies and Toxicology, Board on Atmospheric Sciences and Climate, Commission on Geosciences, Environment, and Resources, National Research Council (1991). Rethinking the Ozone Problem in Urban and Regional Air Pollution, Washington D. C., The National Academies Press.

Cooper O. R., Parrish D. D., Ziemke J., Balashov N. V., Cupeiro M., Galbally I. E., Gilge S., Horowitz L., Jensen N. R., Lamarque J. F., Naik V., Oltmans S. J.: Global Distribution and

Trends of Tropospheric Ozone: An Observation-based Review, *Elementa: Science of the Anthropocene*, 2, 1-29, doi: 10.12952/journal.elementa.000029, 2014.

Crutzen P. J. (1988) Tropospheric Ozone: An Overview. In: Isaksen I.S.A. (eds) Tropospheric Ozone. NATO ASI Series (Series C: Mathematical and Physical Sciences), Dordrecht, Springer.

Ding J., Miyazaki K., Johannes van der A R., Mijling B., Kurokawa J., Cho S., Janssens-Maenhout G., Zhang Q., Liu F, and Levelt P. F.: Intercomparison of NOx Emission Inventories Over East Asia, *Atmos. Chem. Phys.*, 17, 10125–10141, doi: 10.5194/acp-17-10125-2017, 2017.

Doche C., Dufour G., Foret G., Eremenko M., Cuesta J., Beekmann M., and Kalabokas P.: Summertime Tropospheric-Ozone Variability over the Mediterranean Basin Observed with IASI, *Atmos. Chem. Phys.*, 14, 10589–10600, doi: 10.5194/acp-14-10589-2014, 2014.

Duenas C., Fern_andez M.C., Canete S., Carretero J., Liger E: Analyses of Ozone in Urban and Rural Sites in Malaga (Spain), Chemosphere 56, doi: 631–639, 10.1016/j.chemosphere.2004.04.013, 2004.

EKovač-Andrić E., Brana J., Gvozdić V.: Impact of Meteorological Factors on Ozone Concentrations Modelled by Time Series Analysis and Multivariate Statistical Methods, *Ecological Informatics*, 4, 117–122, doi: 10.1016/j.ecoinf.2009.01.002, 2009.

Emmons L. K., Walters S., Hess P. G., Lamarque J.-F., Pfister G. G., Fillmore D., Granier C., Guenther A., Kinnison D.,Laepple T., Orlando J., Tie X., Tyndall G., Wiedinmyer C.,Baughcum S. L., and Kloster S.: Description and Evaluation of the Model for Ozone and Related Chemical Tracers, Version 4 (MOZART-4), *Geosci. Model Dev.*, 3, 43–67, doi: 10.5194/gmd-3-43-2010, 2010.

Eskridge R. E., Jia Yeong Ku J. Y., Rao S. T., Porter P. S, and Zurbenko I. G.: Separating Different Scales of Motion in Time Series of Meteorological Variables, *Bull. Am. Meteorol. Soc.*, 78, 1473-1483, doi: 10.1175/1520-0477(1997)078, 1997.

Galmarini S., Kioutsioukis I., and Solazzo E.: *E pluribus unum_*: Ensemble Air Quality Predictions, *Atmos. Chem. Phys.*, 13, 7153–7182, doi: 10.5194/acp-13-7153-2013, 2013.

Ganzeveld L., and Lelieveld J.: Dry Deposition Parameterization in a Chemistry General Circulation Model and its Influence on the Distribution of Reactive Trace Gases, *Journal of Geophysical Research*, 100, 20,999-21,012, doi: 10.1029/95JD02266, 1995.

Gao D., Stockwell W.R., Milford J. B.: Global Uncertainty Analysis of a Regional-scale Gas Phase Chemical Mechanism, *Journal of Geophysical Research*, 101, 9107-9119, doi: 10.1029/96JD00060, 1996.

Gerasopoulos E., Kouvarakis G., Vrekoussis M., Kanakidou M., and Mihalopoulos N.: Ozone Variability in the Marine Boundary Layer of the Eastern Mediterranean based on 7-year Observations, *Journal of Geophysical Research*, 110, D15309, doi: 10.1029/2005JD005991, 2005.

Goody R., Anderson J., and North G.: Testing Climate Models: An Approach, *Bull. Amer. Meteor. Soc*, 79, 2541– 2549, https://doi.org/10.1175/1520-0477(1998)079<2541:TCMAA>2.0.CO;2, 1998.

Grell G. A. and Freitas S. R.: A Scale and Aerosol Aware Stochastic Convective Parameterization for Weather and Air Quality Modeling, *Atmos. Chem. Phys.*, 14, 5233-5250, doi: 10.5194/acp-14-5233-2014, 2014.

Grell G. A., Peckham S. E., Schmitz R., McKeen S. A., Frost G., Skamarock W. C., and Eder B.: Fully Coupled "online" Chemistry Within the WRF Model, *Atmos. Environ.*, 39, 6957–6975, doi: 10.1016/j.atmosenv.2005.04.027, 2005.

Guenther A., Karl T., Harley P., Wiedinmyer C., Palmer P. I., and Geron C.: Estimates of Global Terrestrial Isoprene Emissions Using MEGAN (Model of Emissions of Gases and Aerosols from Nature), *Atmos. Chem. Phys.*, 6, 3181-3210, doi: 10.5194/acp-6-3181-2006, 2006.

Haman C. L., Couzo E., Flynn J. H., Vizuete W., Heffron B., and Lefer B. L.: Relationship Between Boundary Layer Heights and Growth Rates with Ground-Level Ozone in Houston, Texas, *J. Geophys. Res. Atmos.*, 119, 6230–6245, doi: 10.1002/2013JD020473, 2014.

Hardacre C., Wild O., and Emberson L.: An Evaluation of Ozone Dry Deposition in Global Scale Chemistry Climate Models, *Atmos. Chem. Phys.*, 15, 6419–6436, doi: 10.5194/acp-15-6419-2015, 2015.

Hogrefe C., Rao S. T., Zurbenko I. G., and Porter P. S.: Interpreting the Information in Ozone Observations and Model Predictions Relevant to Regulatory Policies in the Eastern United States, *Bull. Am. Meteorol. Soc.*, 81(9), 2083-2106, doi: 10.1175/1520-0477(2000)081, 2000.

Hogrefe C., Roselle S., Mathur R., Rao S. T., and Galmarini S.: Space-time Analysis of the Air Quality Model Evaluation International Initiative (AQMEII) Phase 1 Air Quality Simulations, *J. Air Waste Manag. Assoc.*, 64(4), 388-405, doi: 10.1080/10962247.2013.811127, 2014.

Hong S. Y., Noh Y., and Dudhia J.: A New Vertical Diffusion Package with an Explicit Treatment of Entrainment Processes, *Monthly Weather Review*, 134, 2318–2341, doi: 10.1175/MWR3199.1, 2006

Hov O., Schmidbauer N.: Atmospheric Concentrations of Nonmethane Hydrocarbons at a North European Coastal site, *J. Atmos. Chem.*, 14, 515–526, doi: 10.1007/BF00115255, 1992.

Iacono M. J., Delamere J. S., Mlawer E. J., Shephard M. W., Clough S. A., and Collins W. D.: Radiative Forcing by Long-lived Greenhouse Gases: Calculations with the AER Radiative Transfer Models, *J. Geophys. Res.*, 113, D13103, doi: 10.1029/2008JD009944, 2008.

Im U., and Kanakidou M.: Impacts of East Mediterranean Megacity Emissions on Air Quality, *Atmos. Chem. Phys.*, 12, 6335–6355, doi: 10.5194/acp-12-6335-2012, 2012.

Jacob D. J. (1999), Introduction to Atmospheric Chemistry, Princeton, Princeton University Press.

Jiménez P. A., Dudhia J., González-Rouco J. F., Navarro J., Montávez J. P., García-Bustamante E.,: A Revised Scheme for the WRF Surface Layer Formulation, *Monthly Weather Review*, 140, issue 3, 898-918, doi: 10.1175/MWR-D-11-00056.1, 2012.

Johnson C. E., Stevenson D. S., Collins W. J., and Derwent R. G.: Role of Climate Feedback on Methane and Ozone Studied with a Coupled Ocean-Atmosphere-Chemistry model, *Geophysical Research Letters*, 28, 1723-1726, doi: 10.1029/2000GL011996, 2001.

Kalabokas P., Hjorth J., Foret G., Dufour G., Eremenko M., Siour G., Cuesta J., and Beekmann M.: An Investigation on the Origin of Regional Springtime Ozone Episodes in the Western Mediterranean, *Atmos. Chem. Phys.*, 17, 3905–3928, doi: 10.5194/acp-17-3905-2017, 2017.

Kang D., Hogrefe C., Foley K. L., Napelenok S. L., Mathur R., Rao S. T.: Application of the Kolmogorove-Zurbenko Filter and the Decoupled Direct 3D Method for the Dynamic Evaluation of a Regional Air Quality Model, *Atmos. Environ.*, 80, 58–69, doi: 10.1016/j.atmosenv.2013.04.046, 2013.

Katragkou E., Zanis P., Tegoulias I., Melas D., Kioutsioukis I., Krüger B. C., Huszar P., Halenka T., Rauscher S.: Decadal Regional Air Quality Simulations over Europe in Present Climate: Near Surface Ozone Sensitivity to External Meteorological Forcing, *Atmos. Chem. Phys.*, 10, 11805–11821, doi: 10.5194/acp-10-11805-2010, 2010.

Kim H. C., Lee P., Ngan F., Tang Y., Yoo H. L., and Pan L.: Evaluation of Modeled Surface Ozone Biases as a Function of Cloud Cover Fraction, *Geosci. Model Dev.*, 8, 2959–2965, doi: 10.5194/gmd-8-2959-2015, 2015.

Kovač-Andrić E., Berna J., Gvozdić V.: Impact of Meteorological Factors on Ozone Concentrations Modelled by Time Series Analysis and Multivariate Statistical Methods, *Ecological Informatics*, 4, 117-122, doi: 10.1016/j.ecoinf.2009.01.002, 2009.

Kuenen J. J. P., Visschedijk A. J. H., Jozwicka M., and Denier van der Gon H. A. C.: TNO-MACC_II Emission Inventory; a Multi-year (2003–2009) Consistent High-resolution European Emission Inventory for Air Quality Modelling, *Atmos. Chem. Phys.*, 14, 10963-10976, doi:10.5194/acp-14-10963-2014, 2014.

Lau K. M., and Weng H.: Climate Signal Detection Using Wavelet Transform: How to Make a Time Series Sing. *Bull. Amer. Meteor. Soc.,* 76, 2391–2402, https://doi.org/10.1175/1520-0477(1995)076<2391:CSDUWT>2.0.CO;2, 1995

Lelieveld J., and Crutzen P. J.: Influence of Cloud and Photochemical Processes on Tropospheric Ozone, Nature, 343, 227–233, doi: 10.1038/343227a0, 1990.

Lelieveld J., Berresheim H., Borrmann S., Crutzen P. J., Dentener F. J., Fischer H., Feichter J., Flatau P. J., Heland J., Holzinger R., Korrmann R., Lawrence M. G., Levin Z., Markowicz K. M., Mihalopoulos N., Minikin A., Ramanathan V., de Reus M., Roelofs G. J., Scheeren H. A., Sciare J., Schlager H., Schultz M., Siegmund P., Steil B., Stephanou E. G., Stier P., Traub M.,Warneke C.,Williams J., and Ziereis H.: Global Air Pollution Crossroads Over the Mediterranean, *Science*, 298, 794–799, doi: 10.1126/science.1075457, 2002.

Lelieveld J., Dentener F. J.: What Controls Tropospheric Ozone?, *Journal of Geophysical Research*, 105, 3531-3551, doi: 10.1029/1999JD901011, 2000.

Li G., Zhang R., Fan J., and Tie X.: Impacts of Black Carbon Aerosol on Photolysis and Ozone, *J. Geophys. Res.*, 110, D23206, doi: 10.1029/2005JD005898, 2005.

Lin X., Trainer M., Liu S. C.: On the Nonlinearity of Tropospheric Ozone, *Journal of Geophysical Research*, 93, 15,879-15,888, doi: 10.1029/JD093iD12p15879, 1988.

Liu S. C., Trainer M., Fehsenfeld F. C., Parrish D. D., Williams E. J., Fahey D. W., Hubler G., Murphy P. C.: Ozone Production in the Rural Troposphere and the Implications for Regional and Global Ozone Distributions, *Journal of Geophysical Research,* 92, 4191-4207, doi: 10.1029/JD092iD04p04191, 1987.

Mar K. A., Ojha N., Pozzer A., and Butler T.: Ozone Air Quality Simulations with WRF-Chem (v3.5.1) over Europe: Model Evaluation and Chemical Mechanism Comparison, *Geosci. Model Dev.*, 9, 3699–3728, doi: 10.5194/gmd-9-3699-2016, 2016.

Marsik F. J., Fischer K. W., McDonald T. D., Samson P. J.: Comparison of Methods for Estimating Mixing Heights Used During the 1992 Atlanta Field Intensive, *Journal of Applied Meteorology*, 34, 1802-1814, doi: https://doi.org/10.1175/1520-0450(1995)034<1802:COMFEM>2.0.CO;2, 1995.

Millán M. M., Mantilla E., Salvador R., Carratalá R., Sanz M. J., Alonso L., Gangioti G., and Navazo M.: Ozone Cycles in the Western Mediterranean Basin: Interpretation of Monitoring Data in Complex Coastal Terrain, J. Appl. Meteorol., 39, 487–508, 1999.

Monks P. S., Archibald A. T., Colette A., Cooper O., Coyle M., Derwent R., Fowler D., Granier C., Law K. S., Mills G. E., Stevenson D. S., Tarasova O., Thouret V., von Schneidemesser E., Sommariva R., Wild O., and Williams M. L.: Tropospheric Ozone and its Precursors From the Urban to the Global, *Atmos. Chem. Phys.*, 15, 8889–8973, doi: 10.5194/acp-15-8889-2015, 2015.

Morrison H., Thompson G., and Tatarskii V.: Impact of Cloud Microphysics on the Development of Trailing Stratiform Precipitation in a Simulated Squall Line: Comparison of One- and Two-moment Schemes. *Monthly Weather Review*, 137, 991–1007, doi:10.1175/2008MWR2556.1, 2009.

Neu J. L., and Prather M. J.: Toward a More Physical Representation of Precipitation Scavenging in Global Chemistry Models: Cloud Overlap and Ice Physics and their Impact on Tropospheric Ozone, *Atmos. Chem. Phys.*, 12, 3289–3310, doi: 10.5194/acp-12-3289-2012, 2012.

Palau J. L., P´erez-Landa G., Di´eguez J. J., Monter C., and Mill´an M. M.: The Importance of Meteorological Scales to Forecast Air Pollution Scenarios on Coastal Complex Terrain, *Atmos. Chem. Phys.,* 5, 2771–2785, doi: 10.5194/acp-5-2771-2005, 2005.

Pfister G. G., Walters S., Emmons L. K., Edwards D. P., and Avise J.: Quantifying the Contribution of Inflow on Surface Ozone Over California During Summer 2008, *J. Geophys. Res.-Atmos.*, 118, 12282–12299, doi: 10.1002/2013JD020336, 2013.

