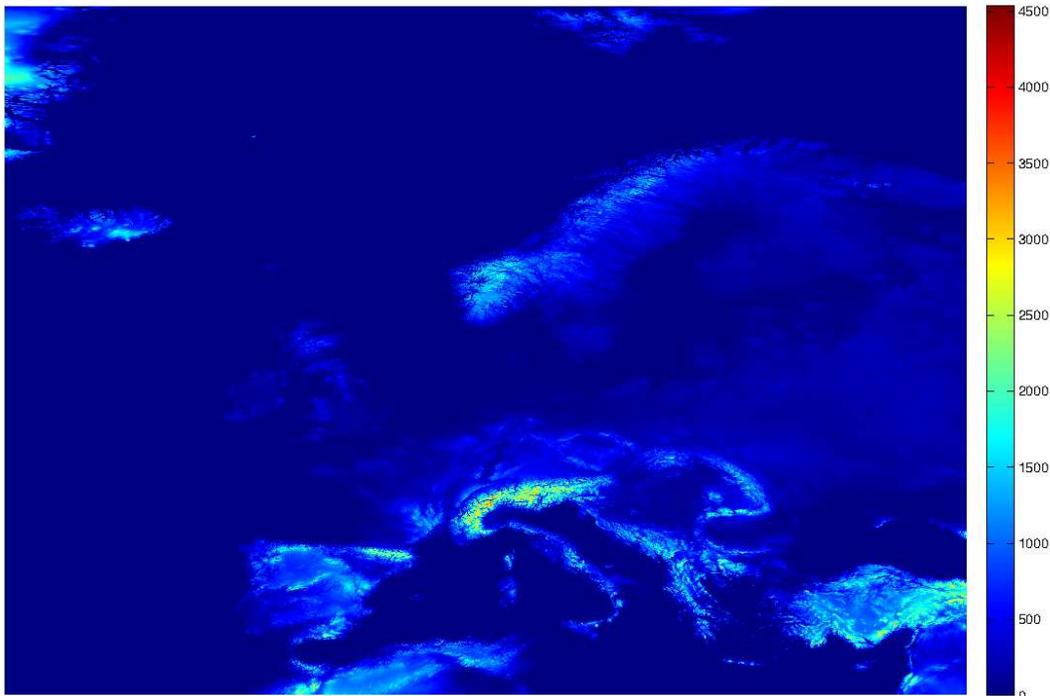


CHALMERS



Analysis and Modelling of Topographic Effects on Ozone and Other Pollutants from a Chemical Transport Model

Master Thesis within the Radio and Space Science Master's Program

Emma Andersson

Jakob Sld

Supervisor: *David Simpson*

Department of Earth and Space Sciences
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden, 4th of June 2012

**Analysis and Modelling of Topographic Effects on Ozone and Other Pollutants from
a Chemical Transport Model**

EMMA ANDERSSON, JAKOB SÜLD

©EMMA ANDERSSON, JAKOB SÜLD

Master Thesis within the Radio and Space Science Master's Program

Department of Earth and Space Sciences
Chalmers University of Technology
412 96 Gothenburg

Cover: The topographic map with a resolution of 30 arcsec from GTOPO30.

Abstract

The modelling of terrain in chemical transport models is a difficult task. Most models have a too coarse resolution to be able to resolve regions with more complex terrain. Problems arise when validation of the models needs to be done with measurements from different stations at different altitudes. The largest deviation is found at stations with a high altitude above sea level in regions with strong altitude variations. This problem has been identified within the the chemical transport model, EMEP MSC-W which is a global air quality model developed for European conditions.

The aim of this thesis is to analyse and post-process 3D data from the EMEP MSC-W model in order to create a sub-grid model that captures the topographic effects in a better way. This will be done for three different pollutants, ozone, nitrogen dioxide, sulphur dioxide and sulphate. The sub-grid model scales the output data from EMEP MSC-W from 50x50 km to 1x1 km and extracts data using four different methodologies. Two of the methods are based on altitude above sea level or absolute altitude and one is on a relative altitude. This relative altitude is defined by subtracting the lowest altitude within a search radius of 3, 5 or 10 km, from a given location. These methods are compared with the output data closest to the model topography and validated with measurements. The model is developed and evaluated with the diurnal variations of ozone and tested for nitrogen dioxide, sulphur dioxide and sulphate. Other possible developments of this model will also be discussed regarding downscaling of dry deposition with respect to land use, temperature and pressure.

The results became different for all the pollutants, but since there are far more measurement stations for ozone the statistical significance of this result becomes important. The best method of extraction was when using the relative height with a search radius of 3 km. A downscaling of the land use, temperature and pressure also points out the possibilities of getting improvements if implementing a higher resolved dry deposition.

Acknowledgements

With this thesis we would like to acknowledge the Norwegian Meteorological Institute MET.NO for giving us an office and letting us use their facilities. A special thank we would like to dedicate to Hilde Fagerli for setting a side time to supervise us at MET.NO. Last but not least we would like to thank our main supervisor David Simpson for giving us this opportunity to gain experience from a new environment and new people and for supporting and supervising us during our master thesis.

Contents

1	Introduction	3
1.1	Aim	5
1.2	Limitations and assumptions	5
2	Theory	6
2.1	Atmospheric properties	6
2.1.1	Vertical distribution	6
2.1.2	Ozone, O ₃	8
2.1.3	Nitrogen dioxide , NO ₂	9
2.1.4	Sulphur dioxide, SO ₂	10
2.1.5	Sulphate, SO ₄	11
2.2	EMEP model	12
2.2.1	Model coverage	12
2.2.2	Model data	15
2.2.3	Model vertical profiles of the pollutants	15
2.2.4	Deposition	16
2.2.5	Ozone in the EMEP MSC-W model	18
2.3	Literature Reviews	20
3	Method	21
3.1	The extraction methods	21
3.1.1	Relative altitude	21
3.1.2	Station selection	22
3.1.3	Interpolation	22
3.2	Analysis of the pollutants	23
3.2.1	O ₃	23
3.2.2	NO ₂	25
3.2.3	SO ₂ and SO ₄	25
3.2.4	Best method selection	25
3.3	Sub-grid extraction model scripts	25
3.3.1	sondes_read.py	25
3.3.2	NILUstat.py	26
3.3.3	heights.py	26
3.3.4	sites_read.py	26
3.3.5	NILUdata.py	27
3.3.6	main.py	27
3.4	Dry deposition downscaling	27
3.4.1	Land use	27
3.4.2	Temperature and pressure	28

4	Results	29
4.1	O ₃	29
4.1.1	Ozone maps	31
4.2	NO ₂	32
4.3	SO ₂	33
4.4	SO ₄	37
4.5	Deposition downscaling	37
4.5.1	Land use	37
4.5.2	Temperature and Pressure	39
5	Discussion	42
5.1	The results	42
5.2	The sub-grid model	43
5.3	Downscaling dry deposition	44
6	Conclusion	45
A	Tables and figures	48
A.1	Stations and land classes	48
A.2	Results O ₃	53
A.3	Results NO ₂	55
A.4	Results SO ₂	57
A.5	Results SO ₄	57
B	Source code	61
B.1	main.py	61
B.2	NILUstat.py	69
B.3	NILUdata.py	71
B.4	heights.py	73
B.5	sondes_read.py	76
B.6	sites_read.py	78

1

Introduction

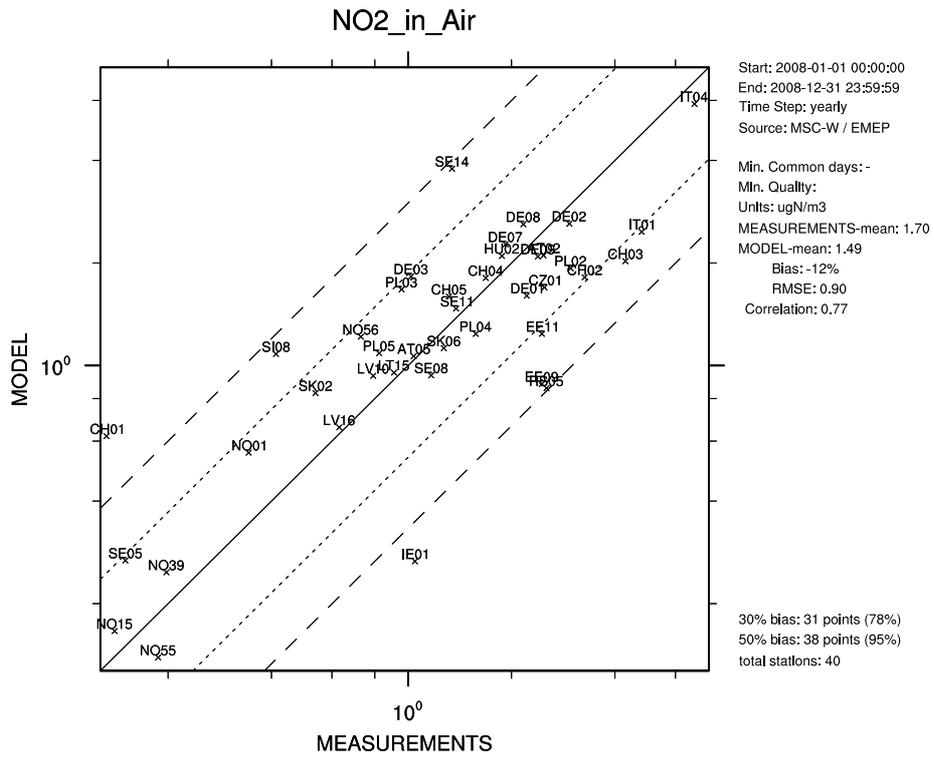
Modelling of topographic effects in chemical transport models is not an easy task. Many models have a too coarse resolution to be able to resolve regions with highly varying terrains and since these regions have a much more complex behaviour when it comes to meteorology and chemistry, the models does not mimic or predict in a desired way. Studies has been done on how the air quality changes due to the terrain by studying measurement from several sites and some of these studies are mentioned later on in section 2.3.

The atmospheric species that are most affected by the topography are the ones with a lifetime of about one day or longer and those who are more abundant. For this thesis four pollutants has been chosen as study objects: ozone (O_3), nitrogen dioxide (NO_2), sulphur dioxide (SO_2) and sulphate (SO_4). All these species has a moderately long lifetimes. O_3 is a secondary pollutant and has a lifetime of about one day, NO_2 and SO_2 are primary pollutants with lifetimes less than a day respectively a few days and SO_4 is an aerosol with a lifetime of a few days (Wallace and Hobbs 2006). These species are also interesting because NO_2 is a precursor gas to O_3 and SO_2 is a precursor gas to SO_4 .

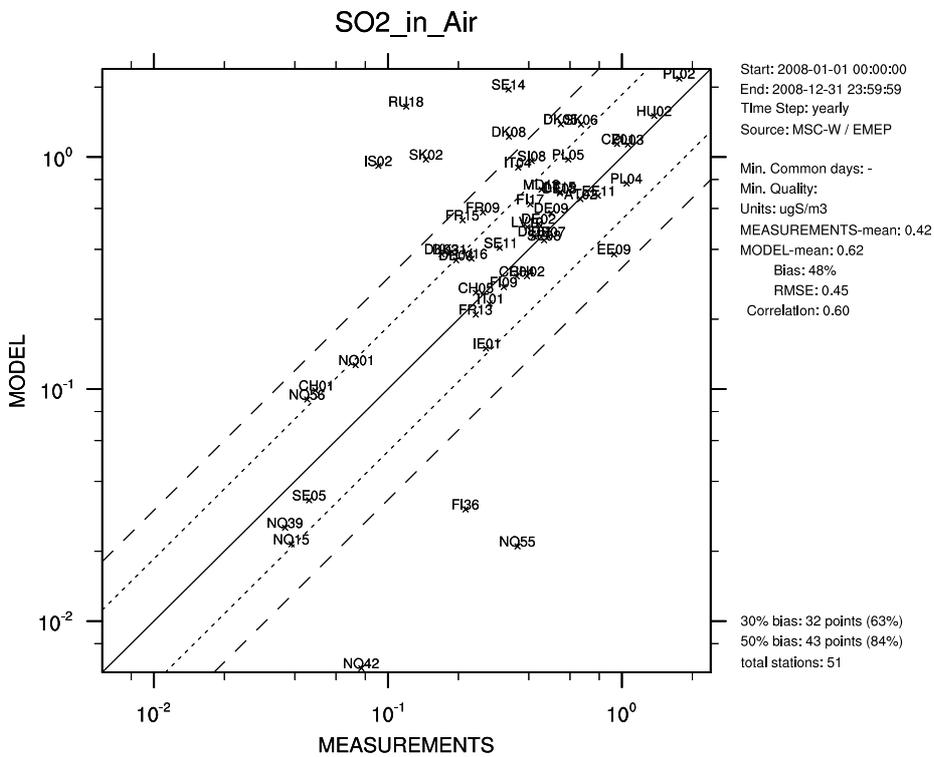
The chemical transport model that is used in this thesis is developed by the Meteorological Synthesizing Centre-West (MSC-W) of the European Monitoring and Evaluation Programme (EMEP). The model is designed for simulating concentrations and depositions of acidifying and eutrophying pollutants (S, N), ground level O_3 and particulate matter with size distribution under 2.5 and 10 μm . These simulations works as a foundation for European air pollution assessments and are supported by the Convention on long-range Transboundary Air Pollution (LRTAP) (Simpson et al. 2012).