Pindyck R. S., and D. L. Rubinfeld. (1998). Econometric Models and Economic Forecasts. New York, Irwin : McGraw-Hill, Singapore, 388 pp.

Rao S. T., Zurbenko I. G., Neagu R., Porter P. S., Ku J. Y., and Henry R. F.: Space and Time Scales in Ambient Ozone Data, *Bull. Am. Meteorol. Soc.*, 78(10), 2153-2166, doi: 10.1175/1520-0477(1997)078, 1997.

Rasmussen D. J., Fiore A. M., Naik V., Horowitz L. W., McGinnis S. J., Schultz M. G.: Surface Ozone-Temperature Relationships in the Eastern US: A Monthly Climatology for Evaluating Chemistry-Climate Models, *Atmos. Environ.*, 47, 142-153, doi: 10.1016/j.atmosenv.2011.11.021, 2012.

Rastigejev Y., R. Park M. Brenner and Jacob D.: Resolving Inter-continental Pollution Plumes in Global Models of Atmospheric Transport, *J. Geophys. Res.*, 115, D02302, doi: 10.1029/2009JD012568, 2010.

Richards N. A. D., Arnold S. R., Chipperfield M. P., Miles G., Rap A., Siddans R., Monks S., and Hollaway M. J.: The Mediterranean Summertime Ozone Maximum: Global Emission Sensitivities and Radiative Impacts, *Atmos. Chem. Phys.*, 13, 2331–2345, doi: 10.5194/acp-13-2331-2013, 2013.

Safieddine S., Boynard A., Coheur P. F., Hurtmans D., Pfister G., Quennehen B., Thomas J. L., Raut J. C., Law K. S., Klimont Z., Hadji-Lazaro G., George M., and Clerbaux C.: Summertime Tropospheric Ozone Assessment Over the Mediterranean Region Using the Thermal Infrared IASI/MetOp Sounder and the WRF-Chem Model, *Atmos. Chem. Phys.*, 14, 10119–10131, doi: 10.5194/acp-14-10119-2014, 2014.

Saikawa E., Kim H., Zhong M., Avramov A., Zhao Y., Janssens-Maenhout G., Kurokawa J. I., Klimont Z., Wagner F., Naik V., Horowitz L. W., and Zhang Q.: Comparison of Emissions Inventories of Anthropogenic Air Pollutants and Greenhouse Gases in China, *Atmos. Chem. Phys.*, 17, 6421,

Saitanis C. J.: Background Ozone Monitoring and Phytodetection in the Greater Rural Area of Corinth-Greece. *Chemosphere*, 51, 913–923, doi: 10.5194/acp-17-6393-2017, 2017.

Salcedo R. L. R., Alvim Ferraz M. C. M., Alves C. A., and Martins F. G.: Time Series Analysis of Air Pollution Data. *Atmos. Environ.*, 33, 2361–2372, doi: 10.1002/clen.201400461, 1999.

Séférian R., Gehlen M., Bopp L., Resplandy L., Orr J. C., Marti O., Dunne J. P., Christian J. R., Doney S. C., Ilyina T., Lindsay K., Halloran P. R., Heinze C., Segschneider J., Tjiputra J., Aumont O., and Romanou A.: Inconsistent Strategies to Spin up Models in CMIP5: Implications for Ocean Biogeochemical Model Performance Assessment, *Geosci. Model Dev.*, 9, 1827-1851, doi: 10.5194/gmd-9-1827-2016, 2016.

Seinfeld J. H., and Pandis S. N. (1998). Atmospheric Chemistry and Physics: From Air Pollution to Climate Change, 2nd ed., New York, John Wiley.

Seo J., Youn D., Kim J. Y., and Lee H.: Extensive Spatiotemporal Analyses of Surface Ozone and Related Meteorological Variables in South Korea for the Period 1999–2010 Atmos. Chem. Phys., 14, 6395–6415, doi:10.5194/acp-14-6395-2014, 2014.

Sillman S.: The Relation between Ozone, NOx and Hydrocarbons in Urban and Polluted Rural Environments, *Atmos. Environ.*, 33, 1821- 1845, doi: 10.1016/S1352-2310(98)00345-8, 1999.

Sillman S., He D., Pippin M., Daum P., Kleinman L., Lee J. H., Weinstein-Lloyd J.: Model Correlations for Ozone, Reactive Nitrogen and Peroxides for Nashville in Comparison with Measurements: Implications for NOx-hydrocarbon Sensitivity, *Journal of Geophysical Research*, 103, 22, 629-22,644, doi: 10.1029/98JD00349, 1998.

Simpson, D.: Long-period Modeling of Photochemical Oxidants in Europe – Model-calculations For July 1985, Atmos. Environ. A-Gen., 26, 1609–1634, doi: 10.1016/0960-1686(92)90061- O, 1992.

Solazzo E. and Galmarini S.: Error Apportionment for Atmospheric Chemistry-Transport Models – a New Approach to Model Evaluation, *Atmos. Chem. Phys.*, 16, 6263–6283, doi: 10.5194/acp-16-6263-2016, 2016.

Solazzo E., Bianconi R., Vautard R., Appel K. W., Moran M. D., Hogrefe C., Bessagnet B., Brandt J., Christensen J. H., Chemel C., Coll I., van der Gon H. D., Ferreira J., Forkel R., Francis X. V., Grell G., Grossi P., Hansen A. B., Jeriˇcevi´c A., Kraljevi´c L., Miranda A. I., Nopmongcol U., Pirovano

Stockwell W. R., Lawson C. V. , Saunders E., and Goliff W. S.: A Review of Tropospheric Atmospheric Chemistry and Gas-Phase Chemical Mechanisms for Air Quality Modeling, *Atmosphere*, 3, 1-32; doi: 10.3390/atmos3010001, 2012.

Solazzo E., Bianconi R., Vautard R., Appel K. W., Moran M. D., Hogrefe C., Bessagnet B., Brandt J., Christensen J. H., Chemel C., Coll I., van der Gon H. D., Ferreira J., Forkel R., Vazhappilly Francis X., Grell G., Grossi P., Hansen A. B., Jericevic A., Kraljevic L., Miranda A. I., Nopmongcol U., Pirovano G., Prank M., Riccio A., Sartelet K. N., Schaap M., Silver J. D., Sokhi R. S., Vira J., Werhahn J., Wolke R., Yarwood G., Zhang J., Rao S. T., and Galmarini S.: Model Evaluation and Ensemble Modelling of Surface-level Ozone in Europe and North America in the Context of {AQMEII}, Atmos. Environ., 53, 60–74, doi: 10.1016/j.atmosenv.2012.01.003, 2012.

Stull R. B. (1988). An Introduction to Boundary Layer Meteorology, Dordrecht, Kluwer Acad.

Tarasova O. A., and Karpetchko A. Y.: Accounting for Local Meteorological Effects in the Ozone Time-Series of Lovozero (Kola Peninsula), *Atmos. Chem. Phys.*, 3, 941–949, doi: 10.5194/acp-3-941-2003, 2003.

Theil H. (1961). Economic Forecasts and Policy, Amsterdam, North-Holland Pub. Co.

Thornton J. A., Wooldridge P. J., Cohen R. C., Martinez M., Harder H., Brune W. H., Williams E. J., Roberts J. M., Fehsenfeld F. C., Hall S. R., Shetter R. E., Wert B. P., and Fried A.: Ozone Production Rates As a Function of NOx Abundances and HOx Production Rates in the Nashville Urban Plume, *Journal of Geophysical Research*, 107, ACH 7-1-ACH 7-17, doi: 10.1029/2001JD000932, 2002.

Tie X., Madronich S., Walters S., Zhang R., Rasch P., and Collins W.: Effect of Clouds on Photolysis and Oxidants in the Troposphere, *J. Geophys. Res.*, 108(D20), 4642, doi: 10.1029/2003JD003659, 2003

Tilmes S. , Lamarque J. F., Emmons L. K., Conley A., Schultz M. G., Saunois M., Thouret V., Thompson A. M., Oltmans S. J., Johnson B., and Tarasick D.: Ozonesonde Climatology Between 1995 and 2009: Description, Evaluation and Applications, *Atmos. Chem. Phys. Discuss.*, 11, 28747–28796, doi:10.5194/acpd-11-28747-2011, 2011.

Trainer M., Parrish D. D., Buhr M. P., Norton R. B., Fehsenfeld F. C., Anlauf K. G., Bottenheim J. W., Tang Y. Z., Wiebe H. A., Roberts J. M., Tanner R. L., Newman L., Bowersox V. C., Maugher J. M., Olszyna K. J., Rodgers M. O., Wang T., Berresheim H., Demerjian K.: Correlation of Ozone With NOy in Photochemically Aged Air, *Journal of Geophysical Research*, 98, 2917-2926, doi: 10.1029/92JD01910, 1993.

Tong, N. Y. O., Leung D. Y. C., and Liu C. H.: A Review on Ozone Evolution and its Relationship with Boundary Layer Characteristics in Urban Environments, *Water Air Soil Pollut.*, 214, 13–36, doi: 10.1007/s11270-010-0438-5, 2011.

Tuccella, P., Curci, G., Visconti, G., Bessagnet, B., Menut, L., and Park, R. J.: Modeling of Gas and Aerosol with WRF/Chem Over Europe: Evaluation and Sensitivity Study, *J. Geophys. Res.- Atmos.*, 117, D03303, doi:10.1029/2011JD016302, 2012.

Val Martin M., Heald C. L., Arnold S. R.: Coupling Dry Deposition to Vegetation Phenology in the Community Earth System Model: Implications for the Simulation of Surface O_3, *Geophysical Research Letters*, 41, 2988-2996, doi: 10.1002/2014GL059651, 2014.

Van Loon M., Vautard R., Schaap M., Bergström R., Bessagnet B., Brandt J., Builtjes P. J. H., Christensen J. H., Cuvelier C., Graff A., Jonson J. E., Krol M., Langner J., Roberts P., Rouil L., Stern R., Tarrasón L., Thunis P., Vignati E., White L., Wind P.: Evaluation of Long-term Ozone Simulations From Seven Regional Air Quality Models and their Ensemble Average, *Atmospheric Environment*, 41, 2083-2097, doi: 10.1016/j.atmosenv.2006.10.073, 2007.

Velchev K., Cavalli F., Hjorth J., Marmer E., Vignati E., Dentener F., and Raes F.: Ozone over the Western Mediterranean Sea – results from Two Years of Shipborne Measurements, Atmos. Chem. Phys., 11, 675–688, doi:10.5194/acp-11-675-2011, 2011.

Wałaszek K., Kryza M., Szymanowski M., Werner M., Ojrzyńska H.: Sensitivity Study of Cloud Cover and Ozone Modeling to Microphysics Parameterization, *Pure Appl. Geophys.*, 174, 491-510, doi: 10.1007/s00024-015-1227-2, 2017.

Wang Y., Zhang Y., Hao J., and Luozone M.: Seasonal and Spatial Variability of Surface Ozone Over China: Contributions From Background and Domestic Pollution, *Atmos. Chem. Phys.*, 11, 3511–3525, doi: 10.5194/acp-11-3511-2011, 2011.

Wentworth G. R., Murphy J. G., Sills D. M. L.: Impact of Lake Breezes on Ozone and Nitrogen Oxides in the Greater Toronto Area, *Atmospheric Environment*, 109, 52-60, doi: 10.1016/j.atmosenv.2015.03.002, 2015.

Wesely M. L. and Hicks B. B.: A Review of the Current Status of Knowledge on Dry Deposition, *Atmos. Environ.*, 34, 2261–2282, doi: 10.1016/S1352-2310(99)00467-7, 2000.

Wesely M. L.: Parameterization of Surface Resistance to Gaseous Dry Deposition in Regional-scale Numerical Models, *Atmos. Environ.*, 23, 1293–1304, doi: 10.1016/0004-6981(89)90153-4, 1989.

Wiedinmyer C., Akagi S. K., Yokelson R. J., Emmons L. K., Al-Saadi J. A., Orlando J. J., and Soja A. J.: The Fire Inventory from NCAR (FINN): A High Resolution Global Model to Estimate the Emissions from Open Burning, *Geosci. Model Dev.*, 4, 625-641, doi: 10.5194/gmd-4-625-2011, 2011.

Wild O.: Modelling the Global Tropospheric Ozone Budget: Exploring the Variability in Current Models, *Atmos. Chem. Phys.*, 7, 2643–2660, doi: 10.5194/acp-7-2643-2007, 2007.

Wilks D. S. (1995). Statistical Methods in Atmospheric Sciences. International Geophysics Series, USA, Academic Press.

Žabkar R., Honzak L., Skok G., Forkel R., Rakovec J., Ceglar A., and Žagar N.: Evaluation of the High Resolution WRF-Chem (v3.4.1) Air Quality Forecast and its Comparison with Statistical Ozone Predictions, *Geosci. Model Dev.*, 8, 2119–2137, doi: 10.5194/gmd-8-2119-2015, 2015.

Zhang J., and Rao S. T.: The Role of Vertical Mixing in the Temporal Evolution of Ground-level Ozone Concentrations, *J. Appl. Meteorol.*, 38, 1674 – 1691, doi: 10.1175/MWR-D-16-0421.1, 2017.

Zhang Y.: Online-coupled Meteorology and Chemistry Models: History, Current Status, and Outlook, *Atmos. Chem. Phys.*, 8, 2895–2932, doi: 10.5194/acp-8-2895-2008, 2008.

Zurbenko, I. G., 1986: *The Spectral Analysis of Time Series.* North Holland, 241 p.

Appendices

Appendix A

Part of the results for the **ACE** station

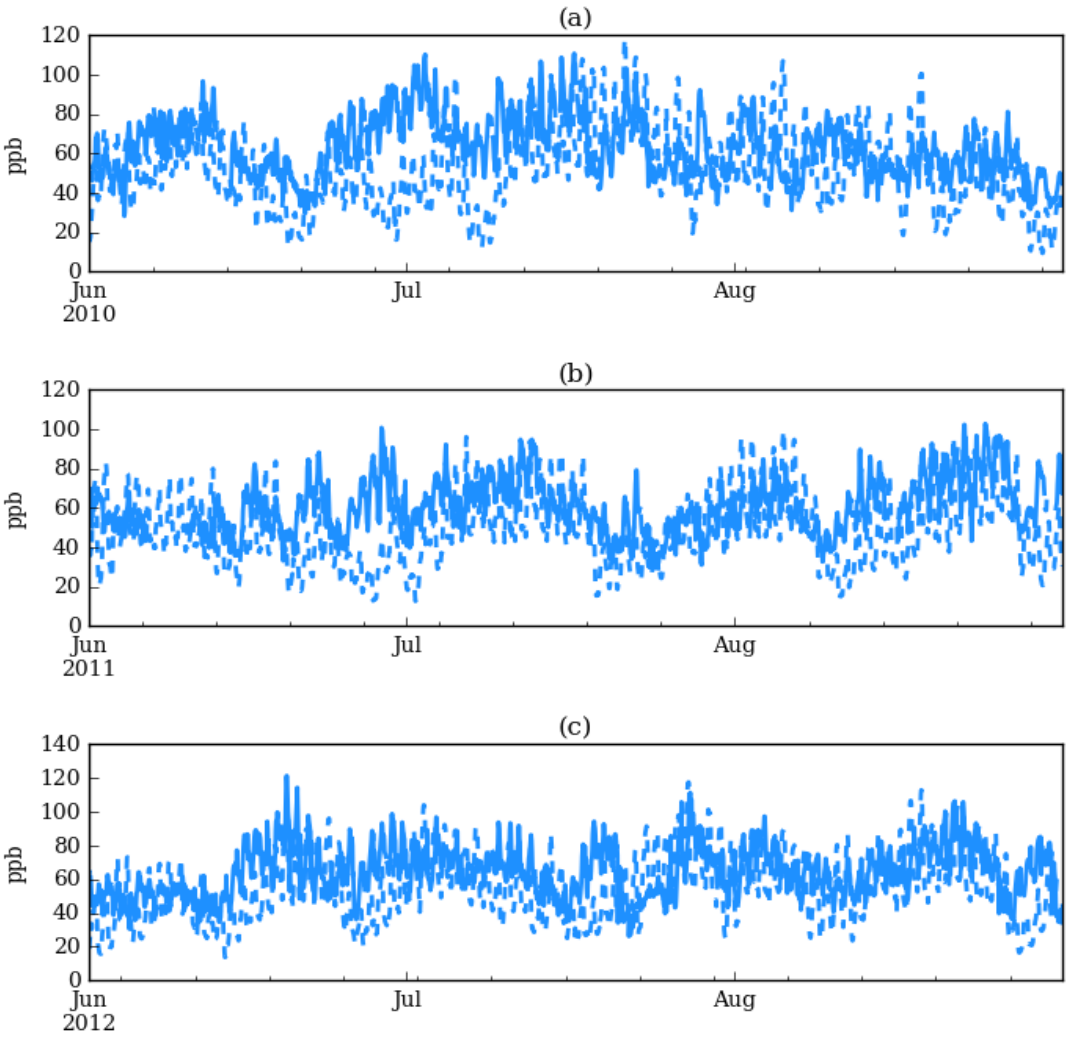

Figure A1. Hourly O_3^{surf} time series for summers **(a)** 2010, **(b)** 2011 and **(c)** 2012 at the ACE station. The solid and dashed lines are used for the observation and the WRF-Chem model simutaion datasets, respectively.

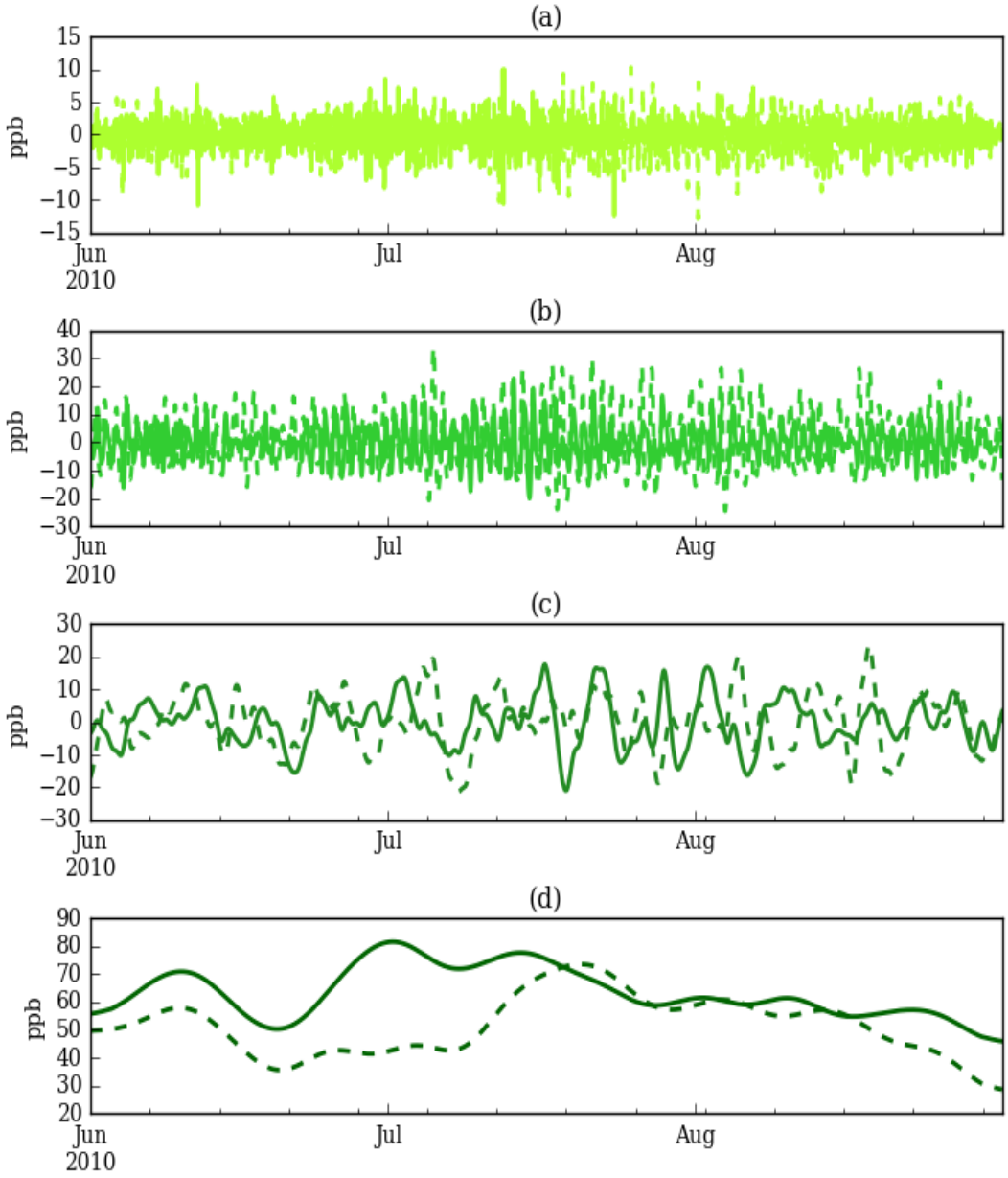

Figure A2. Hourly time series for the **(a)** ID, **(b)** DU, **(c)** SY and **(d)** BL spectral components of O_3^{surf} at the ACE station for summer 2010. The solid and dashed lines are used for the observation and the WRF-Chem model simutaion datasets, respectively.

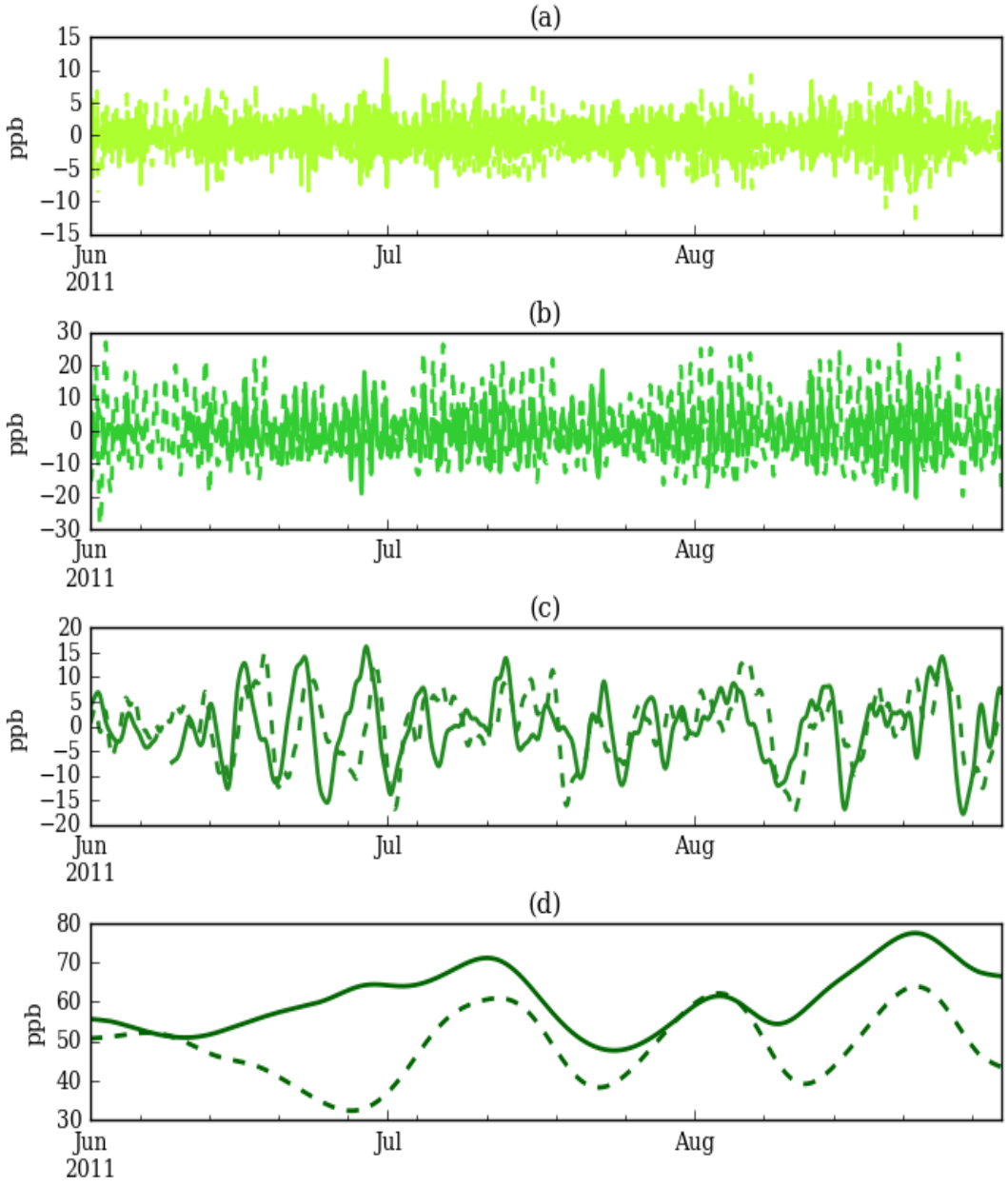

Figure A3. As Figure A2, but for 2011.

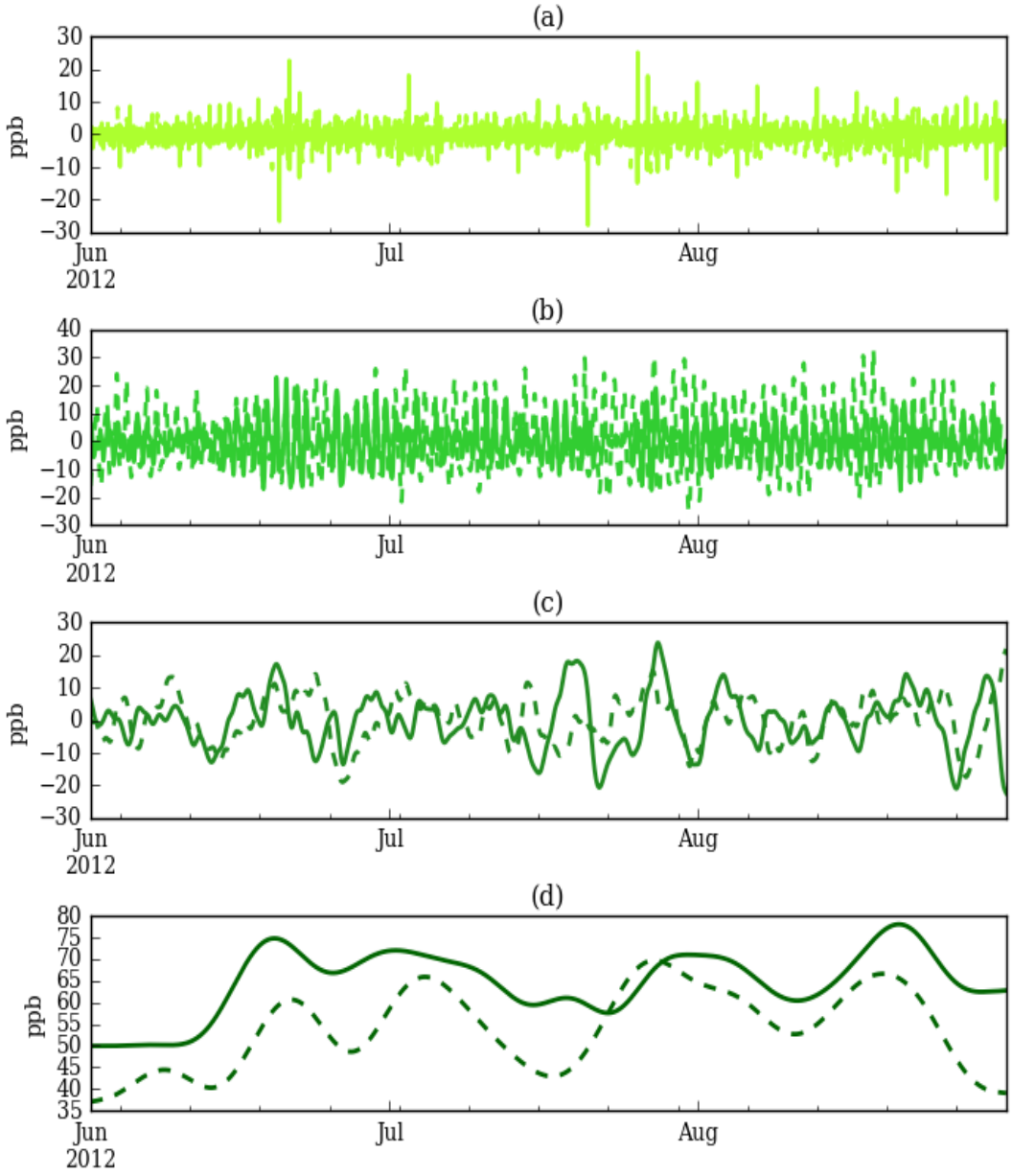

Figure A4. As Figure A2, but for 2012.