The EMEP MSC-W model cover all Europe (see figure 2.3) and have a grid resolution of 50x50 km^2 and extends in the vertical direction from ground level to the tropopause (100 hPa) in 20 distinct layers. In addition to these vertical layers there is a ground layer 3 m above the surface with values interpolated from the lowest vertical layer. As validation of the model the 3 m ground level has been used against measurement from different stations around Europe. Figure 1.1 shows typical scatterplots derived for NO_2 and SO_2 from data for year 2008. Here some of the overestimated outliers have something in common, they are stations situated on an high altitude above sea level or they are coast stations. Some of the corresponding station to each code in the scatter plot will be found in appendix table A.2 and A.3.

The identified problem here is the higher elevated stations and the coast stations. There is a common property of these two stations as well, both have a lower deposition rate of atmospheric species.



(a) NO₂



(b) SO₂

Figure 1.1: Scatterplots for NO₂ and SO₂ from data of 2008 where highly elevated stations clearly can be distinguished as overestimated outliers, CH01 (3578 m.a.s.l.) for NO₂ and SK02 (2008 m.a.s.l.) for SO₂.

1.1 Aim

In this thesis the main focus will lie on the highly elevated stations. The aim is to investigate how the topography affects the air quality by studying ozone, nitrogen dioxide, sulphur dioxide and sulphate in order to create a model that includes such effects. The goal will be to create a sub-grid method for post-processing data from the EMEP MSC-W model to provide a better correlation between model and measurements for stations positioned at high altitudes. This model will be created in the programming language Python.

The specific questions that this thesis aims to get answers to are:

- Is it possible to improve the results from the EMEP model and increase the resolution for ozone concentrations by including more information about topography and how could it be done?
- Is it possible to implement a similar method to get improved results for other pollutants and temperature?
- Could it then be possible to create ozone maps over the EMEP grid?

1.2 Limitations and assumptions

The sub-grid post-processing methods will be developed for O_3 with measurement data from 2008. The methods will then be evaluated and tested for the other three species. The reason is mainly for having a larger set of highly elevated measurement stations with hourly data for O_3 than for the other species. With the hourly data you can look at the diurnal variation and get better statistical significance. O_3 is also well documented in the EMEP MSC-W model which makes it a better target for analysis. This means that the investigation of further development of the developed method will be discussed in terms of O_3 . The results and discussion will also have a larger focus on O_3 .

There will be three different methodologies under investigation in this thesis and further development will be discussed and analysed in terms of dry deposition and its surface resistance with respect to land class, temperature and pressure. The aerodynamic and quasi-boundary layer resistances does not lie in the scope of this thesis, but could with benefit be analysed in future studies.

Some of the used measurement stations did not have measurement throughout the whole year which makes these stations a larger source of error, but stations with a set of data larger than three months will still be used in the analysis. Some stations from certain countries also have a lower credibility since they do not have the same base of regularly calibrations. These stations will still be used in this analysis.

2

Theory

This chapter will first give a short introduction to some atmospheric properties. The vertical structure of the atmosphere will be treated with focus on the temperature and pressure. When the more physical properties of the atmosphere has been surveyed a section handling the chemical properties will be given. The four investigated pollutants ozone, nitrogen dioxide, sulphur dioxide and sulphate will be presented with their major chemistry and how they vary in time and space.

After the introduction of the atmospheric properties, a presentation of the chemical transport model the EMEP MSC-W model will be given and how it models some of these atmospheric properties in terms of the vertical profile of the four species and the dry deposition.

2.1 Atmospheric properties

The atmosphere of Earth is a thin layer of gas surrounding the globe. About 80% of its mass is situated within the lowest 15 km. The constituents of this mass given in fractional concentration by volume is 78.08 % of molecular nitrogen (N_2), 20.95 % of molecular oxygen (O_2), 380 ppm of carbon dioxide (CO_2), 0-0.1 ppm of ozone (O_3) (Finlayson-Pitts and Pitts 2000, Wallace and Hobbs 2006) and the an average abundance of sulphate (SO_4^{2-}) in Europe is, according to measurements from EMEP sites (Torseth et al. 2012), 0.1-1 $\mu g m^{-3}$ (or ~ 0.07 -0.7 ppm).

This section will give an overview of how the terrain can influence the atmospheric properties by looking at how local and regional air pollutant concentrations change due to transport phenomena, i.e. how the wind flow interacts with temperature changes and complex terrain.

2.1.1 Vertical distribution

The distribution of temperature, pressure and mass or density vary much both vertically, horizontally and temporally. Even though the horizontal and the temporal variations can look very different depending on the location on the planet, the vertical distribution have about the same appearance.

Density and pressure

Both density and pressure have about the same exponential decrease with height according to,

$$p = p_0 e^{-z/H} \quad (2.1)$$

where p_0 is the pressure (or density) at some reference height, usually at the surface and H is referred to as the e-folding depth or scale height which is the vertical distance over which the pressure has changed by a factor of e . This scale height varies depending on where you are located on the planet since it depends on the local climate and the topography, but in the lowest 100 km of the atmosphere the value lies between 7 and 8 km (Wallace and Hobbs 2006).

Temperature

The temperature of the atmosphere is overall decided by the radiation balance between the incoming and outgoing radiation on a given location. But when looking on a more regional or local scale closer to the surface, the temperature is not only controlled by the in and outgoing radiation, here the friction of the surface and wind patterns also play an important role. Partly because of the topography, but also the different land types that occurs on different elevated grounds.

The vertical distribution temperature provides the basis for the standard division of the atmosphere into layers. This division speaks of four different layers, the troposphere, stratosphere, mesosphere and thermosphere, see figure 2.1. This thesis will focus on the temperature in the lower part of the troposphere where the air quality plays an important role.

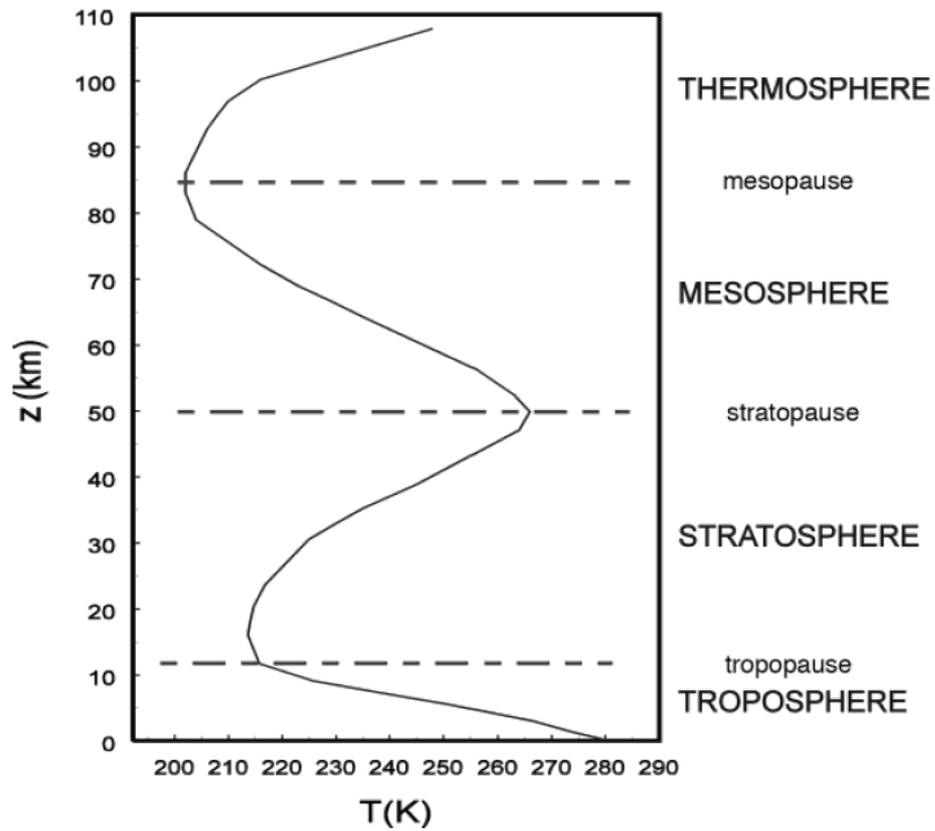


Figure 2.1: Vertical temperature profile for the 'US standard atmosphere' at 40 ° N in December (Marshall and Plumb 2008).

The troposphere is distinguished by a general decrease in temperature with altitude. The highest temperature is found closest to the ground and this is due to the compression from the above laying air, the warming from the surface and the higher absorption of radiation caused by the higher abundance of greenhouse gases near the ground. The average lapse rate, i.e. the rate of which the temperature decreases with height, of the atmosphere is expressed as,

$$\Gamma \equiv \frac{dT}{dz} \approx 6.5^{\circ}\text{C km}^{-1} \quad (2.2)$$

(Wallace and Hobbs 2006) where the lapse rate is defined as positive when having a decrease of temperature with height. This lapse rate is often referred to as the average adiabatic lapse rate and is in reality highly variable depending on the temperature, amount of water saturation and

mixing. For stable conditions of the atmosphere the values of Γ lies between $4.0\text{-}9.8^\circ\text{C km}^{-1}$ in theory, but commonly measured values lies between $6\text{-}7^\circ\text{C km}^{-1}$. Stable conditions means that when a parcel of air is uplifted it will cool down and expand, in order to adjust to the surrounding environment (Wallace and Hobbs 2006, ch. 3).

Different land types contributes both by affecting the air mass flow and how the radiation is being absorbed, i.e. the albedo. E.g. where there is forest, the canopy height creates bigger eddies in the turbulent flow than a field of grass would do. Which contributes to a larger mixing and thereby transport. The land use can also be categorized dependent on whether the land is elevated or not. In regions with complex high terrain there is not likely to be as much forest as down in a weather protected valley. Therefore the deposition is more likely to be much smaller in higher elevated regions and in coastal areas.

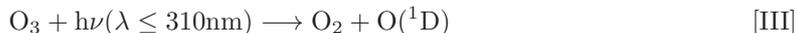
2.1.2 Ozone, O_3

Ozone, O_3 , is a molecule consisting of three oxygen atoms and is distributed over the two lowest layers in the atmosphere, the troposphere (0- ~ 15 km above surface) and the stratosphere ($\sim 15\text{-}50$ km above surface). Ozone is considered a reactive secondary air pollutant, which means that it is not emitted directly into the atmosphere, but photochemically produced through a series of different reactions depending on where the ozone is in the atmosphere.

Formation and removal processes

There are several processes that contributes to the concentration of ozone. Below follows a brief summary of some of the most important processes involved in the formation and removal of ozone in the atmosphere.

90% of the ozone lies in the stratosphere where it functions as UV screener from the sun. Here the life time of ozone is mainly controlled by the Chapman cycle,



where M is an abundant molecule that takes care of the excess energy from the new combined excited molecule. Other cycles also affect the stratospheric ozone, such as the HO_x , ClO_x and BrO_x cycle for example Finlayson-Pitts and Pitts (2000). The ozone in the stratosphere is very important to life on Earth because it absorbs harmful UV-radiation. The other 10% of ozone that exists in the troposphere is important since it is considered a greenhouse gas and a strong oxidant which is hazardous for living beings and vegetation.

The background concentration of ground level ozone is mostly due to downward mixing from the stratosphere and formation through a series of reactions with Volatile Organic Compounds (VOC) and oxides of nitrogen (NO_x). The final reaction that actually forms ozone is the addition of ground state atomic oxygen ($\text{O}({}^3\text{P})$) and molecular oxygen (O_2), see reaction ([VII]). O_2 is an abundant molecule while $\text{O}({}^3\text{P})$ is primarily produced through photolysis of NO_2 , see reaction ([VI]).