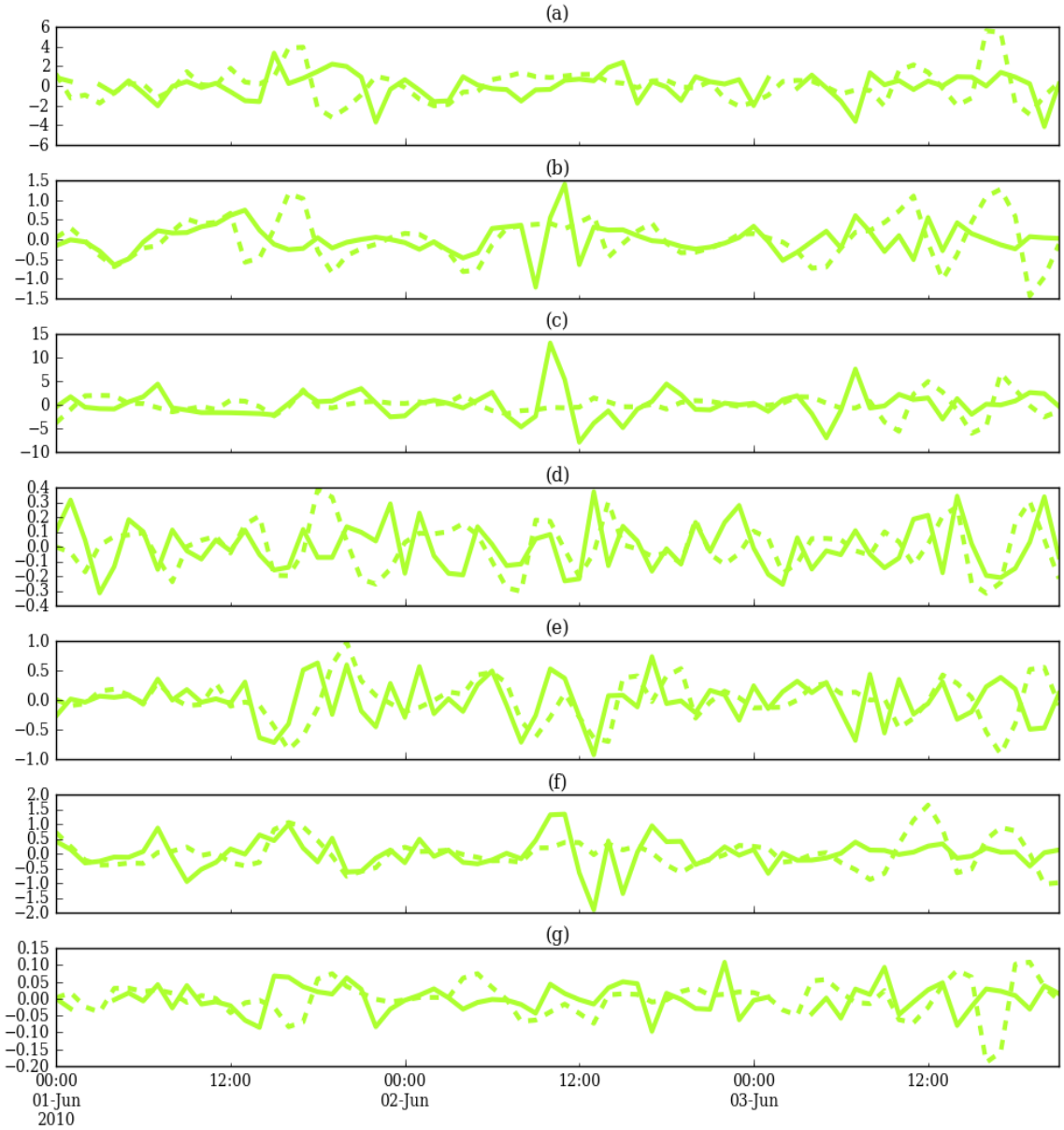

Figure A5. Time series for the ID component of **(a)** O_3^{surf}, **(b)** AT, **(c)** RH, **(d)** SP, **(e)** U, **(f)** V and **(g)** NO_x at the ACE station. The solid and dashed lines are used for the observation and the WRF-Chem model simutaion datasets, respectively.

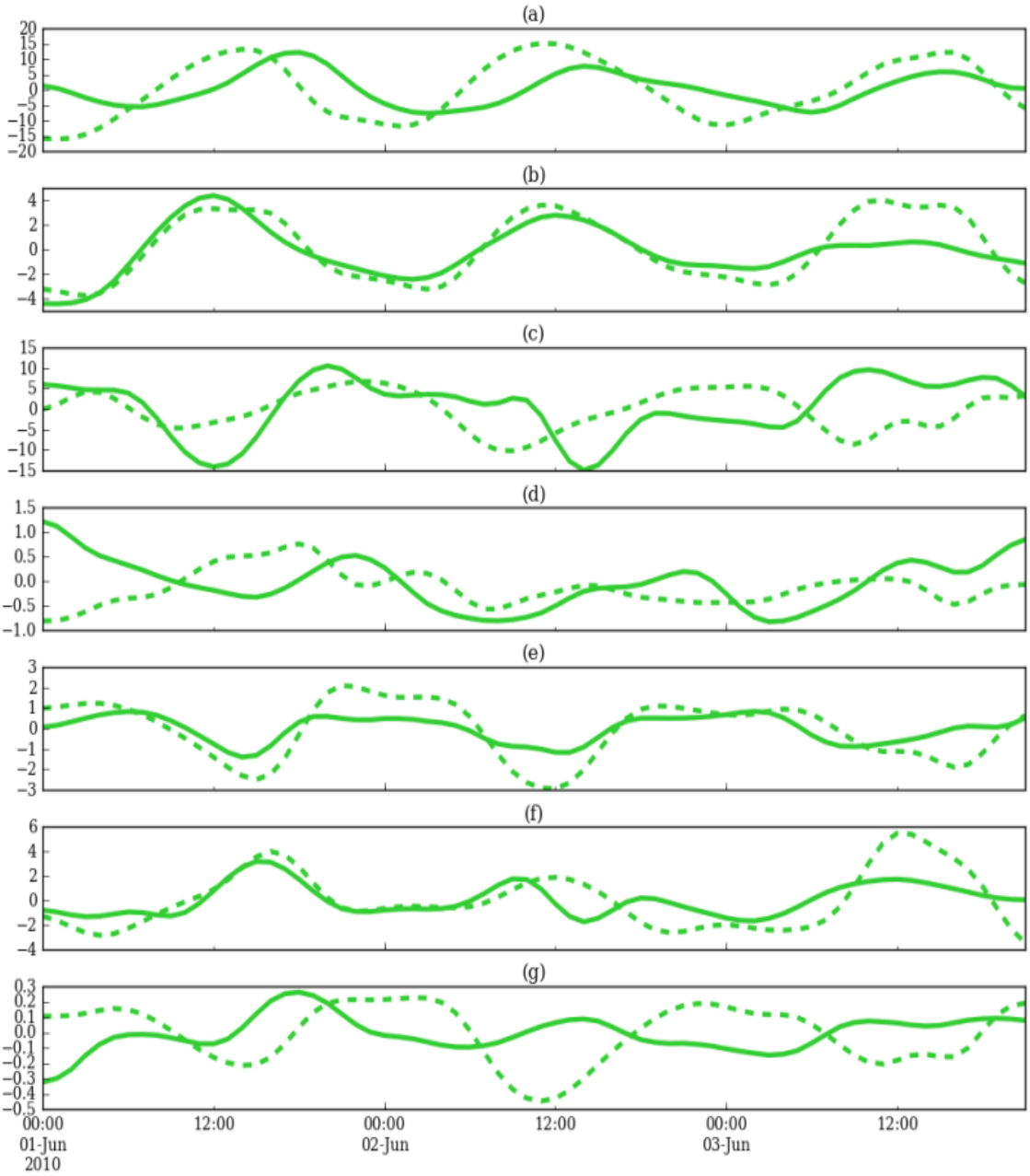

Figure A6. As Figure A5, but for the DU component.

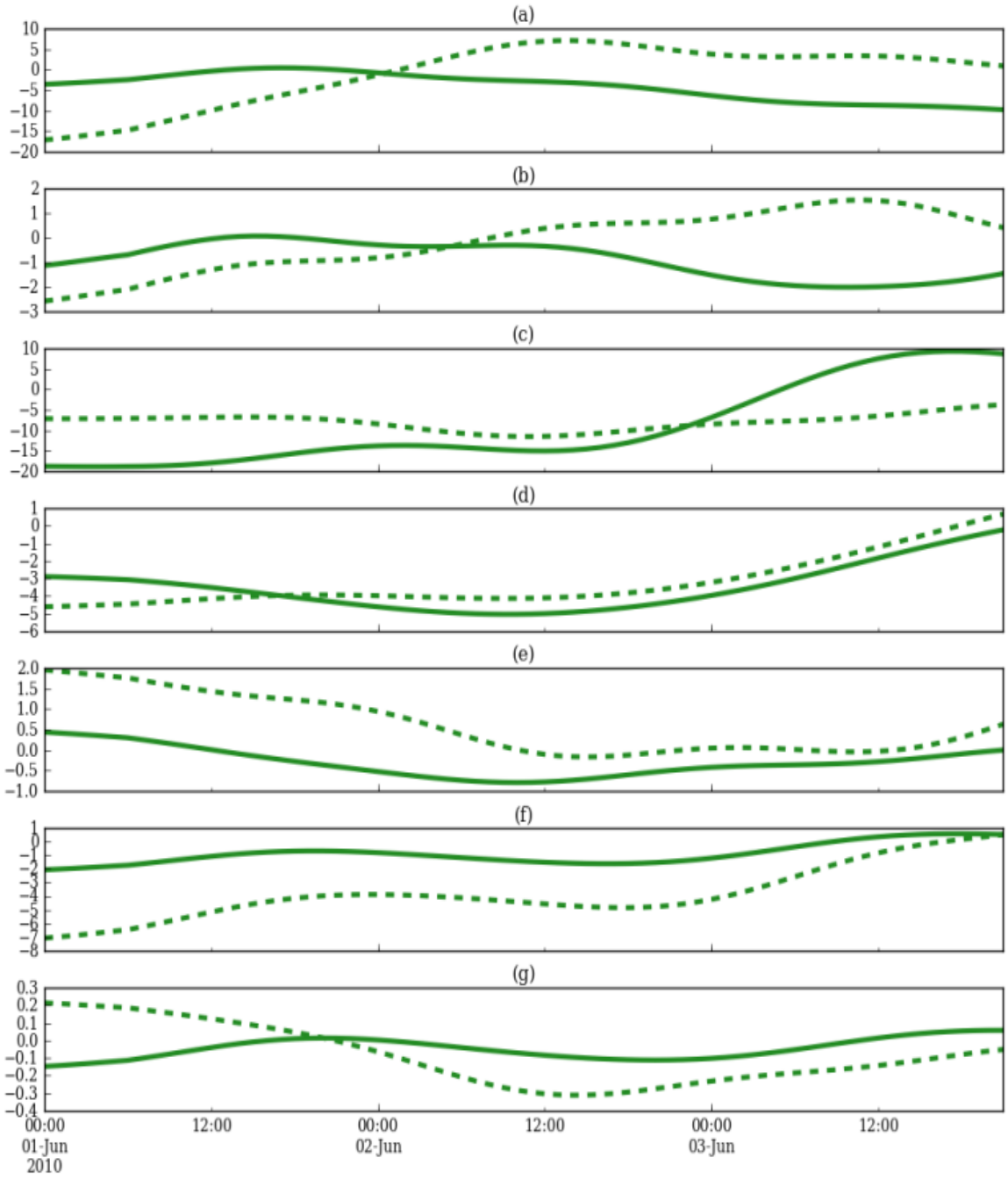

Figure A7. As Figure A5, but for the SY component

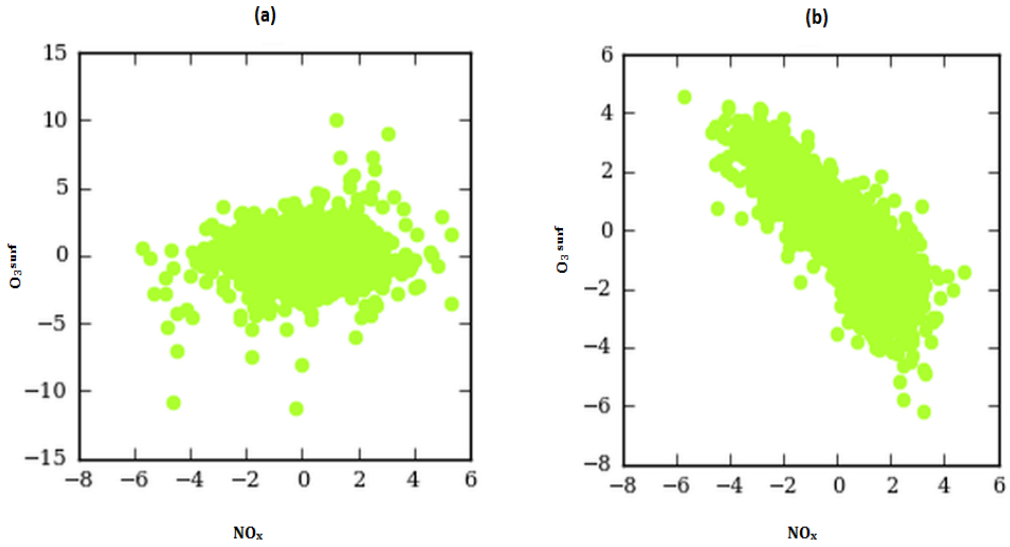

Figure A8. The ID component of the O_3 and NO_x for **(a)** the observationas **(b)** the WRF-Chem model simulation at the ACE station.

Table A1. Standard deviation for the ID spectral component of the variables at the ACE station (averaged over summers 2010, 2011, and 2012).

variables	O_3^{surf}	AT	RH	SP	U	V	NO_x
the observations	2.48	0.32	2.61	0.15	0.32	0.42	0.4
the WRF-Chem model simulation	2.25	0.63	1.92	0.12	0.42	0.53	1.82

Table A2. As in Table A1, but for the DU component.

variables	O_3^{surf}	AT	RH	SP	U	V	NO_x
the observations	6.07	1.43	5.61	0.48	0.52	1.25	0.56
the WRF-Chem model simulation	10.58	3.93	8.61	0.6	1.28	3.14	5.84

Table A3. As in Table A1, but for the SY component.

variables	O_3^{surf}	AT	RH	SP	U	V	NO_x
the observations	7.44	1.31	6.71	2.38	0.52	0.68	0.57
the WRF-Chem model simulation	7.4	1.39	6.11	2.59	1.4	1.4	2.89

Table A4.5 The variation of O_3^{surf} (*ppb*) per one unit change of the spectral components of variables at the ACE station.

datasets	the observations			the WRF-Chem model simulation		
spectral components	**ID**	**DU**	**SY**	**ID**	**DU**	**SY**
AT	-0.25	-1.15	2.99	-0.43	0.61	3.51
RH	-0.4	-0.09	-0.02	-0.14	0.09	0.27
SP	-1.2	-1.37	-0.13	-1.66	-2.86	-0.47
U	0.37	1.98	-0.24	0.69	0.35	0.25
V	-0.09	0.56	1.77	0.29	0.7	1.23
NO_x	0.22	5.07	4.76	-0.89	-0.98	-1.30

Table A4. As in Table 4.8**,** but for using ($m_{i\text{-}opt}$, $K_{i\text{-}opt}$) instead of ($m_{i\text{-}ref}$, $k_{i\text{-}ref}$) in the KZ filter. The values within the parenthesis show the percentage of the errors with respect to the sum of the model variance error (= 226.97).

spectral components	ID	DU	SY	SUM
variance error	11.09 (4.89%)	132.1 (58.2%)	83.79 (36.92%)	226.97 (100%)
explained error	4.94 (2.1%)	94.56 (41.66%)	31.83 (14.02%)	131.32 (57.86%)
unexplained error	6.15 (2.7%)	37.54 (16.54%)	52.96 (22.89)	95.65 (42.14%)

Table A5. As in Table 4.4**,** but for using ($m_{i\text{-}opt}$, $K_{i\text{-}opt}$) instead of ($m_{i\text{-}ref}$, $k_{i\text{-}ref}$) in the KZ filter.

datasets	the observation			the WRF-Chem model simulation		
spectral components	**ID**	**DU**	**SY**	**ID**	**DU**	**SY**
R^2	0.01	0.32	0.68	0.61	0.86	0.73
a_1	-0.03	-0.27	0.53	-0.12	0.22	0.67
a_2	-0.04	-0.08	-0.02	-0.12	0.07	0.22
a_3	-0.07	-0.1	-0.04	-0.09	-0.16	-0.17
a_4	0.04	0.17	-0.02	0.13	0.04	0.04
a_5	-0.01	0.11	0.18	0.07	0.21	0.25
a_6	0.04	0.47	0.36	-0.72	-0.54	-0.53

Appendix B

Part of the results for the **ALC** station

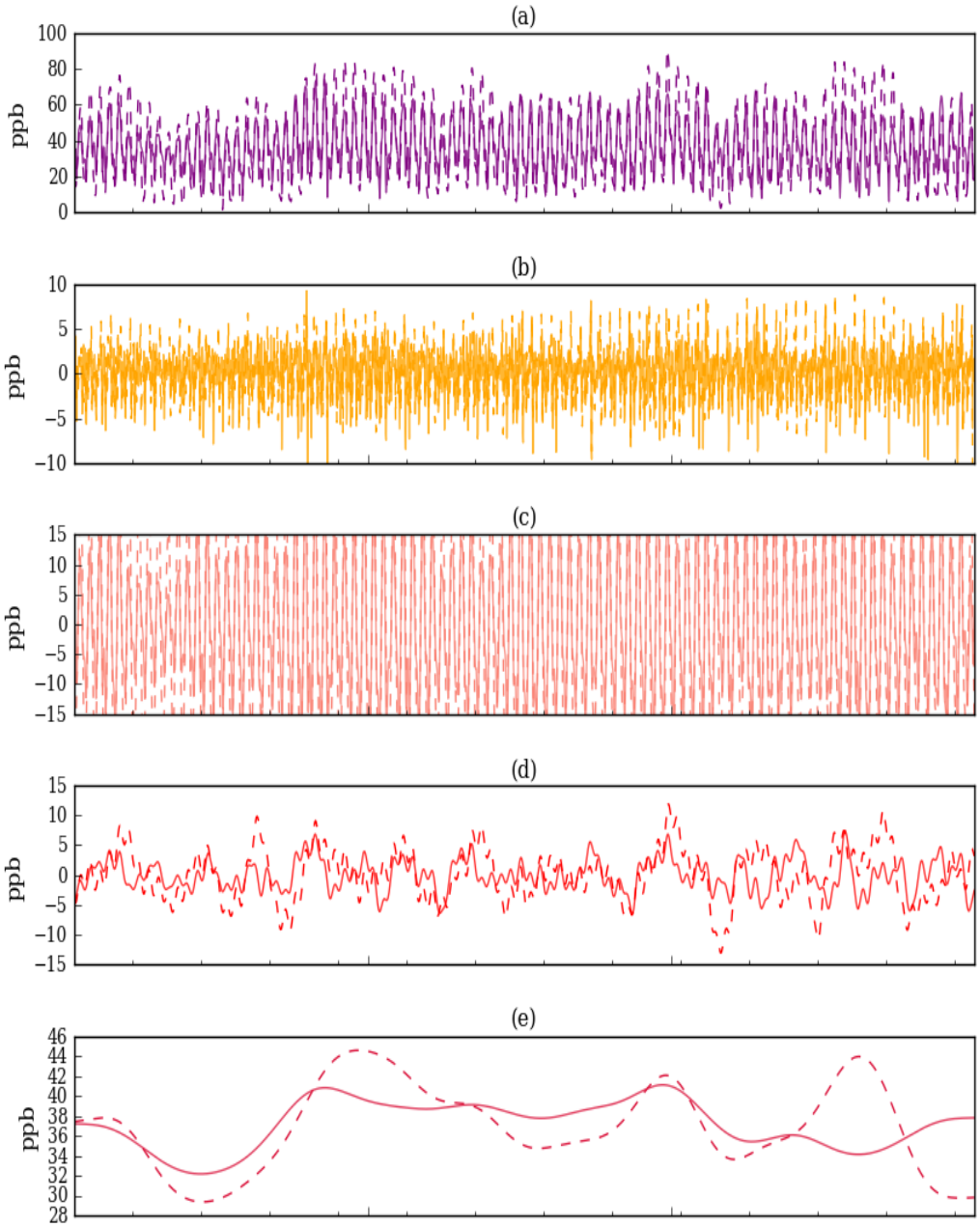

Figure B1. As Figure A1, but for the ALC station. Time series of **(a)** ORG, **(b)** ID, **(c)** DU, **(d)** SY, and **(e)** BL spectral components for the O_3^{surf} at the ALC station. The values are hourly averaged over summers of 2010, 2011, and 2012.

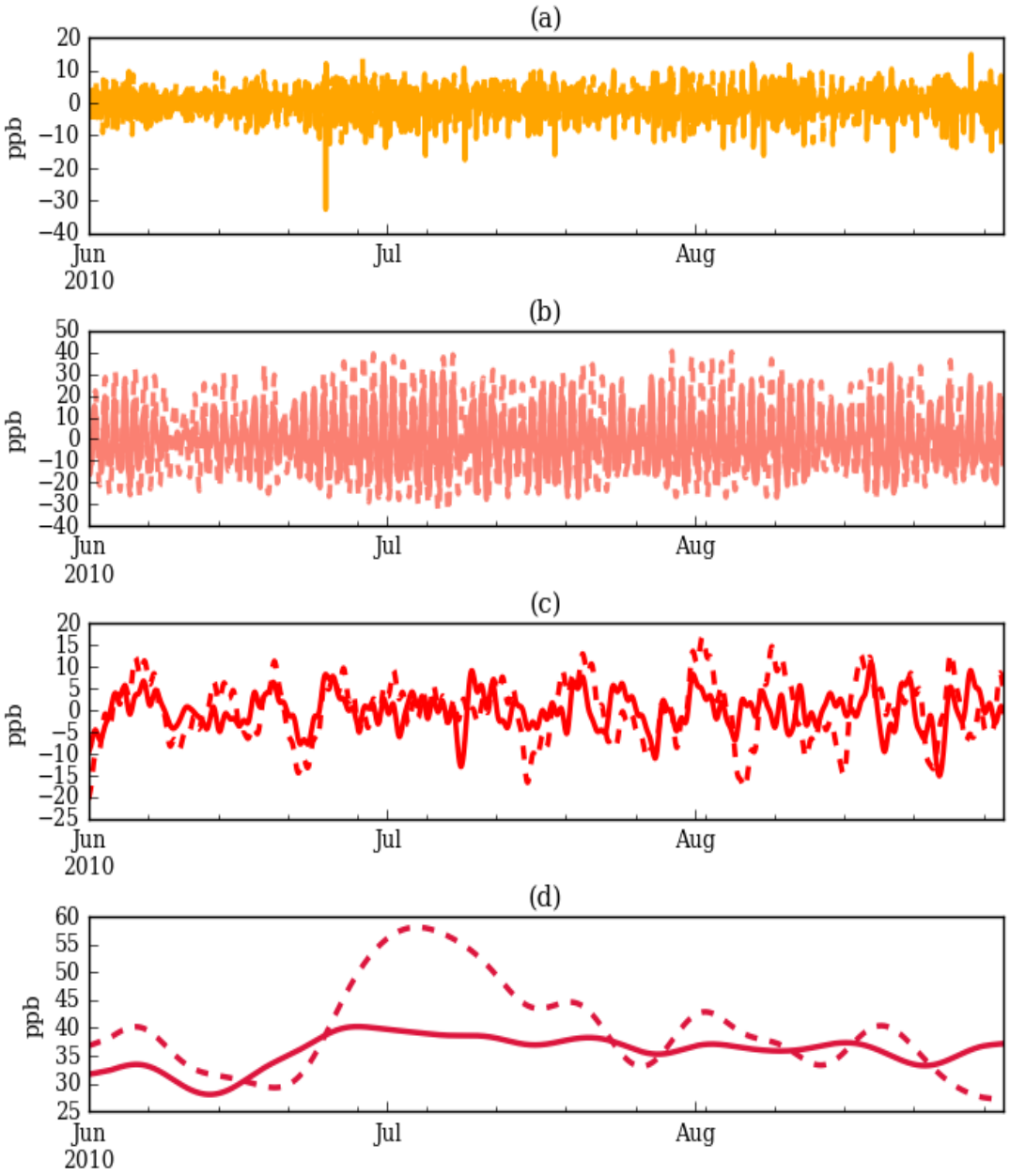

Figure B2. As Figure A2 but for the ALC station. Hourly time series for the **(a)** ID, **(b)** DU, **(c)** SY and **(d)** BL spectral components of O_3^{surf} at the ALC station for summer 2010. The solid and dashed lines are used for the observation and the WRF-Chem model simutaion datasets, respectively.

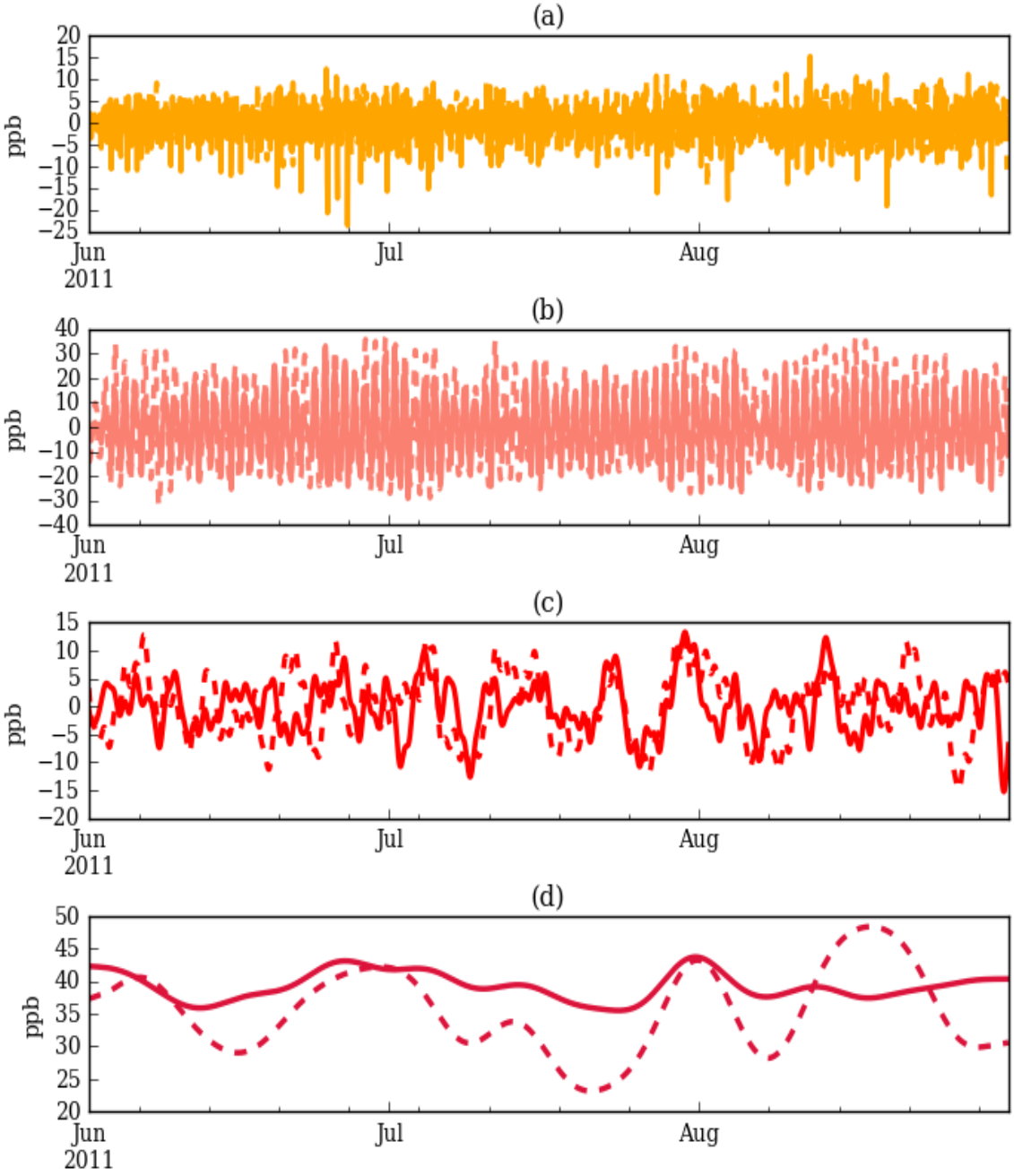

Figure B3. As Figure B2, but for the summer 2011.

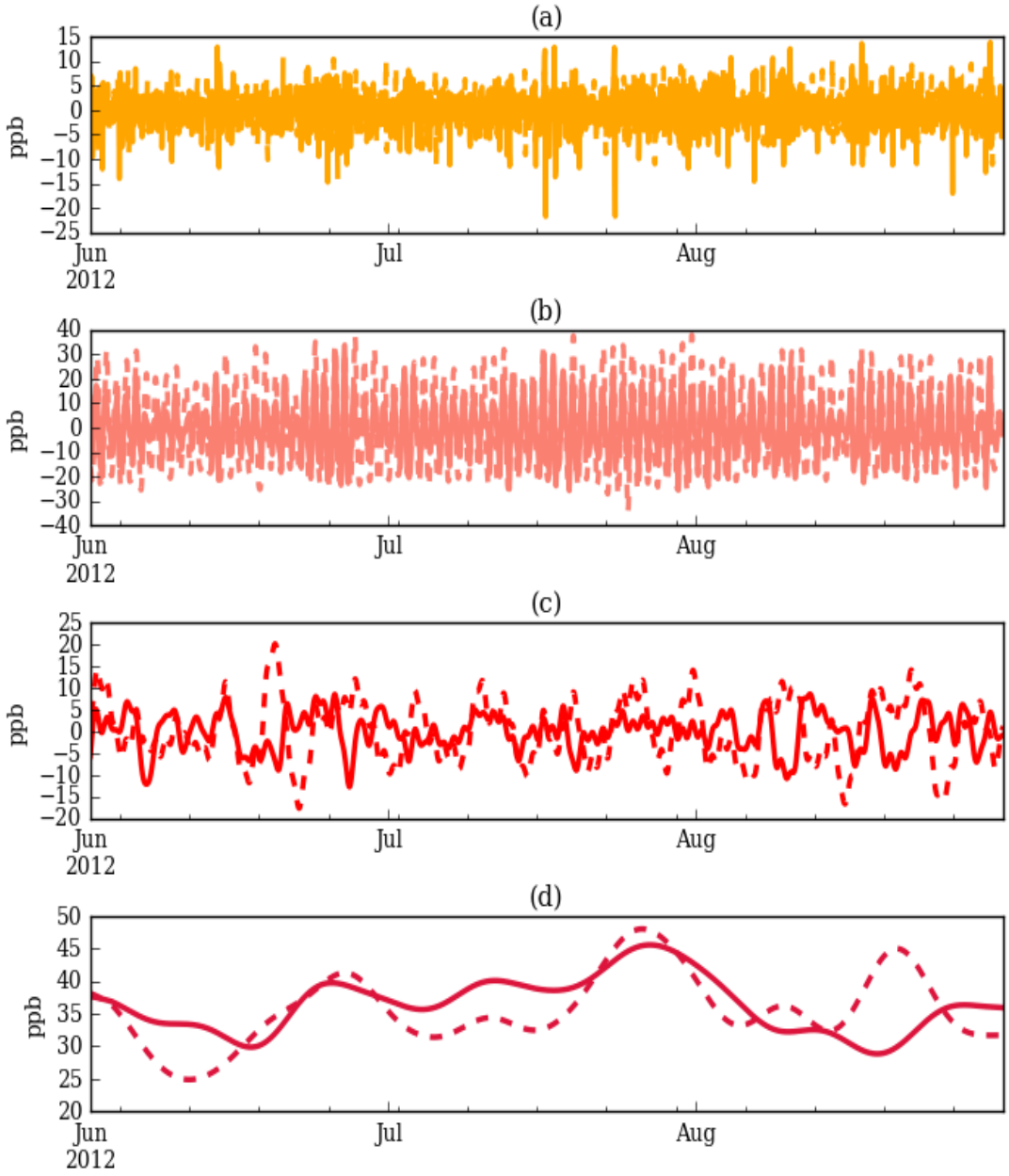

Figure B4. As Figure B2, but for the summer 2012.