NO_2 is then either produced through oxidation of VOC's or emitted from some source (often anthropogenic).

There are then several different sinks of ground level ozone. One major chemical reaction is the conversion back to NO_2 ,



This reaction together with reaction ([VII]) are very fast compared to ([VI]), which means that the photolysis is the rate limiting step in this reaction chain. Another sink is dry deposition onto vegetation such as the surface of leaves and needles. Depending on the time a parcel of air with pollutants, is in contact with a leaf or a needle, how big the leaf area is and the time of the day (affecting the openings of the stomata), the dry deposition rate will vary.

Temporal and spatial variations

The peak concentrations of ozone during a day occurs on the early afternoon when there has been a build up of ozone through the photolysis of NO_2 and the mixing of O_3 rich air from higher altitudes. During the afternoon as the sun sets, the radiation intensity drops of the photolysis will decrease as well as the mixing of air. When the night comes the concentration will fall off down to its minimum value and the two dominant sinks consists of reaction ([VIII]) and the dry deposition onto the ground and vegetation. This diurnal variation depends strongly on the location and on the topography. Like for example, the diurnal variation on a mountain top is much smaller than the variation down in a valley. Since there in the valleys are more likely to be denser vegetation which during night times dominates the sink of ozone. On a mountain top there is more likely to be stronger winds, which contributes to the regional transport of ozone and NO_x away from the mountain top. Measurement of ozone during one day of a mountain top and a valley can then look like in figure 2.2 where measurements from two stations in Great Britain has been used.

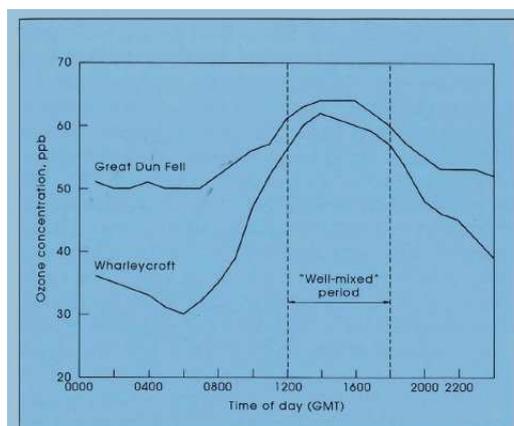


Figure 2.12 Mean ozone concentrations during photochemical episodes at Great Dun Fell and Wharleycroft.

Figure 2.2: Diurnal mean variation of ozone from Fowler et al. (1993)

2.1.3 Nitrogen dioxide , NO_2

Nitrogen oxides, NO and NO_2 referred to as NO_x , plays an important role in atmospheric chemistry. Nitrogen oxides are produced by fossil fuel combustion, biomass burning, transport from stratosphere, lightning, NH_3 oxidation and emitted from soils (Wallace and Hobbs 2006). NO_2 emissions results in a number of harmful secondary pollutant such as nitric acid, photo oxidants (e.g. ozone) and the nitric part in secondary inorganic aerosols (WHO 2003).

Formation and removal processes

The NO_x emissions in the troposphere consist primarily of NO in rural areas, which during the day establishes an equilibrium with NO_2 according to [VI], [VII] and [VIII]. There is no NO present during the night due to [VI]. The principal sink of NO_2 produces nitric acid, HNO_3



During night, when N_2O_5 is present, NO_2 is removed through the following processes



where HNO_3 is removed through wet and dry deposition. The resulting life time of NO_2 will be approximately 1 day (Wallace and Hobbs 2006). NO_2 is also produced from reactions with alkyl peroxy radicals (RO_2), which occurs due to emissions from soils (Finlayson-Pitts and Pitts 2000).



Temporal and spatial variations

The concentration of NO_2 is strongly dependent on urban activities and will therefore be high close to busy roads or other sources of NO_x . Close to roads, the amount of traffic can often be seen as two peaks in NO and NO_2 concentrations during the morning and afternoon traffic. The tropospheric NO_2 decreases with height because a large part is produced due to emissions on the ground. The concentrations are higher during the winter because of more emissions and lower concentrations of OH.

2.1.4 Sulphur dioxide, SO_2

Sulphur dioxide is the dominant anthropogenic air pollution containing sulphur. It contributes to the formation of acid aerosols and acidic precipitation and plays an important role in the sulphur cycle.

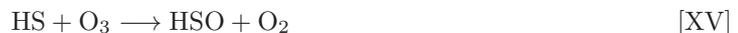
Formation and removal processes

The main natural sources of tropospheric SO_2 is volcanoes and dimethylsulfide (DMS). The main anthropogenic source is combustion of fossil fuels containing sulphur. The major removal process in the troposphere is wet deposition but it is also removed by dry deposition. SO_2 emissions leads to formation of HSO_3 due to fast reactions with OH. HSO_3 reacts with O_2 and forms SO_3 which in aerosols and clouds reacts with H_2O_2 and forms H_2SO_4 which is removed from the atmosphere in the form of acidic rain. (Speidel et al. 2006)

The formation of SO_2 through oxidation of DMS and H_2S is done for example through the following processes (Wallace and Hobbs 2006)



which forms HSO through the reaction with either NO_2 or O_3



HSO is then converted to SO₂



Volcanic activities can inject SO₂ into the stratosphere where it converts to H₂SO₄ which results in stratospheric aerosols.

Temporal and spatial variations

SO₂ concentrations are highest close to its source and decreases fast with the distance from the source. This is due to a short lifetime of a few days. The lifetime in the stratosphere is longer, several weeks, due to lower concentrations of OH. (Eisinger and Wegener 1998)

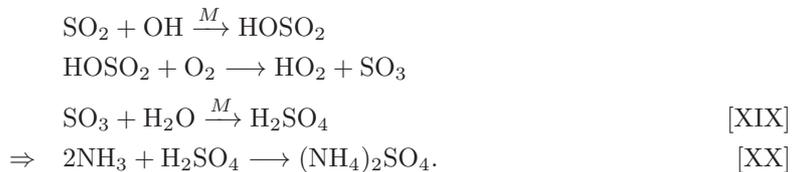
2.1.5 Sulphate, SO₄

Sulphate is a molecule consisting of a sulphur atom surrounded by four oxygen atoms and exists in the atmosphere primarily as aerosols and is considered a secondary air pollutant. The vertical distribution of sulphate is highly variable since it is strongly dependent on precursor gases and their emissions. The lifetime depends on where in the vertical it is situated. In the troposphere the lifetime is of about 5 days (Ménégoz et al. 2009) and in the stratosphere it can survive for several weeks. The sulphate aerosol have a huge impact on the radiative balance of the atmosphere since it is affecting the albedo both by scattering radiation on itself or by working as a cloud condensation nuclei and thereby creating water droplets that scatters the radiation.

Formation and removal processes

The most important sources of sulphate in the atmosphere is the oxidation of sulphur dioxide (SO₂), dimethyl sulphide (DMS) and hydrogen sulphide H₂S. These precursor gases comes from different emission sources. SO₂ originates from both anthropogenic, such as combustion of S-containing fuels in power plants and vehicles, and volcanic emissions. DMS is emitted in the marine boundary layer from dying phytoplankton in the upper layer of the ocean. H₂S comes from anaerobic decomposition of S-containing compounds by micro organisms and is emitted mostly into the atmosphere through volcanoes, hot/cold springs and petroleum refineries.

All the three above mentioned precursor species will go through the oxidation of SO₂. Because both DMS and H₂S will react with the hydroxyl radical and eventually turn into SO₂. The oxidation of SO₂ to sulphate can then be either in a gaseous or aqueous way. The gaseous reaction scheme includes reactions with different radicals to form H₂SO₄ which then reacts with NH₃, see reaction ([XIX]) and ([XX]). This is a rather slow process which takes about one day (Liang and Jacobson 2009)



The aqueous way involves dissolved SO₂ which oxidise with different oxidants depending on the pH, content of liquid water and the availability of the oxidants, to sulphuric acid. The two most dominant oxidants are O₃ and H₂O₂ with pH<5 or pH>5. The sulphuric acid from both the gaseous and aqueous pathways then undergoes nucleation with some other gas species and forms a particle. The most relevant removal process of sulphate is wet deposition.

Temporal and spatial variations

The diurnal variations of sulphate is not as strong as the one for ozone and the temporal variations depends a lot on the regional emissions of precursor gases. The spatial variation also depends strongly on the regional precursor gas emissions. If one looks at the sulphate from the non marine boundary layer the spatial distribution over Europe is about the same, but in some regions where there are a lot of topographical changes like in the Alps and in southern Norway local maxima with increased wet deposition is observed (Torseth et al. 2012).

2.2 EMEP model

The EMEP MSC-W model is a huge model that takes many physical and chemical principles into account. In this section a short presentation of the 3D coverage of the model and what type of input and output data the model needs and gives. A more thorough description will then follow about how the model treats the dry deposition, since this is an important process depending on the topography that will be analysed.

2.2.1 Model coverage

The EMEP MSC-W model provides data divided into 50x50 km (at 60 degrees N) grids. The grids are defined on a polar stereographic projection. (The EMEP MSC-W Model - User's Guide, 2011) The EMEP area with its 132x159 grid points is shown in figure 2.3. These grid squares, or coordinates, are labelled i (east west) and j (south-north) related to latitude (ϕ) and longitude (λ) accordance to the following,

$$i = i_{pol} + M \tan \left[\frac{\pi}{4} - \frac{\phi}{2} \right] \sin(\lambda - \lambda_0) \quad (2.3)$$

$$j = j_{pol} - M \tan \left[\frac{\pi}{4} - \frac{\phi}{2} \right] \cos(\lambda - \lambda_0) \quad (2.4)$$

where M is the number of grid distances from the North pole to the equator and depends on the size of one grid space. i_{pol} and j_{pol} are the i and j coordinates of the north pole. λ_0 is the longitude parallel to the j axis and ϕ_0 defines latitude.

Each grid has a vertical resolution of 20 layers defined with σ coordinates.

$$\sigma = \frac{p - p_t}{p_s - p_t} \quad (2.5)$$

where p_s , p_t and p are the pressure at the surface, at the top of the model (100hPa) and at pressure surface σ (Simpson et al. 2012). The different layers are shown in figure 2.4. The heights in a standard atmosphere for the boundaries is shown on the right side while the sigma levels are shown on the left. The different levels are indexed with different k -values with $k=20$ for the lowest layer and $k=1$ for the top layer. The values for the lowest layer correspond to an altitude of about 45 m and from this layer a 3 m layer can be defined by extrapolating down by using the equality of the depositoin flux at the boundary of the vertical layers of 3m and 45m.