Table B1. As in Table A1, but for the ALC station. Standard deviation for the ID spectral component of the variables at the ALC station (averaged over summers 2010, 2011, and 2012).

variables	O_3^{surf}	AT	RH	SP	U	V	NO_x
the observations	3.6	0.54	2.79	0.21	0.49	0.49	5.88
the WRF-Chem model simulation	3.08	0.7	1.56	0.12	0.4	0.4	4.97

Table B2. As in Table B1, but for the DU component.

variables	O_3^{surf}	AT	RH	SP	U	V	NO_x
the observations	11.78	3.87	11.83	0.8	1.28	0.94	9.96
the WRF-Chem model simulation	17.45	6.17	10.46	1.07	1.78	1.28	22.68

Table B3. As in Table B1, but for the SY component.

variables	O_3^{surf}	AT	RH	SP	U	V	NO_x
the observations	4.34	1.77	6.6	2.16	1.29	0.7	5.57
the WRF-Chem model simulation	6.68	1.64	4.45	2.24	1.23	1.01	5.25

Table B4.10. O_3^{surf} variation (ppb) per one unit change of the spectral component of the measured and the simulated variables (AT, RH, SP, U, V, and NO_x) at the ALC station.

datasets	the observations			the WRF-Chem model simulation		
spectral components	**ID**	**DU**	**SY**	**ID**	**DU**	**SY**
AT	1.02	1.69	0.15	0.3	0.9	2.7
RH	-0.04	-0.07	-0.14	-0.01	-0.01	0.35
SP	-0.03	-0.13	0.13	-3.23	-0.8	0.11
U	0.23	0.08	-0.59	0.05	-0.17	1.05
V	0.13	0.8	0.84	-0.4	0.32	0.69
NO_x	-0.45	-0.5	-0.46	-0.47	-0.48	-0.39

Appendix C

Part of the results for the **rural** stations

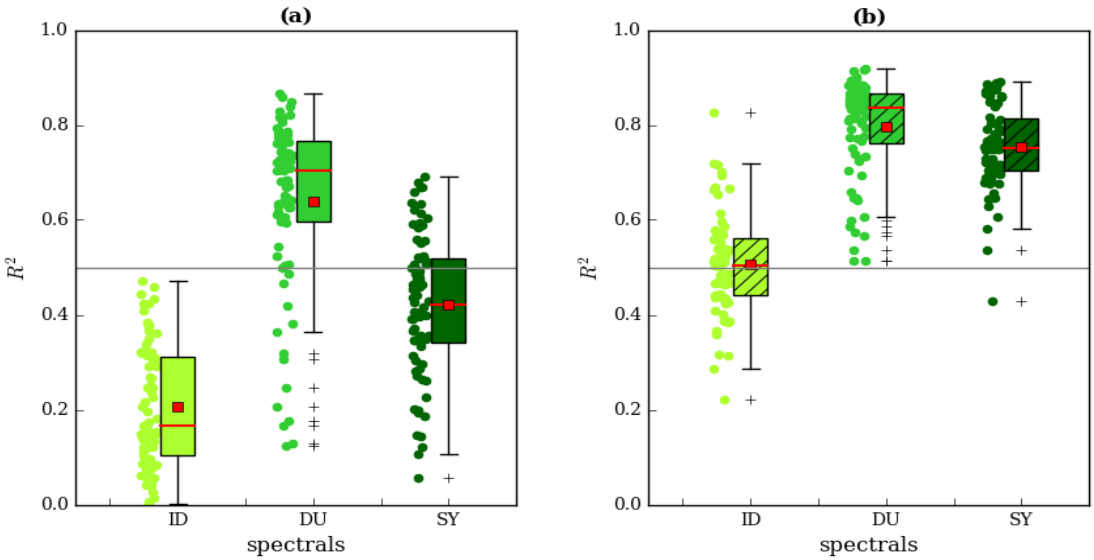

Figure C1. As Figure 4.10, but for using ($m_{i\text{-}opt}$, $K_{i\text{-}opt}$) instead of ($m_{i\text{-}ref}$, $k_{i\text{-}ref}$) in the KZ filter.

Table C1. As in Table A1, but for all rural stations. Standard deviation for the ID spectral component of the variables at the rural stations (averaged over summers 2010, 2011, and 2012).

variables	O_3^{surf}	AT	RH	SP	U	V	NO_x
the observations	2.35	0.43	2.66	0.2	0.46	0.47	1.41
the WRF-Chem model simulation	2.06	0.67	2.17	0.11	0.38	0.4	1.11

Table C2. As in Table C1, but for the DU component.

variables	O_3^{surf}	AT	RH	SP	U	V	NO_x
the observations	8.68	2.71	10.58	0.67	1.02	1.01	1.9
the WRF-Chem model simulation	9.37	4.6	11.11	0.65	1.59	1.77	4.3

Table C3. As in Table C1, but for the SY component.

variables	O_3^{surf}	AT	RH	SP	U	V	NO_x
the observations	5.13	1.89	7.3	2.51	1.02	1.05	1.16
the WRF-Chem model simulation	8.14	2	7.47	2.5	1.5	1.71	2.41

Table C4. As in Table 4.14**,** but for using ($m_{i\text{-}opt}$, $K_{i\text{-}opt}$) instead of ($m_{i\text{-}ref}$, $k_{i\text{-}ref}$) in the KZ filter. The values within the parenthesis show the percentage of the errors with respect to the sum of the model variance error (= 133.81).

stations type	errors	ID	DU	SY	SUM
rural	**variance error**	9.9 (7.4%)	70.37 (52.59%)	53.55 (40.01%)	133.81 (100%)
	explained error	3.87 (2.89%)	20.25 (15.13%)	35.86 (26.8%)	59.99 (44.83%)
	unexplained error	6.03 (4.5%)	50.12 (37.45%)	17.68 (13.21%)	73.83 (55.17%)

Appendix D

Part of the results for the **urban** stations

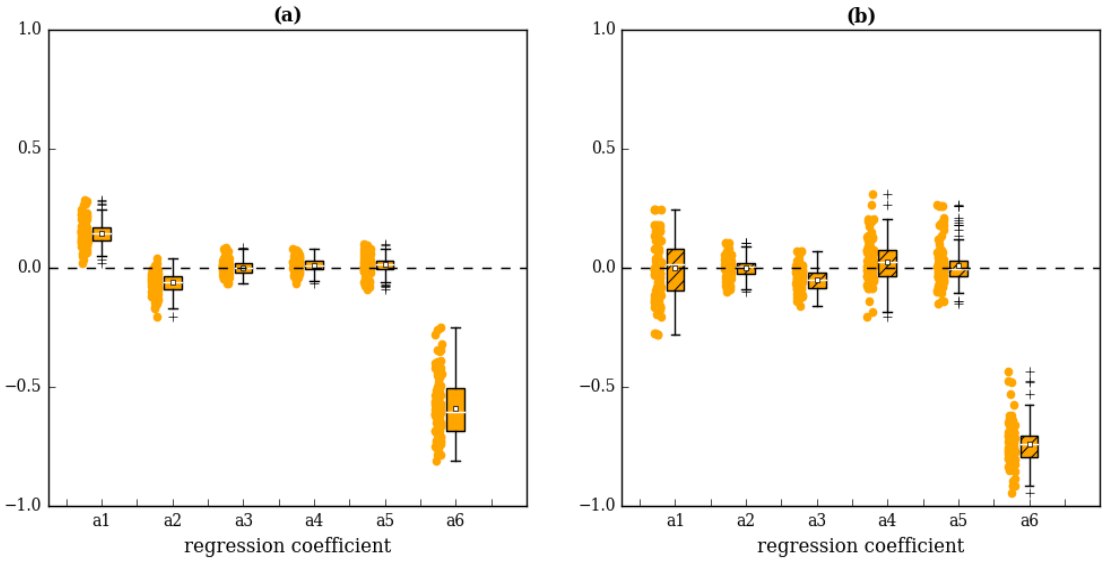

Figure D1. As Figure 4.12, but for urban sites. The regression coefficients (a_1, a_2, a_3, a_4, a_5, and a_6) of the MLR model, where AT, RH, SP, U, V, and NO_x used as predictors, respectively, for the ID components of **(a)** the observed and **(b)** the WRF-Chem model simulation datasets at the urban station.

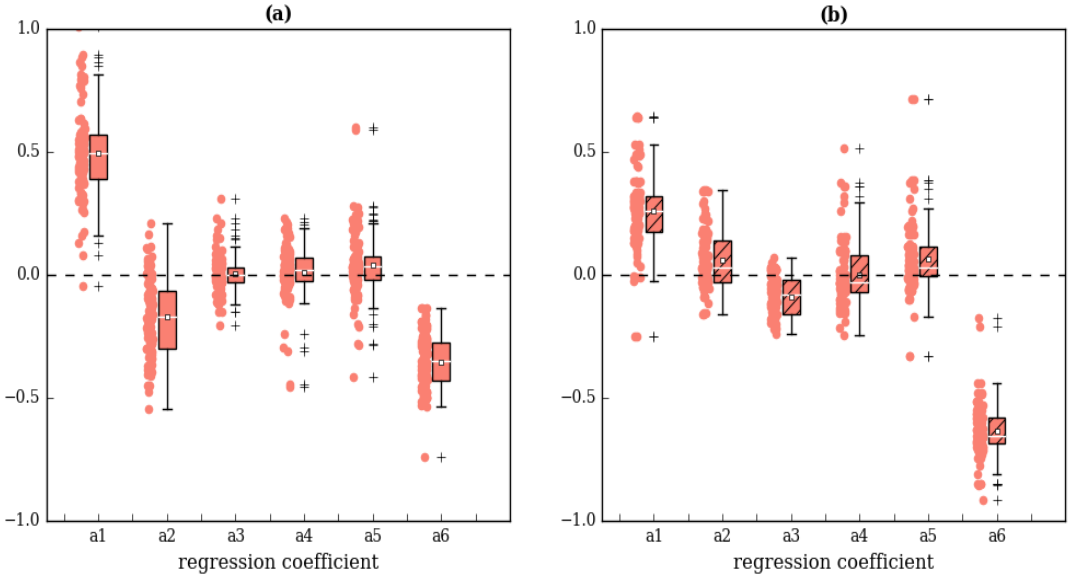

Figure D2. As Figure D1, but for the DU component.

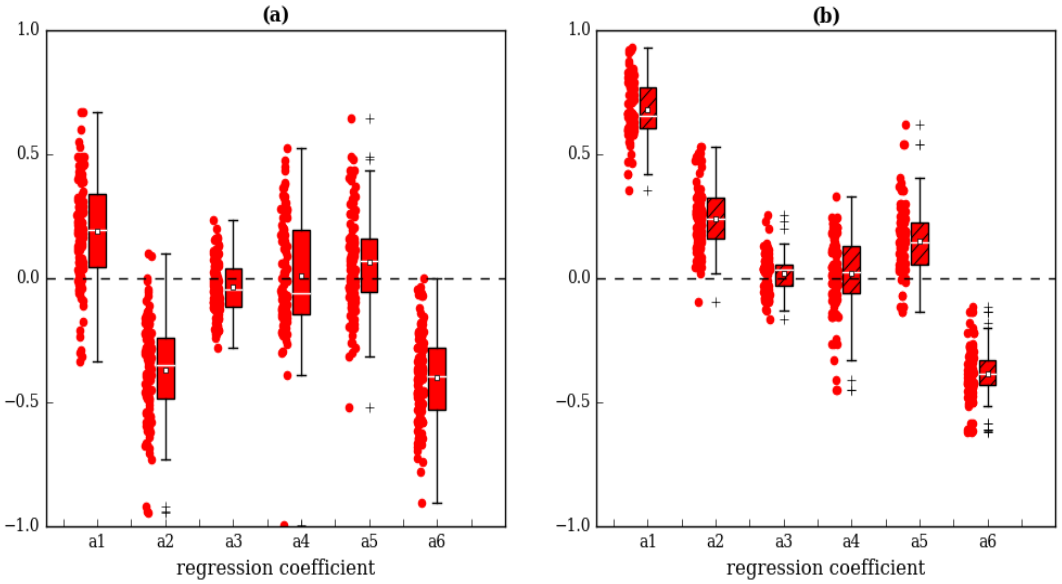

Figure D3. As Figure D1, but for the SY component.

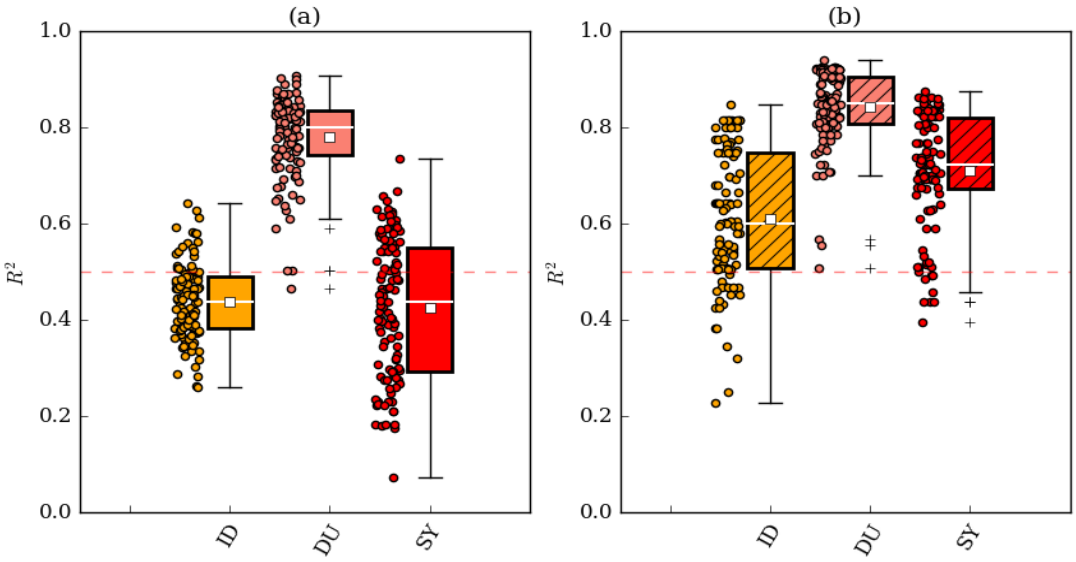

Figure D4. As Figure 4.11, but for using ($m_{i\text{-}opt}$, $K_{i\text{-}opt}$) instead of ($m_{i\text{-}ref}$, $k_{i\text{-}ref}$) in the KZ filter.