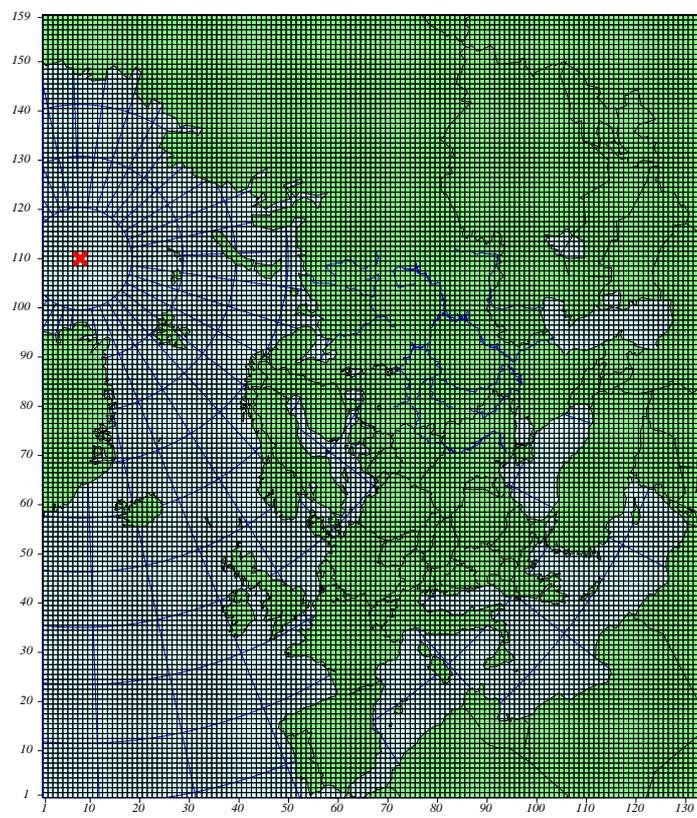


Figure 2.3: EMEP grid

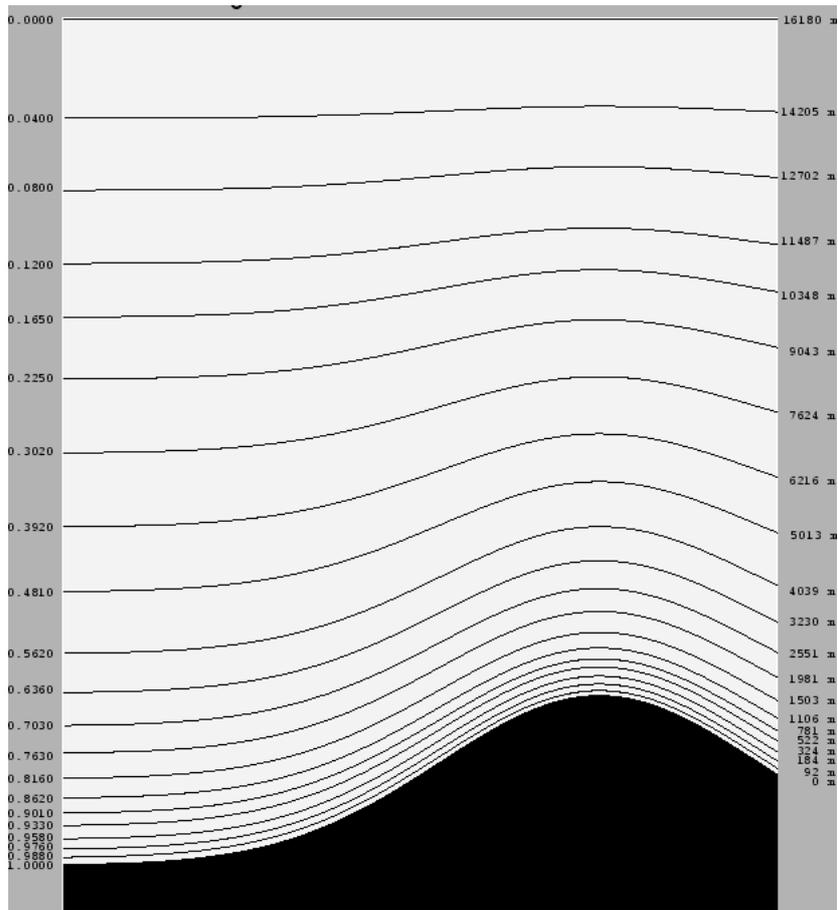


Figure 2.4: Vertical layers in the EMEP model

2.2.2 Model data

The EMEP model needs several different inputs, among them meteorology data, boundary conditions and so on. Some of the input files to the EMEP model are listed below,

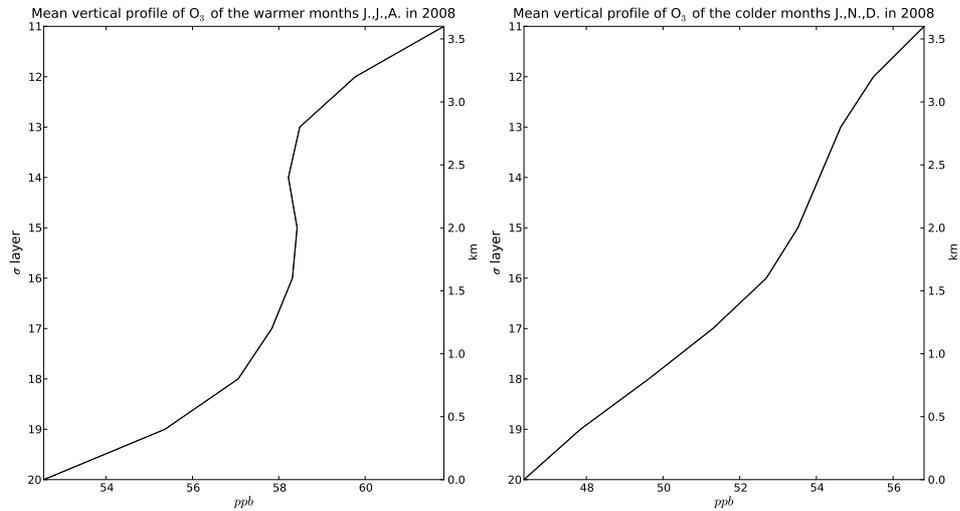
- Meteorology data based on runs with the global operational forecasting model called Integrated Forecast Systems (IFS) from the European Centre for Medium Range Weather Forecast (ECMWF)
- Global O₃ as initial concentrations of O₃
- Land use-fractional land use for each grid
- Gridded emissions as annual national emissions

for a full list of input files, see the user guide (MSC-W 2011).

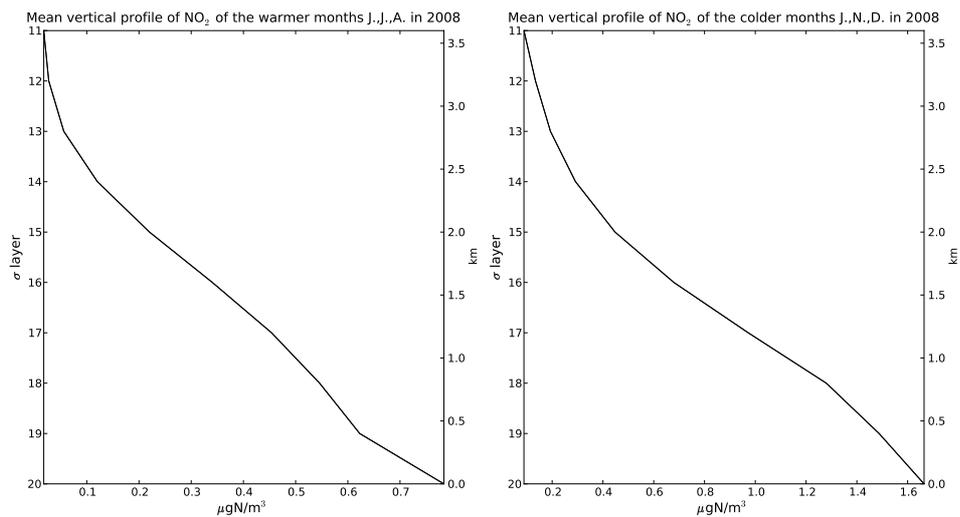
There can be several different outputs from the model. For validation, model data for coordinates corresponding to the coordinates of measurement stations can be chosen. The output files are then two ascii files, one containing the concentration for each vertical layer corresponding to the σ -layer and one containing the 'surface concentration' at 3 m above ground. Gridded concentrations are written to netCDF files which can contain either 2D or 3D data.

2.2.3 Model vertical profiles of the pollutants

In order to be able to analyse the results in the end, the vertical profiles of the different pollutants needs to be derived from the model. Vertical profiles of O₃, NO₂ and SO₂ for the station Jungfrauoch, CH01 in Switzerland can be seen in the figures 2.5(a), 2.5(b) and 2.6(a). The vertical profile for SO₄ can be seen in figure 2.6(b) for the station SK06 in Slovakia. The reason for having another station for SO₄ is because there are no measurement of this pollutant for CH01. See table A.2 and A.3 for altitude information about the stations.



(a)

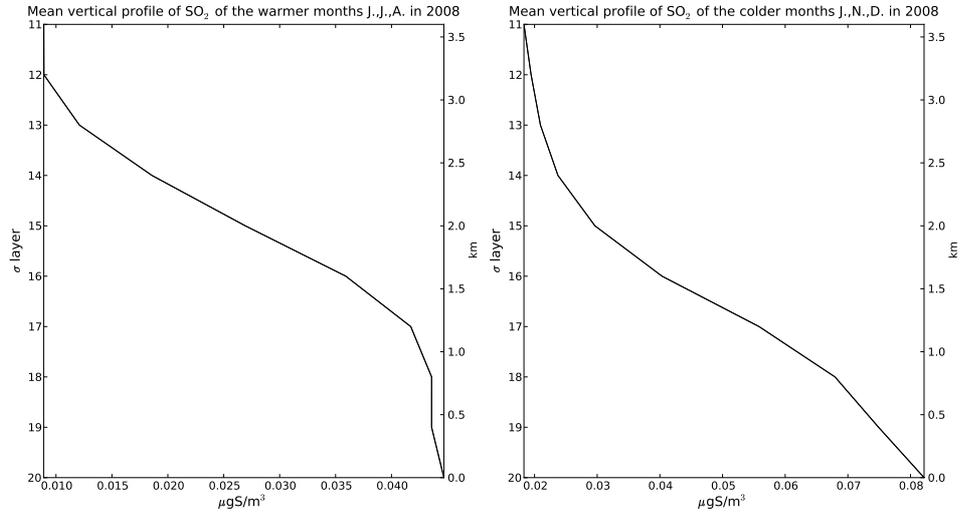


(b)

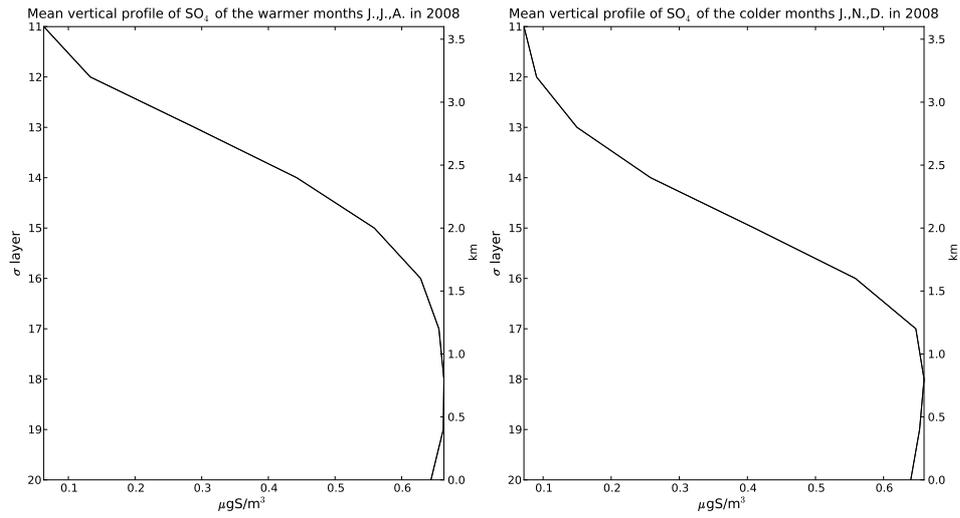
Figure 2.5: The average vertical profile of O₃ and NO₂ at the stations Jungfraujoeh CH01 in Switzerland.

2.2.4 Deposition

One of the most important topography specific processes is the deposition which is divided into dry and wet deposition. This division refers to the way something gets transported to the surface where the deposition takes place, either by a dry process where the pollutant or the aerosol reaches the surface by getting transported as a gas in the air or by a wet process where it deposits to wet surfaces. In this thesis the dry deposition will be in focus and analysed in terms of O₃ deposition, therefore the following sections will treat how this is modelled in the EMEP MSC-W model.



(a)



(b)

Figure 2.6: The average vertical profile of SO_2 at the station Jungfrauoch CH01 in Switzerland and SO_4 at the station Starina SK06 in Slovakia.

Dry deposition

The dry deposition of a compound i over a surface within a volume of unit area and height Δz is given by

$$\frac{dC_i(z_{ref})}{dt} = \frac{-V_g(z_{ref}) \cdot C_i(z_{ref})}{\Delta z} \quad (2.6)$$

where $V_g(z_{ref})$ is the deposition velocity (m/s) at height z_{ref} , which is calculated using

$$V_g = \frac{1}{R_a + R_b + R_c} \quad (2.7)$$

where R_a is the aerodynamic resistance which handles the transport of the turbulent layer, R_b is the quasi-laminar layer resistance which handles the movement of the air mass closest to the surface and R_c is the vegetation canopy or surface resistance. R_a and R_b depends mostly on meteorological parameters and will not be treated here, while R_c is strongly dependent on the land use. The surface conductance $G_c = 1/R_c$ is calculated (Simpson et al. 2012)

$$G_c = LAI \cdot g_{sto} + G_{ns} \quad (2.8)$$

where LAI is the Leaf Area Index (m^2/m^2), G_{ns} is the bulk non-stomatal conductance and g_{sto} is the stomatal conductance. LAI depends on the land use and time of the year. There are 16 land classes defined in the model which are shown in table A.5. The start and end of growing season (SGS, EGS), the minimum and maximum LAI and the length of the LAI increase and decline is defined as inputs to the model. These parameters are shown in figure 2.7 and the values for the different land classes are shown in table A.5. The stomatal conductance, g_{sto} , depends on land class and also temperature and pressure. G_c is calculated different depending on the atmospheric gas. How to calculate it for O_3 is described in chapter 2.2.5. Other gases are scaled from the $G_c^{\text{O}_3}$ with diffusivity ratios given in Simpson et al. (2012, ch. 8).