Table D1. As in Table A1, but for all urban stations. Standard deviation for the ID spectral component of the variables at the urban stations (averaged over summers 2010, 2011, and 2012).

variables	O_3^{surf}	AT	RH	SP	U	V	NO_x
the observations	2.96	0.46	2.7	0.23	0.47	0.48	4.56
the WRF-Chem model simulation	2.29	0.69	2.06	0.11	0.38	0.4	2.56

Table D2. As in Table D1, but for the DU component

variables	O_3^{surf}	AT	RH	SP	U	V	NO_x
the observations	10.35	2.94	11.62	0.74	1.01	0.99	7.31
the WRF-Chem model simulation	11.08	5.08	11.5	0.75	1.69	1.77	9.72

Table D3. As in Table D1, but for the SY component.

variables	O_3^{surf}	AT	RH	SP	U	V	NO_x
the observations	5.02	1.75	6.98	2.56	1	0.97	4.6
the WRF-Chem model simulation	7.44	1.93	7.18	2.53	1.42	1.57	3.67

Table D4. As Table 4.14, but for using ($m_{i\text{-}opt}$, $K_{i\text{-}opt}$) instead of ($m_{i\text{-}ref}$, $k_{i\text{-}ref}$) in the KZ filter. The values within the parenthesis show the percentage of the errors with respect to the sum of the model variance error (= 142.27). The sum of the values in this table may not be matched, mathematically. That is because of using a dot product instead of a cross product among spectral components. A similar table when ($m_{i\text{-}ref}$, $k_{i\text{-}ref}$) use in the KZ filter are given in Table 4.14.

stations type	errors	ID	DU	SY	SUM
urban	**variance error**	12.61 (8.86%)	83.74 (58.86%)	45.92 (32.28%)	142.27 (100%)
	explained error	3.71 (2.61%)	22.81 (16.03%)	29.96 (21.05)	56.48 (39.7%)
	unexplained error	8.89 (6.25%)	60.93 (42.83%)	21.06 (11.22%)	85.79 (60.30%)

Appendix E

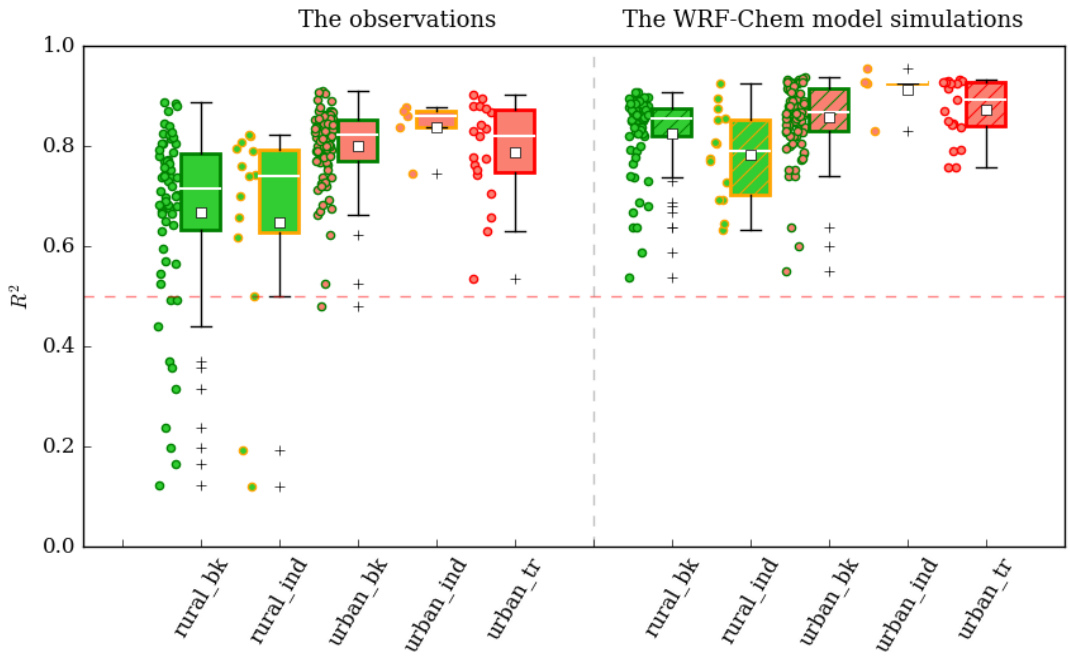

Figure E1. As Figure 5.4, but for the DU components.

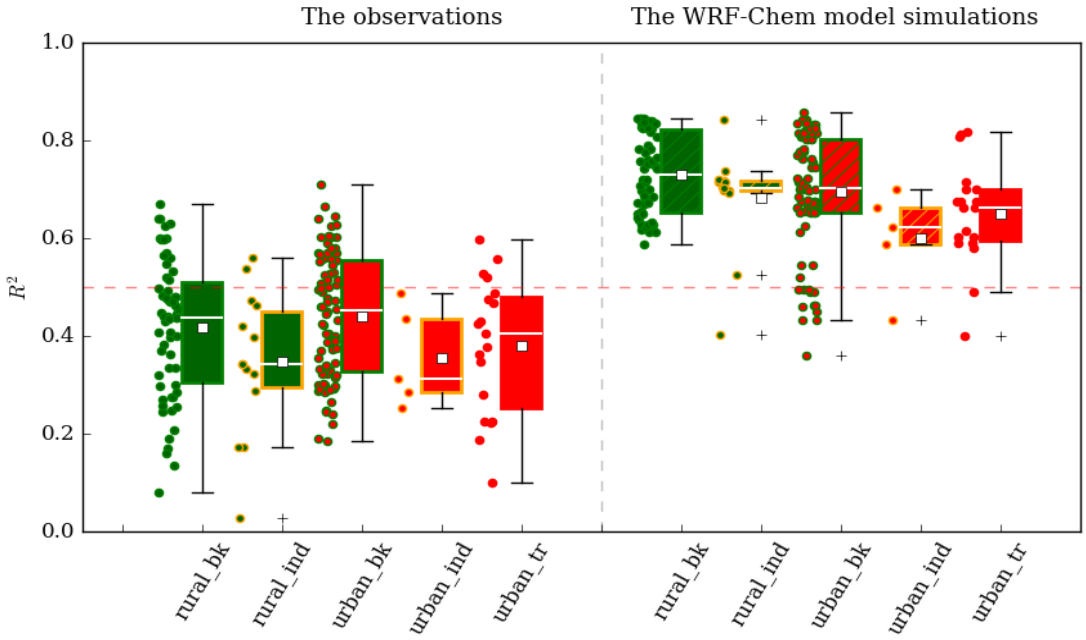

Figure E2. As Figure 5.4, but for the SY components.

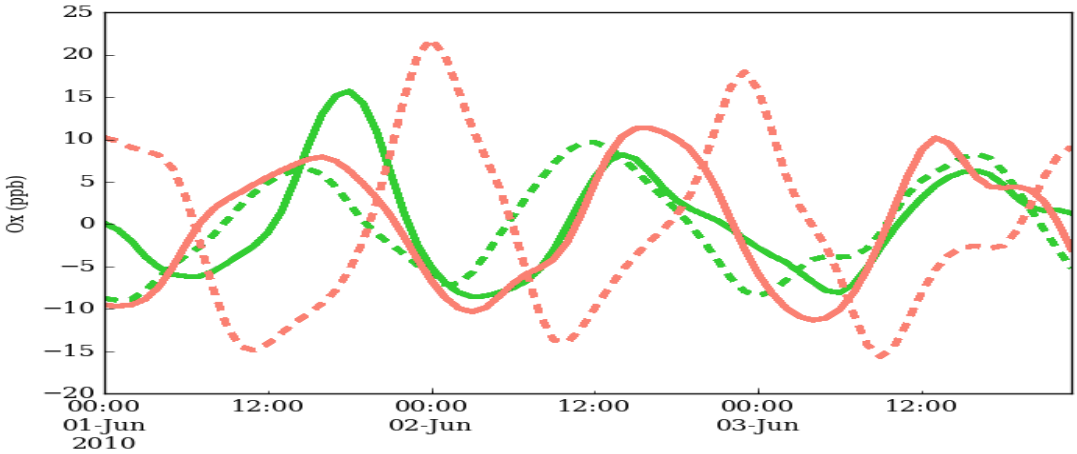

Figure E3. Hourly time series of the DU spectral component of O_x (O_3^{surf} + NO_2) at the ALC (salmon colour) and ACE (lime-green colour) stations. The solid and dashed lines show the measurement and the WRF-Chem model simulation datasets, respectively.

Appendix F

The following *settings list* had been used for the WRF-Chem model simulation in this thesis.

```
&time_control
 run_days                 = 100,
 run_hours                = 0,
 run_minutes              = 0,
 run_seconds              = 0,
 start_year               = 2010,   2014,     2014,
 start_month              = 05,     05,       05,
 start_day                = 01,     01,       01,
 start_hour               = 00,     00,       00,
 start_minute             = 00,     00,       00,
 start_second             = 00,     00,       00,
 end_year                 = 2010,   2014,     2014,
 end_month                = 10,     06,       06,
 end_day                  = 01,     03,       03,
 end_hour                 = 00,     00,       00,
 end_minute               = 00,     00,       00,
 end_second               = 00,     00,       00,
 interval_seconds         = 21600,
 input_from_file          = .true.,   .true.,   .true.,
 history_interval         = 60,     60,     60,
 frames_per_outfile       = 1,      1,      1,
 restart                  = .false.,
 restart_interval         = 180,
 io_form_history          = 2,
 io_form_input            = 2,
 io_form_boundary         = 2,
 io_form_auxinput2        = 2,
 io_form_auxinput4        = 2,
 debug_level              = 0,
 auxinput5_inname         = 'wrfchemi_d<domain>_<date>',
 auxinput6_inname         = 'wrfbiochemi_d<domain>',
```

```
 auxinput7_inname             = 'wrffirechemi_d<domain>_<date>',
 frames_per_auxinput5         = 1,  1,  1,
 frames_per_auxinput7         = 1,  1,  1,
 auxinput5_interval_m         = 60, 60, 60,
auxinput7_interval_h          = 1,
 io_form_auxinput5            = 2,
 io_form_auxinput6            = 2,
 io_form_auxinput7            = 2,
/
&domains
 time_step                    = 120,
 time_step_fract_num          = 0,
 time_step_fract_den          = 1,
 max_dom                      = 1,
 e_we                         = 130, 146, 154,
 e_sn                         = 130, 156, 154,
 e_vert                       = 35,      35,      35,
 p_top_requested              = 5000,
 num_metgrid_levels           = 38,
 eta_levels                   = 1.000, 0.993, 0.983, 0.97,
                                0.954, 0.934, 0.909, 0.88, 0.845,
                                0.807, 0.765, 0.719, 0.672, 0.622,
                                0.571, 0.52, 0.468, 0.42, 0.376,
                                0.335, 0.298, 0.263, 0.231, 0.202,
                                0.175, 0.15, 0.127, 0.106, 0.088,
                                0.07, 0.055, 0.04, 0.026, 0.013,
                                0.000,
num_metgrid_soil_levels       = 4,
 dx                           = 30000,    3000,    1000,
 dy                           = 30000,    3000,    1000,
 grid_id                      = 1,       2,       3,
 parent_id                    = 1,       1,       2,
 i_parent_start               = 1,      61,      43,
 j_parent_start               = 1,      60,      55,
 parent_grid_ratio            = 1,       5,       3,
 parent_time_step_ratio       = 1,       5,       3,
 feedback                     = 0,
```

```
 smooth_option                = 0,
/
&physics
 mp_physics                   = 10,    10,    10,
 ra_lw_physics                = 4,     4,     4,
 ra_sw_physics                = 4,     4,     4,
 radt                         = 15,    15,    15,
 sf_sfclay_physics            = 2,     2,     2,
 sf_surface_physics           = 2,     2,     2,
 bl_pbl_physics               = 2,     2,     2,
 bldt                         = 0,     0,     0,
 cu_physics                   = 3,     3,     3,
 ishallow                     = 1,
 cudt                         = 5,     5,     5,
 isfflx                       = 1,
 ifsnow                       = 0,
 icloud                       = 1,
 surface_input_source         = 1,
 num_soil_layers              = 4,
 naer                         = 1e9
 num_land_cat                 = 24,
 mp_zero_out                  = 0,
 mp_zero_out_thresh           = 1.e-8
 maxiens                      = 1,
 maxens                       = 3,
 maxens2                      = 3,
 maxens3                      = 16,
 ensdim                       = 144,
 cu_rad_feedback              = .true.,
/
&fdda
 grid_fdda                    = 1,
 gfdda_inname                 = "wrffdda_d<domain>",
 gfdda_interval_m             = 360,
 gfdda_end_h                  = 9000,
 io_form_gfdda                = 2,
 if_no_pbl_nudging_uv         = 1,
```

```
 if_no_pbl_nudging_t      = 1,
 if_no_pbl_nudging_q      = 1,
 if_zfac_uv               = 1,
 k_zfac_uv                = 10,
 if_zfac_t                = 1,
 k_zfac_t                 = 10,
 if_zfac_q                = 1,
 k_zfac_q                 = 10,
 guv                      = 0.0003,
 guv                      = 0.0003,
 guv                      = 0.0003,
 if_ramping               = 1,
 dtramp_min               = 360,
/
&dynamics
 rk_ord                   = 3,
 w_damping                = 1,
 diff_opt                 = 1,
 km_opt                   = 4,
 diff_6th_opt             = 0,    1,  1,
diff_6th_factor           = 0.12,  0.12, 0.12,
 damp_opt                 = 0,
 dampcoef                 = 0.2,   0.2,   0.2,
 zdamp                    = 5000.,  5000.,  5000.,
 khdif                    =  0,    0,     0,
 kvdif                    =  0,    0,     0,
 smdiv                    = 0.1,   0.1,   0.1,
 emdiv                    = 0.01,  0.01,  0.01,
 epssm                    = 0.1,   0.1,   0.1,
 mix_full_fields          = .true., .true., .true.,
 non_hydrostatic          = .true., .true., .true.,
 use_baseparam_fr_nml     = .false.
 time_step_sound          = 0,     0,     0,
 h_mom_adv_order          = 5,     5,     5,
 v_mom_adv_order          = 3,     3,     3,
 h_sca_adv_order          = 5,     5,     5,
 v_sca_adv_order          = 3,     3,     3,
```

```
 moist_adv_opt                       = 2,     2,     2,
 chem_adv_opt                        = 2,     2,     2,
 scalar_adv_opt                      = 2,     2,     2,
 tke_adv_opt                         = 2,     2,     2,
/
&bdy_control
 spec_bdy_width                      = 5,
 spec_zone                           = 1,
 relax_zone                          = 4,
 specified                           = .true., .false., .false.,
 nested                              = .false., .true., .true.,
/
&grib2
/
&namelist_quilt
nio_tasks_per_group                  = 0,
 nio_groups                          = 1,
/
&chem
 kemit                               = 1,
 ne_area                             = 1400,
 chem_opt                            = 111, 113, 113,
 bioemdt                             = 15, 15, 15,
 photdt                              = 15, 15, 15,
 chemdt                              = 5, 15, 15,
 emiss_inpt_opt                      = 1, 1, 1,
 emiss_opt                           = 7, 7, 7,
 io_style_emissions                  = 2,
 chem_in_opt                         = 1, 1, 1,
 phot_opt                            = 3, 3, 3,
 gas_drydep_opt                      = 1, 1, 1,
 bio_emiss_opt                       = 3, 3, 3,
 seas_opt                            = 0,
 dust_opt                            = 0,
 gas_bc_opt                          = 1, 1, 1,
 gas_ic_opt                          = 1, 1, 1,
 gaschem_onoff                       = 1, 1, 1,
```

```
 vertmix_onoff             = 1, 1, 1,
 wetscav_onoff             = 1, 1, 1,
 cldchem_onoff             = 0, 0, 0,
 biomass_burn_opt          = 2,
 plumerisefire_frq         = 30,
 scale_fire_emiss          = .true.,
 have_bcs_chem             = .true., .false., .false.,
/
```