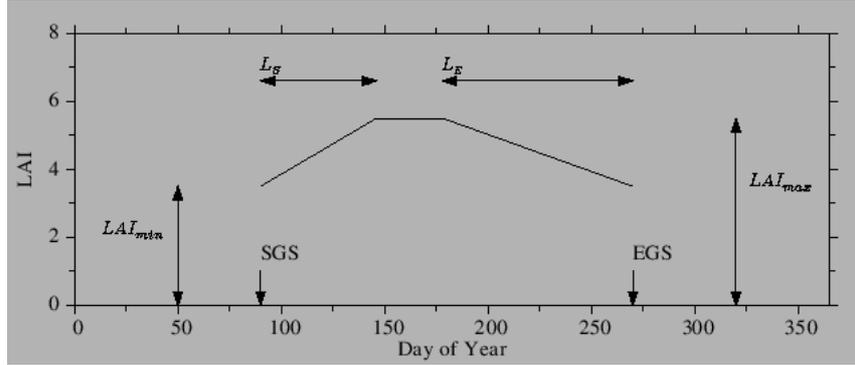


Figure 2.7: The definition of the Leaf Area Index taken from Simpson et al. (2012).

2.2.5 Ozone in the EMEP MSC-W model

In order to start a model run, initial O_3 concentrations needs to be provided. The boundary conditions for O_3 are obtained using 3D fields from either global observational data sets or Chemical Transport Models (CTM). The observational data set used are based on sonde measurements together with surface and satellite measurements (Simpson et al. 2012).

The used data sets have a limited number of O_3 sondes and underestimates concentrations on the west coast of Europe. The data is therefore adjusted depending on measurements done at a site called Mace Head on the west coast of Ireland (Simpson et al. 2012). It has been shown that this site is well suited for measurements on background O_3 for mid-latitude air masses. After taking into account the origin of the air masses, the measured concentrations are compared with concentrations from global dataset corresponding to the south west quadrant of the EMEP-domain. The difference between the global data set and Mace Head measurement is added to the O_3 boundary conditions. This corrections shows good results for almost all stations on the west coast of Europe (Simpson et al. 2012).

Dry deposition of O_3

The stomatal conductance, g_{sto} in equation (2.12), is calculated according to:

$$g_{sto} = g_{max} \cdot f_{phen} \cdot f_{light} \cdot \max\{f_{min}, f_T \cdot f_{VPD} \cdot f_{SWP}\} \quad (2.9)$$

where g_{max} is the maximal stomatal conductance specific for each land class. The factors f_x has values between 0 and 1 and are used to modify g_{max} with seasonal variations. The definition of these factors are the time of year leaf phenology (f_{phen}), light (or photon density flux, f_{light}), leaf temperature (f_T), leaf-to-air vapour pressure deficit (f_{VPD}), soil water potential (f_{SWP}) and the minimum stomatal conductance f_{min} ensures that there always will be some minimal stomatal uptake (Simpson et al. 2012, Emberson et al. 2000, Simpson et al. 2003) during daytime when $f_{light} \neq 0$. All these f parameters are changing with the growing season and for a detailed description of how they change see Emberson et al. (2000) and see table A.1 for values for the different land classes. Since the limitations of the analysis of the dry deposition lies within the downscaling of land class, temperature and pressure only the expression that contains temperature and pressure will be presented here.

The maximal stomatal conductance values for each land class are found in Table A.1 and has the unit of $\text{mmol m}^{-2} \text{s}^{-1}$. In order to use this within the resistance it has to be converted to m s^{-1} , which is done by the ideal gas law and thereby the g_{max} is temperature and pressure dependent.

$$g_{max} = g_{max}^m \cdot \frac{RT}{P} \quad (2.10)$$

Two of the f factors in g_{max} are temperature dependent, f_T and f_{VPD} . The leaf temperature factor is calculated as,

$$f_T = 1 - \left(\frac{T - T_{opt}}{T_{opt} - T_{min}} \right)^2 \quad (2.11)$$

where T_{opt} and T_{min} are land class dependent parameters and T is supposed to be the leaf temperature, but is assumed to be the same as the 2 m temperature T_{2m} (Emberson et al. 2000). The leaf-to-air vapour deficit is defined as the difference between the vapour pressure in the air and vapour pressure within the sub-stomatal cavity and is therefore dependent on the surrounding temperature. Modelling of this VDP parameter is difficult and there is likely to be uncertainties (Simpson et al. 2003). Therefore the temperature dependence of this parameter will not be considered.

The non-stomatal conductance, G_{ns} used to calculate the canopy conductance, G_c in equation (2.8), for O_3 is calculated

$$G_{ns}^{O_3} = SAI \cdot g_{ext} + \frac{1}{R_{inc} + R_{gs}^{O_3}} \quad (2.12)$$

where g_{ext} is the external leaf-conductance ($g_{ext}=1/2500 \text{ ms}^{-1}$), R_{inc} is the in-canopy resistance, R_{gs} is the ground surface resistance and SAI is a surface area index (m^2/m^2). R_{gs} depends on the land use and modified due to snow-cover and low temperatures.

$$R_{gs}^{O_3} = \hat{R}_{gs}^{O_3} + R_{low} + 2000 \cdot \delta_{snow} \quad (2.13)$$

where values for $\hat{R}_{gs}^{O_3}$ can be found in table A.6, δ_{snow} is 1 for snow cover and 0 otherwise, R_{low} is a correction term for lower temperatures, $T < 0^\circ \text{C}$ and is defined a

$$R_{low} = 1000e^{-(T+4)}, \quad T < 0^\circ \text{C} \quad (2.14)$$

R_{inc} is the in canopy resistance and is dependent on the land use and the friction velocity u_* of the atmosphere,

$$\frac{b \cdot SAI \cdot h}{u_*} \quad (2.15)$$

where $b=14\text{s}^{-1}$ and h is the canopy height. The surface area index, SAI, for forest is set to LAI+1 and for non-crop vegetation to LAI.

2.3 Literature Reviews

Several other papers have been investigating how the topography affects the air quality of the lower part of the atmosphere. Among them are Klingberg et al. (2012) who conducted a measurement survey of ozone and temperature in southern parts of Sweden with the aim of investigating the influence of topography. In their analysis they defined a relative altitude as the average height within an area of 3 km around the site subtracted from the altitude of the site, for each measurement site. They concluded that inland sites with a low relative altitude experience stronger diurnal variations of ozone than coastal sites and sites with higher relative altitude. This was then also correlated with the diurnal temperature range, where at the low land sites there were stronger nocturnal inversions.

The same results about the diurnal variations for ozone at different altitudes were found years earlier by Fowler et al. (1993). They used measurement data for ozone from a large measurement network in the United Kingdom and developed an interpolation method that derived the ozone levels between the sites using the absolute altitude as the variable. With this technique they created ozone maps over the United Kingdom.

Similar work was also done by Loibl et al. (1994), who used a large number of monitoring sites (99) distributed over Austria. Measurements from these stations were used together with topographic data in order to do an interpolation over the complex terrain in Austria. Instead of using the absolute altitude, Loibl et al. (1994) used a relative altitude defined as the minimum height within a radius of 5 km surrounded a given point subtracted from the absolute altitude. Ozone maps over Austria were created where the uncertainty usually was within 10 ppb.

3

Method

In this chapter the methodology of the sub-grid modelling will be described. First three new different extraction methods will be described together with new 1x1 km topography and how this will be used in the different methods. There are stations in the set of available measurement stations that are not located on high altitudes, therefore there has to be a selection to what stations should be used, this is described in the following section. The results from the different extraction methods has to be analysed and a 'best' method has to be chosen, that analysis and how a method is chosen are described in terms of the four different pollutants since they have either hourly or daily data.

3.1 The extraction methods

The most frequent extraction method for validation is, as mentioned in the introduction, the interpolated '3 m' level, which from now on will be called the 'site' level. Three new extraction methods will be tested based on three different altitude information set, the absolute altitude h_s , the difference between h_s and the EMEP grid altitude h_E or the relative altitude h_r which is described in the section below. The three extraction methods are defined as,

$$h_s = \text{Absolute altitude} \quad (3.1)$$

$$h_{s,E} = h_s - h_E, \quad h_E = \text{EMEP grid altitude} \quad (3.2)$$

$$h_r = \text{Relative altitude.} \quad (3.3)$$

The altitude information of the used station can be found in table A.2 and A.3. These three extraction methods can be motivated in different ways. The absolute altitude or h_s method looks at the real topography and assumes that the most important geographical information regarding topographic effects is the altitude above sea level and that the EMEP topography as no affect on the outgoing pollutant concentrations. The $h_{s,E}$ method differs from the h_s method by assuming that the pollutant concentrations do depend on the EMEP topography, since both the T_2 temperature (temperature 2 m above the used topography) and the surface pressure P_s is calculated accordingly. The last method about the relative altitude h_r is described below.

3.1.1 Relative altitude

The relative altitude in this thesis is defined as the altitude of a given location (x_0, y_0) subtracted with the lowest altitude within a search radius R ,

$$h_r(x_0, y_0) = h(x_0, y_0) - \min(h(x, y)) \quad (3.4)$$
$$(x, y) \in \{(x, y) : \|(x, y) - (x_0, y_0)\| \leq R\}.$$

What is claimed when this relative altitude is used as an extraction height is that it is more important to look at the relative surrounding when looking at the topographical effects. The

question is then how much of the surrounding is supposed to be accounted for? In this thesis three different search radius has been used, 3, 5 and 10 km in order to see where the topographic effects are best captured.

In order to create this relative altitude a finer resolution topography has to be used since the EMEP grid only has a resolution of 50x50 km. The used topography for this relative altitude derivation comes from The Oak Ridge National Laboratory Distributed Active Archive Center (ORNL DAAC) for biogeochemical dynamics (for Biogeochemical Dynamics)) and is a global 30 Arc-Second Elevation Data Set (GTOPO30). The coverage of this data is: latitude from 33 to 79.99 ° N and longitude 25° E to 39.99° W. Each grid cell corresponds to 0.00833 degrees and can approximately be converted to meters with data from table A.1 which comes from the documentation of the GTOPO30 data (USGS). Figure 3.2 shows the GTOPO30 topography and figure 3.1 shows the relative altitude for a search radius of 5km.

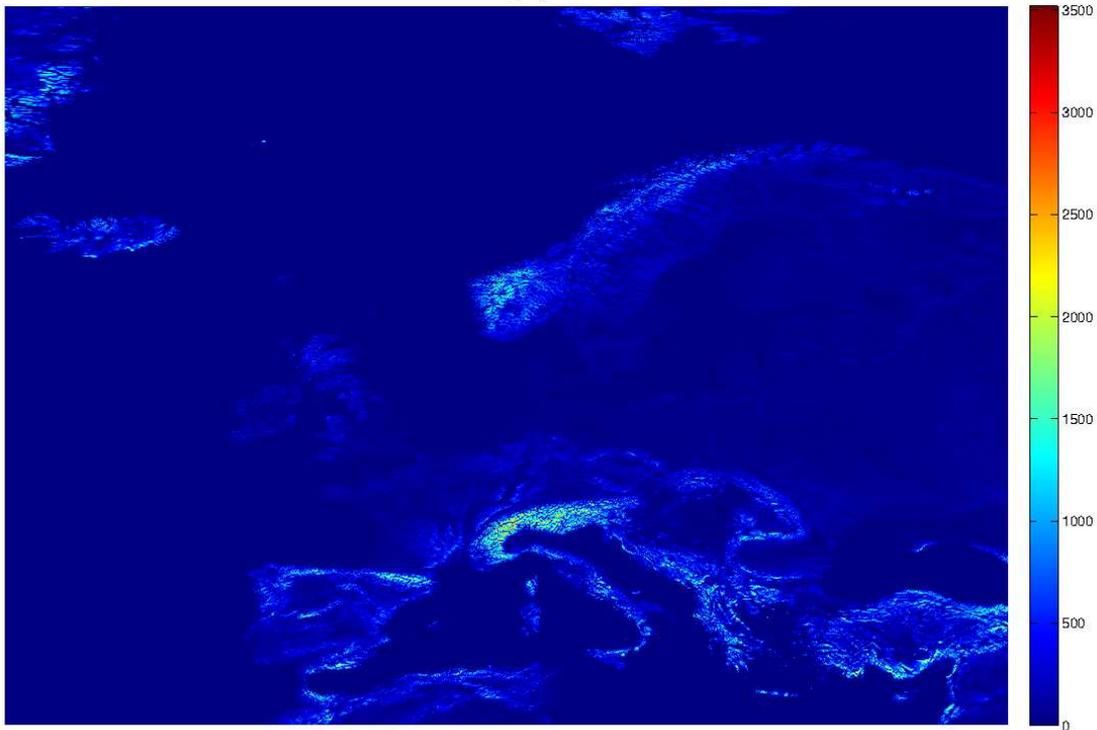


Figure 3.1: The relative altitude for central Europe with a search radius of 5 km.

3.1.2 Station selection

The stations that are used in this validation process with different extraction methods are stations with a relative altitude, $h_{r,10} > 45\text{m}$. If the relative altitude would be smaller than this the extraction would show little or no difference. Also since the main validation problems lies within the higher elevated stations it is better to choose a higher h_r . Most of the measurement stations that are used for validation of the EMEP model are rural sites with low influence from roads and cities. The stations selection with a $h_{r,10} > 45\text{m}$ resulted in a total number of 89 measurement stations which can be seen in figure 3.2.

3.1.3 Interpolation

This thesis also aims to do sub-grid modelling when it comes to the vertical layers, therefore a linear interpolation between vertical layers is implemented for the concentrations. The concentration for

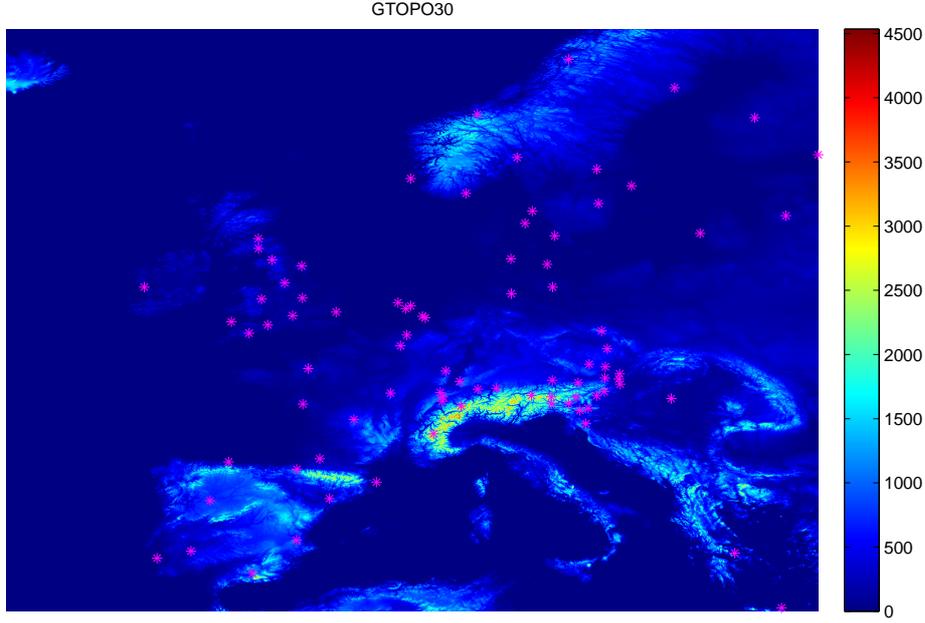


Figure 3.2: The distribution of the used stations.

each level corresponds to the mid altitude of that layer (Simpson et al. 2003). If the station altitude were lower than the mid altitude of the used layer, the interpolation was done with respect to the layer below and vice versa for altitudes above. The concentration for a station s with corresponding layer k is calculated

$$C_s = C_k + \Delta C \quad (3.5)$$

where

$$\Delta C = \begin{cases} \frac{C_{k+1} - C_k}{h_{mid,k+1} - h_{mid,k}} (h_s - h_{mid,k}) & \text{for } h_s > h_{mid,k} \\ \frac{C_{k-1} - C_k}{h_{mid,k-1} - h_{mid,k}} (h_s - h_{mid,k}) & \text{for } h_s < h_{mid,k} \end{cases} \quad (3.6)$$

where C_k , C_{k-1} and C_{k+1} is the concentration for the current layer, the layer below and above. h_s is the (relative) altitude for the station s and h_k , h_{k-1} and h_{k+1} is the mid altitude of the different layers.

3.2 Analysis of the pollutants

There were different amount of measurement data available for the pollutants and they show different vertical behaviours, therefore the analysis of them became somewhat different. How the analysis for each pollutant was performed is described in this section.

3.2.1 O₃

Hourly O₃ measurements was available for 87 stations with a relative altitude higher than 45 m for the search radius of 10 km. The measurement data have the unit $\mu\text{g}/\text{m}^3$ while the model data

has ppb. The conversion factor In order to convert from $\mu\text{g}/\text{m}^3$ to ppb, the following a conversion was used,

$$c[\text{ppb}] = c[\mu\text{g}/\text{m}^3] \cdot 10^{-6} \frac{RT}{PM} 10^9 \quad (3.7)$$

where c is the concentration and M the molecular weight.

ΔO_3 analysis

The EMEP model has generally too high O_3 levels, which can be seen when validating the model data of the lowest layer (site) with measurements, see figure 3.3. The new extraction methods would then result in even higher O_3 values. In order to analyse the new extraction methods the diurnal variations is therefore investigated instead. As validation of the different extraction methods a mean day was created for each station over the year and the difference between the maximum ($O_{3,max}$) and minimum ($O_{3,min}$) value was derived,

$$\Delta O_3 = O_{3,max} - O_{3,min} \quad (3.8)$$

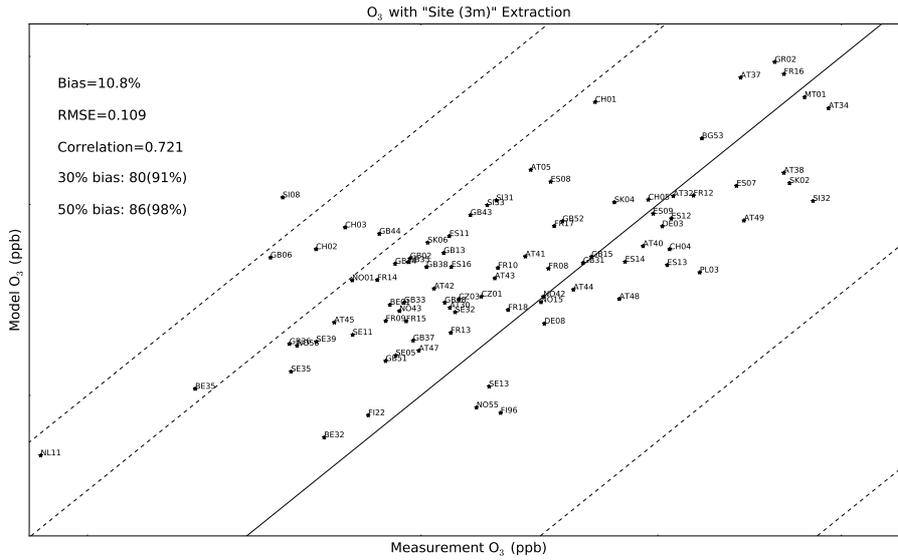


Figure 3.3: The yearly mean of O_3 for the 87 stations with site extraction.

An analysis of the even higher stations with a relative altitude above 200 m was also conducted to see whether there had been any improvements for these stations. An error analysis was conducted on the four new methods where the difference between the error from the site extraction method and the other methods were derived according to,

$$\Delta \text{Error} = \left| \Delta O_{3,meas} - \Delta O_{3,site} \right| - \left| \Delta O_{3,meas} - \Delta O_{3,i} \right| \quad (3.9)$$

where i is some of the other four extraction methods. In order to see how the different extraction methods behave during the summer months of June, July and August and the winter months January, February and December, an analysis of the ΔO_3 will also be conducted for these periods.

Ozone maps

One of the aims is to create maps of O₃ over Europe with one of the extraction methods. In order to create a map 3D data over Europe has to be used. The EMEP MSC-W model provided us with hourly 3D data for 9 of the lowest vertical layers and the 3 m surface layer. The land coverage of this data with EMEP MSC-W model coordinates i and j are between 1 and 100.

3.2.2 NO₂

Hourly measurement data for NO₂ was available for 25 of the selected stations. In order to determine then best method for extraction of NO₂ concentrations, the yearly mean value was investigated. The analysis was done on mean concentrations instead of diurnal variations because the diurnal variations depends strongly on the urban activities in the area. The mean values retrieved from the model is also closely related to the measurements which makes it possible to do a good analysis using mean values.

3.2.3 SO₂ and SO₄

Both SO₂ and SO₄ had only daily measurements available. For SO₄ only 10 of the chosen stations had measurement data, 21 had SO₂ measurements. The yearly mean for the stations was calculated and analysed. Daily measurements of SO₄ was available for 10 of the selected high stations. As for SO₂ the yearly mean is investigated for the different extraction methods and the measurements.

3.2.4 Best method selection

In order to choose the best method, three statistical parameters were calculated. The Pearson product-moment coefficient which measure the linear dependence between model and measurement data. The root mean square error (RMSE) sums the squares of the differences between model and measurements. The total bias of the results shows the amount of systematically errors. The statistical parameters regarded as most important are RMSE and the amount of stations within 30% bias. However, the method aims at improving the results for sites positioned at high altitudes, so the stations are investigated individually as well. Even though a method shows good results over all, it might not perform as wanted for the highest stations. This analysis is important for the species were the number of stations with measurement data is limited.

3.3 Sub-grid extraction model scripts

A flowchart of the scripts used to make scatterplots of the model and measurement concentrations is shown in figure 3.4. The source code can be seen in B and descriptions of the function of the scripts in the sections below. The used EMEP model output for this sub-grid model are one file containing the 20-layered data for each selected station which is called 'sondes' and one file containing the site layer value for each station which is called 'sites'. Two of the scripts are named with NILU, which corresponds to the Norwegian Institute for Air Research from where we have been getting the measurement data.

3.3.1 sondes_read.py

This script contains a function that reads the elevated model data from the model output. This program has to read the file two times with `numpy.loadtxt`, first in order to extract names and (i,j) coordinates about the stations and the second time is to read all the data, but here it skips the first rows containing all the above information. This had to be done since the file did not have the same amount of columns everywhere, which led to loss of data further down. The input parameters are which month and species to get data from.

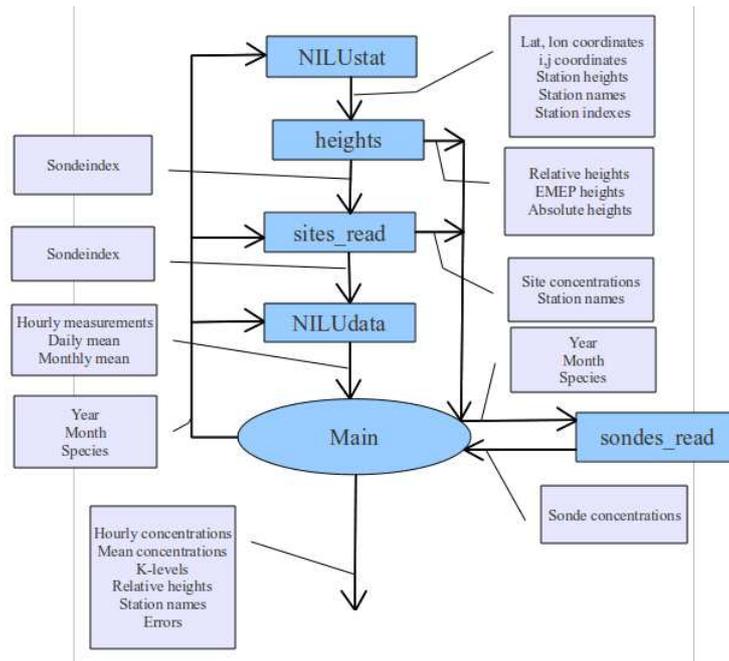


Figure 3.4: Flowchart of the scripts used to create scatterplots

3.3.2 NILUstat.py

NILUstat provides information about the stations which have both measurement data and model data. The parameters provided are: coordinates (both latitude, longitude and EMEP (i,j)), altitudes, names and indexes for the different stations. The function loads a text file with the names of each station and compare these with the station names in a netCDF file containing NILU measurements. When the station names are confirmed, the i,j coordinates are extracted from the text file with the stations and the latitude and longitude coordinates are extracted from another text files containing names and coordinates for all EMEP stations.

3.3.3 heights.py

This script contains a function that gives relative altitudes for four different requisition regions with a radius of 3, 5 and 10 km. It also gives EMEP altitudes, EMEP coordinates (i,j) and indexes for stations that both have levelled model and measurement data. The incoming data is (lat, lon) and (i,j) coordinates, altitude and indexes for these above mentioned stations.