Characteristic of the **rural** stations

Name	ID	Latitude	Longitude	Types
station name	AT60107	47.045	15.15	background
PIEVE D"ALPAGO 502507	IT1790A	46.1625	12.360833	background
ES0010R-CABO DE CREUS	ES0010R	42.3189	3.31582	background
ES1531A-LOS TOJOS	ES1531A	43.15361	-4.256944	background
ES1662A-CTCC-FUNES	ES1662A	42.30833	-1.808333	background
station name	ES0009R	41.27417	-3.1425	background
Kramsach Angerberg	AT72538	47.45917	11.910563	background
FONTECHIARI 1206005	IT0992A	41.67	13.674445	background
ES1991A-MURIEL DE LA FUENTE	ES1991A	41.72361	-2.856944	background
AMS Rakovsky-Dimitrovgrad	BG0041A	42.05638	25.59345	background
AIRVAULT 2	FR09301	46.82437	-0.135461	industrial
BALLIRANA 803922	IT1927A	44.64139	11.982222	background
Sulzberg im Bregenzerwald	AT80503	47.52833	9.92722	background
ASIAGO - CIMA EKAR 502415	IT1791A	45.84833	11.569166	background
ES0365A-VILLALBA DE GUARDO	ES0365A	42.70361	-4.827222	industrial
Haunsberg	AT53055	47.97278	13.015847	background
ST JUNIEN - Fontaine	FR35002	45.89	0.9	background
station name	AT10002	47.30417	16.193064	background
ES1297A-LA ROBLA	ES1297A	42.80278	-5.623611	industrial
ES1489A-VALDEREJO	ES1489A	42.87528	-3.231389	background
Lenzing	AT4S418	47.97195	13.600842	industrial
Liezen	AT60182	47.56722	14.24362	background
LEONESSA 1205701	IT0989A	42.5725	12.961945	background
ES1808A-SAN MARTIN DE VALDEIGLESIAS	ES1808A	40.36889	-4.396944	background
ETROUBLES 200709	IT0979A	45.81583	7.235278	background
station name	ES0016R	42.63472	-7.70472	background
ES1811A-VILLAREJO DE SALVANES	ES1811A	40.16722	-3.276667	background
ES0012R-ZARRA	ES0012R	39.08278	-1.10111	background
ES1881A-PUIGMORENO	ES1881A	41.10583	-0.253889	industrial
station name	AT0ZOE2	47.83861	14.441395	background

LA TARDIERE	FR23124	46.65668	-0.743681	background
ES0001R-SAN PABLO DE LOS MONTES	ES0001R	39.54667	-4.35056	background
ES1432A-CANGAS DE NARCEA	ES1432A	43.17778	-6.547222	industrial
GR-MAREMMA 905304	IT1942A	42.67056	11.094167	background
ES0761A-CUADROS	ES0761A	42.71556	-5.638889	industrial
station name	ES0013R	41.23889	-5.8975	background
ES1746A-GALILEA	ES1746A	42.34167	-2.231667	industrial
PEYRUSSE	FR12020	43.62871	0.17911	background
ES0360A-LILLO	ES0360A	42.78139	-6.603889	industrial
ES1660A-CTCC-ARGUEDAS	ES1660A	42.20833	-1.585556	background
ES1805A-GUADALIX DE LA SIERRA	ES1805A	40.78083	-3.7025	background
ES1579A-LOS CORRALES DE BUELNA	ES1579A	43.26444	-4.062778	background
BORGO VAL 402201	IT0703A	46.05111	11.454444	background
station name	FR36005	44.52074	5.089872	background
ES1535A-ALBACETE	ES1535A	38.98167	-1.956667	background
Charavines	FR15001	45.4276	5.51795	background
Payerbach	AT31502	47.67	15.854726	background
St. Georgen im Lavanttal - Herzogberg	AT2WOZONE5	46.70806	14.891953	background
Deutschlandsberg Rathausgasse	AT60195	46.81389	15.212789	background
Chamusca	PT03096	39.21139	-8.281958	background
Bockberg	AT60151	46.87306	15.495011	background
ES1671A-VILLAR DEL ARZOBISPO	ES1671A	39.70806	-0.831944	background
ES1802A-EL ATAZAR	ES1802A	40.91028	-3.466667	background
ES1793A-EL ARENOSILLO	ES1793A	37.10409	-6.734081	background
station name	ES1753A	42.34	-2.059722	industrial
ES1835A-LA JOYA	ES1835A	36.95	-1.961481	industrial
station name	AT60185	46.75278	15.956672	background
ES0008R-NIEMBRO	ES0008R	43.43917	-4.85	background
Saxon	CH0024A	46.13899	7.148333	background
ES0011R-BARCARROTA	ES0011R	38.47278	-6.92361	background
station name	ES1806A	40.28694	-3.222222	background

ES0014R-ELS TORMS	ES0014R	41.39389	0.73472	background
ES0622A-AGUA AMARGA	ES0622A	36.94046	-1.939911	industrial
station name	FR09302	46.14317	-0.396042	background
station name	AT60198	47.04694	16.078901	background
station name	AT72912	46.81917	12.765831	background
OSTELLATO 803813	IT1919A	44.74167	11.942223	background
Bludenz Herrengasse	AT82708	47.15611	9.82723	background
AMS Rail. Station-Vratza	BG0043A	43.20798	23.557414	background
station name	AT60137	47.05917	15.016666	background
ES0813A-LA CEROLLERA	ES0813A	40.84194	-0.0575	industrial
Saliceto - Moizo 10401	IT1519A	44.41361	8.1675	background
Hartberg Zentrum	AT60179	47.28278	15.971944	background
ES1287A-GUARDO	ES1287A	42.79528	-4.840833	industrial
ES1654A-SIERRA NORTE	ES1654A	37.99435	-5.666869	background
station name	AT60194	47.60472	15.673061	background

Characteristic of the **urban** stations

Name	ID	Latitude	Longitude	Types
GAILLARD	FR33212	46.19369	6.214859	background
AGEN	FR31032	44.19222	0.623889	background
ES1891A-GRANOLLERS	ES1891A	41.60083	2.288333	traffic
MONZA via MACHIAVELLI 301569	IT1743A	45.58111	9.274445	background
MAZADES	FR12021	43.62361	1.438611	background
ES1268A-PLAZA DE TOROS	ES1268A	43.35778	-5.86472	traffic
VILLA ADA 1205820	IT0953A	41.93278	12.50695	background
VOTANIKOS	GR0029A	37.98363	23.70679	industrial
ES1422A-PLAZA DEL CARMEN	ES1422A	40.41917	-3.70333	background
ES1638A-BERMEJALES	ES1638A	37.34719	-5.97963	background
ES0126A-FAROLILLO	ES0126A	40.39472	-3.73194	background
GRAND PARC	FR31001	44.86139	-0.57944	background
Beato	PT03070	38.73333	-9.11362	background
GERLAND	FR20017	45.73529	4.829794	background
ES1564A-BENDAS	ES1564A	40.54056	-3.64472	industrial
Station DAIX	FR26010	47.34555	5.002222	background
LYON Centre	FR20062	45.75773	4.854214	background
GRASSE CLAVECIN	FR24015	43.65694	6.919444	background
station name	FR35003	45.8311	1.256875	background
Station TARNIER	FR26002	47.30416	5.02	background
PD - Mandria 502808	IT1453A	45.37194	11.84194	background
Ermesinde-Valongo	PT01023	41.20611	-8.55168	background
Winterthur-Obertor	CH0014A	47.50023	8.731838	background
Sobreiras-Lordelo do Ouro	PT01050	41.1475	-8.65889	background
Epalinges	CH0037A	46.54229	6.662928	background
station name	ES1563A	40.47917	-3.37778	traffic
Station PEJOCES	FR26005	47.30806	5.066111	background
ES1946A-JUAN CARLOS I	ES1946A	40.465	-3.60889	background
ES1645A-BARAJAS - PUEBLO	ES1645A	40.47472	-3.575	background
LIOSIA	GR0027A	38.07681	23.69784	background
AMS Kamenitza	BG0051A	42.14289	24.76524	background
APPENTIS	FR23150	47.44739	-0.55007	background

Place du Marche	FR09009	46.58305	0.344889	background
Laranjeiro	PT03083	38.66361	-9.15778	background
ES1829A-PARC BIT-PALMA	ES1829A	39.63472	2.630556	background
station name	ES1635A	38.35944	-0.47194	traffic
ES1521A-BARRIO DEL PILAR	ES1521A	40.47833	-3.71139	traffic
St Martin d"Heres	FR15038	45.1828	5.7531	background
station name	CH0048A	46.21613	6.1513	traffic
ES1804A-GETAFE	ES1804A	40.32472	-3.71417	traffic
ES1747A-ROTXAPEA	ES1747A	42.82694	-1.64944	background
Fontaine Les Balmes	FR15017	45.19044	5.686964	background
Bourg-en-Bresse	FR33305	46.21167	5.226389	background
Delille	FR07008	45.78089	3.094583	background
Piazza Repubblica 2009234	IT1634A	39.21722	9.126111	traffic
Restelo	PT03087	38.705	-9.21029	background
ES1536A-AZUQUECA DE HENARES	ES1536A	40.57361	-3.26333	background
ALBI DELMAS	FR12026	43.92806	2.146111	background
ANNEMASSE	FR33211	46.19583	6.240556	background
ES1018A-E1-TERRASSA	ES1018A	41.55556	2.007778	traffic
BUFALOTTA 1205884	IT1835A	41.94778	12.53361	background
TV - Via Lancieri di Novara 502608	IT1590A	45.6725	12.23861	background
ES1869A-COSLADA	ES1869A	40.43	-3.54583	traffic
station name	ES1038A	43.34611	-4.04972	background
ES1807A-RIVAS-VACIAMADRID	ES1807A	40.36	-3.54222	background
station name	CH0013A	47.38677	8.539811	traffic
ALCORCÓN	ES1890A	40.34194	-3.83389	background
station name	ES1567A	40.33972	-3.73583	traffic
ES0118A-ESCUELAS AGUIRRE	ES0118A	40.42167	-3.68222	traffic
ES1658A-CAMPUS DEL CARMEN	ES1658A	37.27148	-6.92462	industrial
TALENCE	FR31002	44.80056	-0.58806	background
ES1593A-SAN FERNANDO	ES1593A	36.46059	-6.20307	background
BERTHELOT	FR12030	43.58722	1.443889	background
SAINT-CROUTS	FR31016	43.47972	-1.48806	background
LOVERCHY	FR33201	45.89528	6.118333	background
ST JUST	FR20004	45.75532	4.820331	background

ES1630A-ALJARAFE	ES1630A	37.34075	-6.04246	background
station name	ES0694A	41.39333	2.010556	industrial
CAGNES LADOUMEGUE	FR24020	43.65834	7.157778	background
OSIO SOTTO 301621	IT1463A	45.62055	9.612778	background
station name	CH0049A	46.19888	6.132222	background
station name	CH0005A	47.40295	8.613414	background
TENUTA DEL CAVALIERE 1205817	IT0957A	41.93139	12.65917	background
MISTERBIANCO 1908701	IT1899A	37.51639	15.02694	unknown
CIM BOUTEILLERIE	FR23188	47.2225	-1.5375	background
Meyrin-Vaudagne	CH0050A	46.23106	6.074183	background
Bern-Brunngasshalde	CH0020A	46.94915	7.449575	traffic
ES1208A-AF-REUS	ES1208A	41.15222	1.121389	traffic
Montferrand	FR07004	45.79728	3.113833	background
Les Couronneries	FR09015	46.58692	0.358583	background
ES1752A-TORREJON DE ARDOZ II	ES1752A	40.455	-3.48417	background
ARENULA 1205887	IT1906A	41.89389	12.47528	traffic
station name	ES1568A	40.32417	-3.87639	background
ES1941A-ENSANCHE DE VALLECAS	ES1941A	40.37278	-3.61194	background
Les Carmes	FR08714	42.6958	2.8997	background
ES1924A-PALACIO CONGRESOS	ES1924A	37.16568	-3.6001	background
PERISTERI	GR0028A	38.02081	23.68835	background
ES1740A-PLAZA DE LA CRUZ	ES1740A	42.81222	-1.64	traffic
ES0124A-ARTURO SORIA	ES0124A	40.44	-3.63917	background
BILLERE	FR31013	43.31083	-0.39111	background
NOVEL	FR33202	45.91694	6.135556	background
BASSENS	FR31007	44.90028	-0.51472	background
ES0890A-TORNEO	ES0890A	37.39452	-6.0031	traffic
ES1269A-PALACIO DE DEPORTES	ES1269A	43.36722	-5.83139	traffic
station name	ES1885A	39.48028	-0.33639	background
station name	CH0046A	46.1956	6.211264	background
Station BALZAC	FR26007	47.34139	5.046944	background
ES1750A-CARRANQUE	ES1750A	36.71964	-4.4475	background
La Seyne sur mer 2	FR03060	43.10159	5.875383	background
Jardin Lecoq	FR07009	45.77217	3.0875	background

place de VERDUN	FR09003	46.16256	-1.1536	background
Graz Nord	AT60138	47.09555	15.41418	background
HAMEAU	FR31014	43.30778	-0.32278	background
CANTEPAU ALBI	FR12017	43.92806	2.1625	background
AVIGNON MAIRIE	FR03080	43.94973	4.805081	background
ES1565A-FUENLABRADA	ES1565A	40.28111	-3.80167	industrial
CINECITTA 1205804	IT0956A	41.85778	12.56861	background
ECOLE M.JACQUIER	FR12004	43.57556	1.418611	background
station name	CH0010A	47.37759	8.530419	background

Acknowledgements

My first appreciation to both of the great supervisors of mine PD. Dr. Martin Schultz and Prof. Dr. Andreas Hense for giving me an opportunity to freely ask questions, to have so many opening discussions, and leading me in the right direction.

My sincere thanks also go to the directors of IEK-8, Institute of Energy and Climate Research Troposphere, Prof. Dr. Andreas Wahner and Prof. Astrid Kiendler-Scharr for their clear inquiries leading to a light toward the right direction. Moreover, many thanks to them for providing facilities at the institute, Friday seminar, local server, data, friendly atmosphere, and an office, where I spent my working time there.

Besides, I would like to thank Prof. Dr. Tim Butler and Dr. Aura Lupascu who provided me the model simulation data.

I greatly acknowledge the funding source, the Helmholtz Association, which made my Ph.D. work and conference participation possible.

Thanks to the HITEC, Helmholtz Graduate School of Forschungszentrum Jülich, where provided me many great training programs.

Thanks to the community of Python programming language, Pandas library, and stack overflow for providing fantastic tools and documentation that made my scientific study and communication feasible.

I would also like to thank my colleagues at the FZJ, Forschungszentrum Jülich, for having many enjoyable times together.

My last and foremost thanks to my lovely family members Iman, Ali, and Zahra who always give me motivation, support and encourage me to pass this stage.