The script starts by reading topography data from GTOPO30. Then it finds the stations located in the area and extract their positions in the map. The next step is to read the EMEP topography data and find the corresponding 'EMEP altitudes' for the stations. Then it calculates the relative altitude for each stations, by using the right amount of grid cells that corresponds to the different search radius. The number of grid cells is decided by using table A.1.

3.3.4 sites_read.py

This script reads model data from the files that contain the lowest layered data at 3 m called sites. The input parameters are month, year and the indexes for the stations that has passed through NILUstat.py and heights.py. The outputs are hourly or daily concentrations for each station together with the station names and indexes.

3.3.5 NILUdata.py

This function provides NILU measurement data for the station which have both sonde data, NILU data and site data. The indexes of the stations are provided by sites_read.py. The measurement data is extracted from the NILU netCDF files using the names corresponding to the sonde indexes. The function provides hourly and/or daily concentrations as well as the corresponding station names.

3.3.6 main.py

This script calls upon all the above mentioned scripts/functions. Here both measurement and model data is collected for each station over the year 2008. The output consists of hourly or daily data for each of the four new extraction methods, lowest layer extraction (site extraction) and measurement data for all the stations.

The script is running the above smaller functions for each month for the chosen pollutant and prints out hourly or daily data files for the different extraction methods. The information regarding the stations and their absolute, relative and EMEP altitudes together with the extracted layer of the model data are also printed out when the complete year has been gone through.

3.4 Dry deposition downscaling

As an investigation of a possible further development of this sub-grid model, the dry deposition is downscaled to a finer resolution in order to see if there are big changes. The reason for looking at the dry deposition is the high EMEP model mean values of O_3 , see section 3.2.1 and since the dry deposition is a very important sink for O_3 . If a downscaling of the dry deposition would differ much from the previous values it might be interesting to develop a correction term to the otherwise high model means. The diurnal variation of O_3 is also likely to be affected by a downscaling of O_3 , because at higher altitude there it is less likely to be dense vegetation and therefore there would be less surfaces for O_3 to deposit on. The downscaling will therefore only be analysed in terms of O_3 in this thesis.

Downscaling of the dry deposition means a downscaling of the land use, temperature and pressure. These three factors implies changes in the dry deposition velocity V_g in equation (2.7).

3.4.1 Land use

The land use dependent part of the dry deposition is investigated by comparing the EMEP land use with a higher resolution land use data set. Three groups of land classes are investigated; water, grass and forest. Water for EMEP land use includes; water (W), ice (I) and wetlands (WE). The forests group includes; temperate/boreal coniferous forests (CF), temperate/boreal deciduous forests (DF), Mediterranean needleleaf forests (NF) and Mediterranean broadleaf forests (BF). The land classes used as grass are seminatural/moorland (SNL), grassland (GR) and Mediterranean scrub (MS). The classes regarded as grass is the classes which is not forest and crops and have a LAI which is not zero. The water classes all have $LAI = 0$. Information about the land classes can be found in table A.5 and A.1 in Appedix A.

The EMEP land use is compared to the 1x1 km land use data set MODIS(MCD12Q1) land cover type 1 which uses 17 different land classes. The data set is for 2007 and the land classes are defined by the International Geosphere Biosphere Programme (IGBP). The classes regarded as forest are; evergreen needleleaf forest, evergreen broadleaf forest, deciduous needleleaf forest, deciduous broadleaf forest and mixed forests. The classes regarded as water are; snow and ice, permanent wetlands and water. Grass classes are; closed shrubland, open shrubland and grassland.

The relationship between errors and land use difference between EMEP and the high resolution dataset was investigated. It is assumed that the land class which gives largest difference in deposition is water, therefore the relationship between errors and the difference in water is investigated. The land use difference for forest and grass depending on altitude is also analysed. The

land use from the 1x1 km data set was extracted from an area with radius 5 km. The difference is calculated by subtracting the amount (percentage) of the land class group for the high resolution data set by the amount for EMEP land use.

3.4.2 Temperature and pressure

The temperature dependence of the canopy conductance, R_c first of all lies within the stomatal conductance g_{sto} , but when temperatures ranges below 0° C also the non-stomatal conductance, G_{ns} will be dependent on the temperature. The pressure dependence is assumed to only lie within g_{sto} , with accordance to (2.10).

When downscaling the temperature and pressure a fixed lapse rate and scale height is assumed, see the equations (2.2) and (2.1). The lapse rate is assumed to have the value of the average adiabatic lapse rate Γ and the scale height H is assumed to be 7 km. The altitude difference z will correspond to the extraction method that shows best results for ΔO_3 .

In order to see how R_c changes with pressure and temperature, three land classes were used in order to get parameter values, coniferous forest, grassland and water (including ice). The parameters were chosen to fit an ideal phenology situation, where all the f-factors in (2.10) and LAI is assumed to be at maximum vales. In order to see how the canopy resistance R_c changed with the temperature and pressure, they were allowed to change within the range of downscaled temperature and pressure of the selected stations.

4

Results

4.1 O₃

The extraction method for the diurnal variation of O₃ which shows largest improvements according to most station within the limit of 30% is the method that extracts O₃ with the h_{r,3}, i.e. relative altitude with a search radius of 3 km. This method shows 48 stations within the limit of 30%, see figure 4.2. Compared to the site extraction figure 4.1 this is not a big improvement, but the amount of stations within 50% is a lot more and the other statistical parameters has also been improved, see table 4.1.

Table 4.1: The statistical parameters of the extraction methods for the diurnal variation of O₃.

	bias	RMSE	Correlation	30% Bias	50% Bias
h _s	-48%	0.257	0.505	18(20%)	45(51%)
h _{s,E}	-20.7%	0.168	0.706	47(54%)	76(87%)
h _{r,3}	-20.8%	0.156	0.784	48(55%)	77(88%)
h _{r,5}	-26.2%	0.169	0.765	44(50%)	76(87%)
h _{r,10}	-31-2%	0.185	0.749	37(42%)	66(75%)

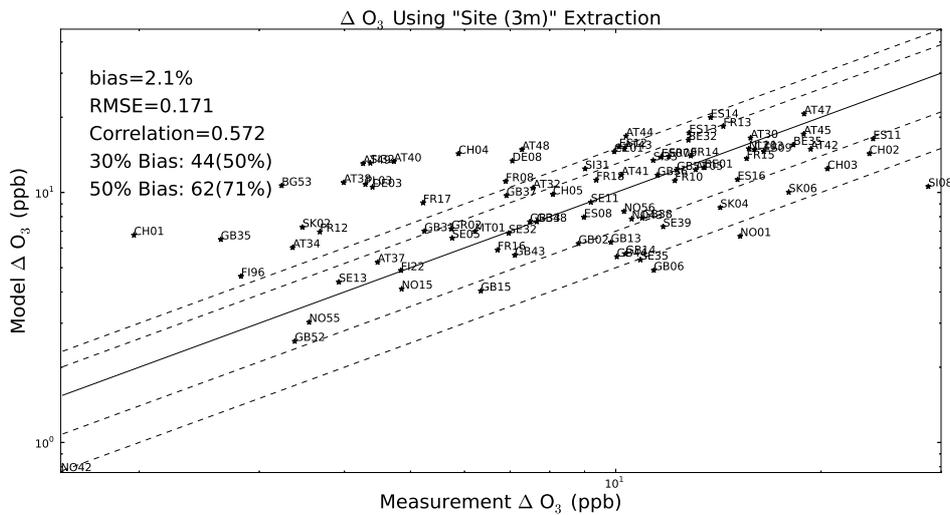


Figure 4.1: Scatterplot with the 'site' extraction method.

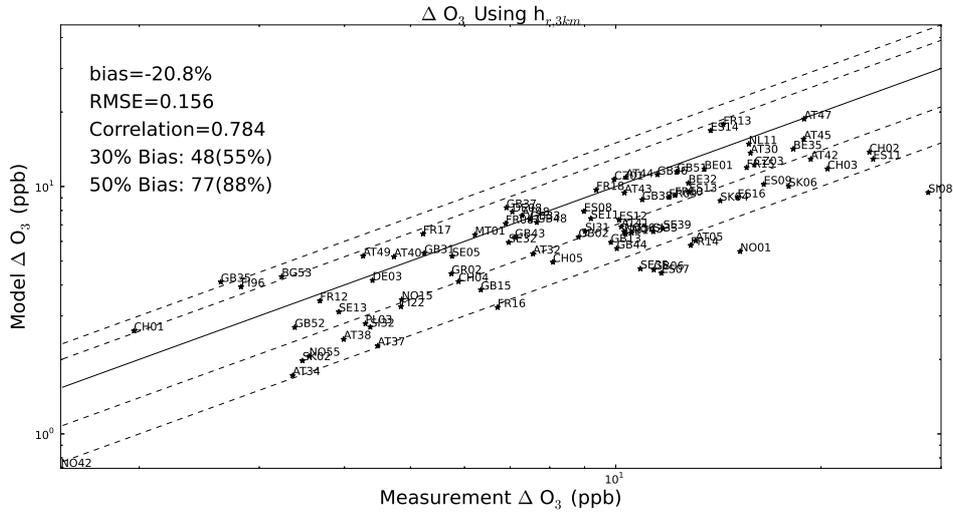


Figure 4.2: Scatterplot with the $h_{r,3}$ extraction method.

The error analysis of the higher stations for the two best extraction methods are shown in figure 4.3. Here stations with an higher relative altitude than 200 m are plotted and several of the highly elevated stations do get improved.

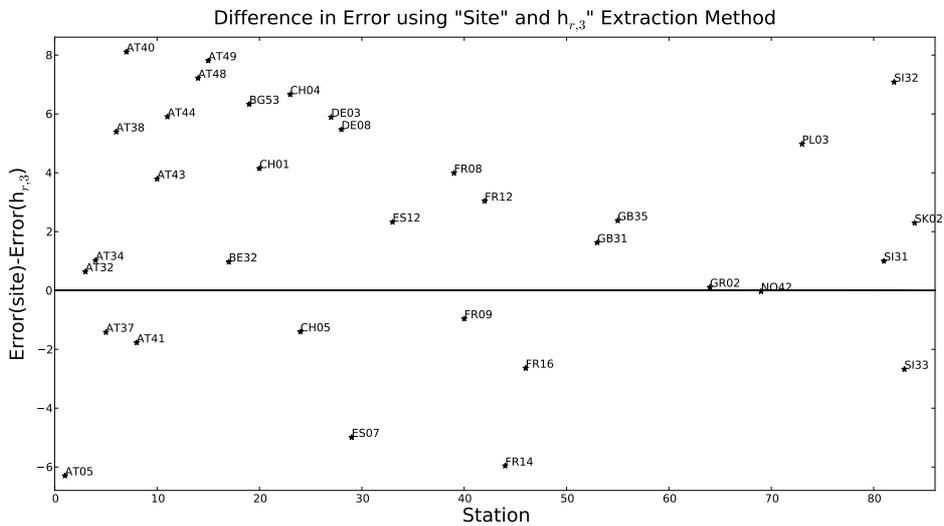


Figure 4.3: Comparison between the measurement difference of the extraction method 'site' and $h_{r,3}$.

For the summer months June, July and August the RMSE is calculated for site extraction, $h_{r,5}$ and $h_{r,10}$ to be 0.164, 0.157 and 0.159. For the winter months December, January and February the results are 0.198, 0.177 and 0.175. The best method, looking at RMSE, for summer was $h_{r,3}$ and $h_{r,5}$ with RMSE=0.157. The method for winter with lowest RMSE (0.146) and largest amount of stations within 30% bias (57) is $h_{r,3}$.

4.1.1 Ozone maps

The resulting O₃ maps are shown in figure 4.4, 4.5 and 4.6. The maps show higher values on higher altitudes.

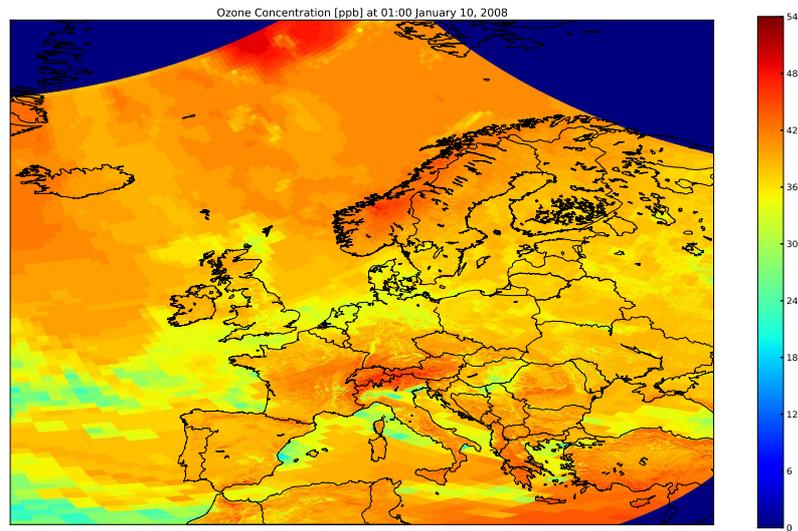


Figure 4.4: O₃ map of a part of the EMEP grid $i:\{0,100\}$, $j:\{0,100\}$ at the 10th of January 2008 at 01.00.

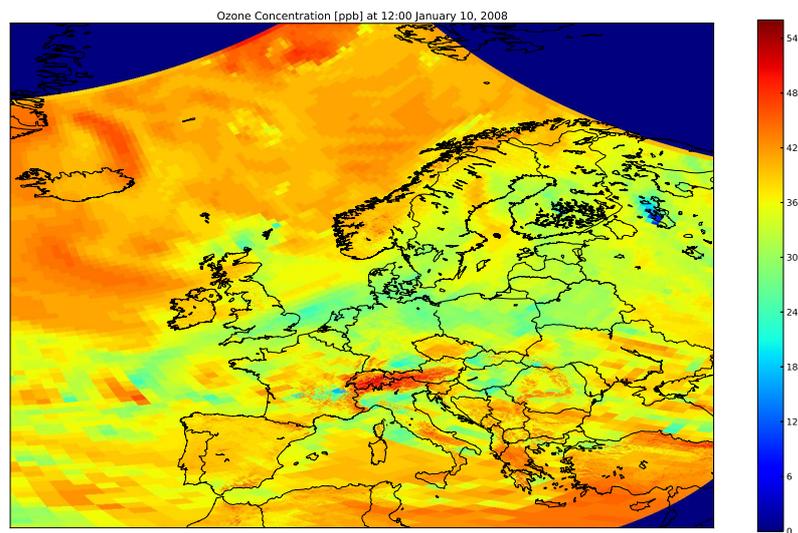
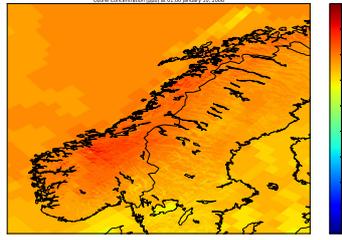
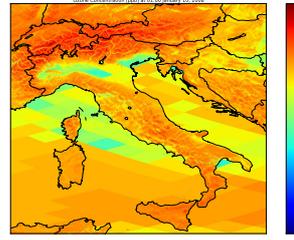


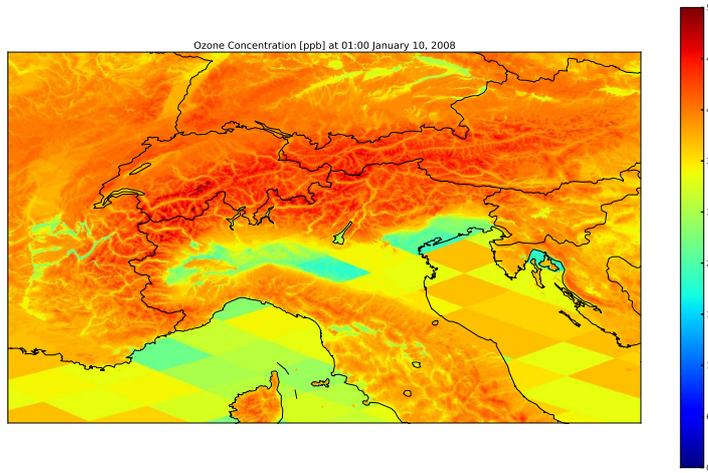
Figure 4.5: O₃ map of a part of the EMEP grid $i:\{0,100\}$, $j:\{0,100\}$ at the 10th of January 2008 at 12.00.



(a) Ozone map of Norway



(b) Ozone map of the Alps and Italy



(c) Ozone map of the Aples region

Figure 4.6: Maps of ozone concentrations.

4.2 NO₂

The results for each method applied on NO₂ shows improvements for all statistical parameters. The methods showing lowest RMSE are $h_{r,5}$ and $h_{r,10}$ with a normalised RMSE of 0.232 closely followed by $h_{r,3}$ with RMSE=0.233. The methods having largest amount of stations within 50 and 30% bias are $h_{r,10}$ (10 and 18), $h_{s,E}$ (10 and 17) and $h_{r,5}$ (9 and 17). See table 4.2 for all the statistical parameters of the different extraction methods.

Table 4.2: The statistical parameters of the extraction methods for NO₂.

	bias	RMSE	Correlation	30% Bias	50% Bias
h_s	-15.9%	0.238	0.550	10(40%)	15(60%)
$h_{s,E}$	18.1%	0.238	0.574	10(40%)	17(68%)
$h_{r,3}$	25.0%	0.230	0.572	8(32%)	14(56%)
$h_{r,5}$	13.8%	0.229	0.574	9(36%)	17(68%)
$h_{r,10}$	6.6%	0.229	0.572	10(40%)	18(72%)

The station CH01, positioned at 3578 m above sea level, is the largest outlier when using the site extraction method. It is also the most elevated station and is therefore an indicator of how good the method is at taken topographic effects into account. The two methods with which CH01

gets within a 30% bias are $h_{s,E}$ and $h_{r,5}$.

The best method for NO_2 is chosen to be $h_{r,5}$. It shows best results for RMSE (together with $h_{r,10}$), good correlation, a large percentage of stations within 30 and 50% bias and very good results for CH01. The results are shown in figure 4.8. Figure 4.7 shows the results using site extraction method. All 9 scatterplots are presented in figure A.4 and A.5 in appendix A.

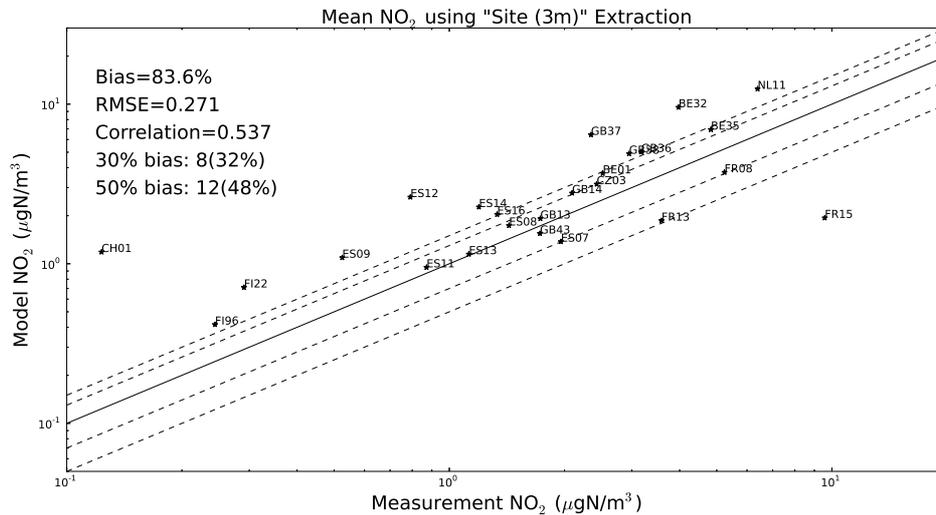


Figure 4.7: Scatterplot with the 'site' extraction method.

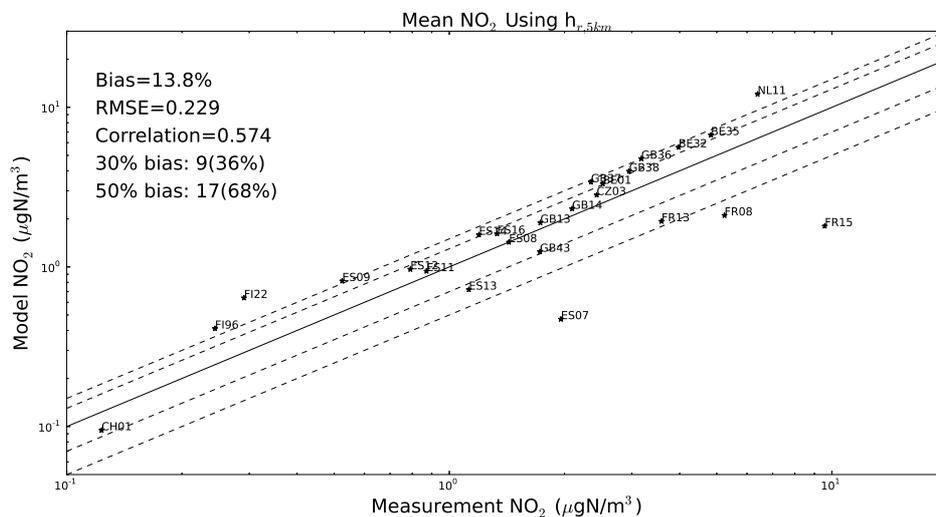


Figure 4.8: Scatterplot with the 'relative altitude 5 km' extraction method.

4.3 SO_2

The results for SO_2 shows worsening for all parameters except the total bias where there are improvements for all methods. The best results for RMSE is obtained when the change from site extraction is small, for $h_{r,3}$ with $\text{RMSE}=0.226$. For site extraction the RMSE was 0.219. The

lowest total bias is obtained with $h_{r,5}$ (-0.8%) closely followed by $h_{r,3}$ and $h_{r,10}$ (2.6% and -4.2%). $h_{s,E}$ gives largest amount of stations within 50 and 30% bias (13 and 9). See table 4.3 for all the statistical parameters of the different extraction methods.

Table 4.3: The statistical parameters of the extraction methods for SO_2 .

	bias	RMSE	Correlation	30% Bias	50% Bias
h_s	-26.2%	0.315	0.515	5(23%)	6(28%)
$h_{s,E}$	-19.2%	0.258	0.651	9(42%)	13(62%)
$h_{r,3}$	-10.4%	0.226	0.739	6(28%)	9(42%)
$h_{r,5}$	-13.5%	0.259	0.678	6(28%)	11(52%)
$h_{r,10}$	-16.2%	0.277	0.630	6(28%)	11(52%)

A station that is largely affected by extraction method change is SK02 which is positioned at an altitude of 2008 m above sea level. The result for this station is improved for each method.

The method showing best results for RMSE is $h_{r,3}$. The method showing best results for SK02 as well as number of stations within 30% and 50% is $h_{s,E}$. The scatterplots for site extraction, $h_{r,3}$ and $h_{s,E}$ are shown in figure 4.10, 4.9 and figure 4.11. The difference in error between site extraction and $h_{s,E}$ is shown in figure 4.12 in order to see which stations that are affected by the method change.

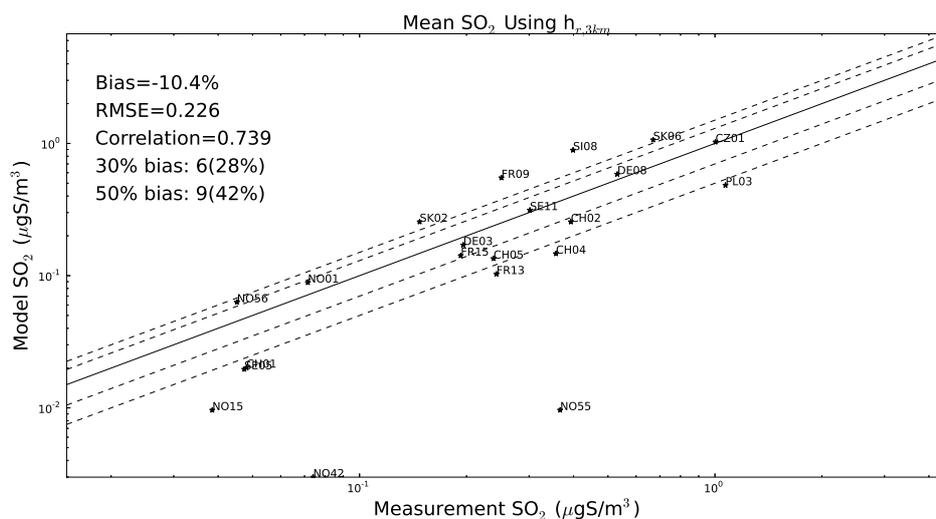


Figure 4.9: Scatterplot with the 'site' extraction method.

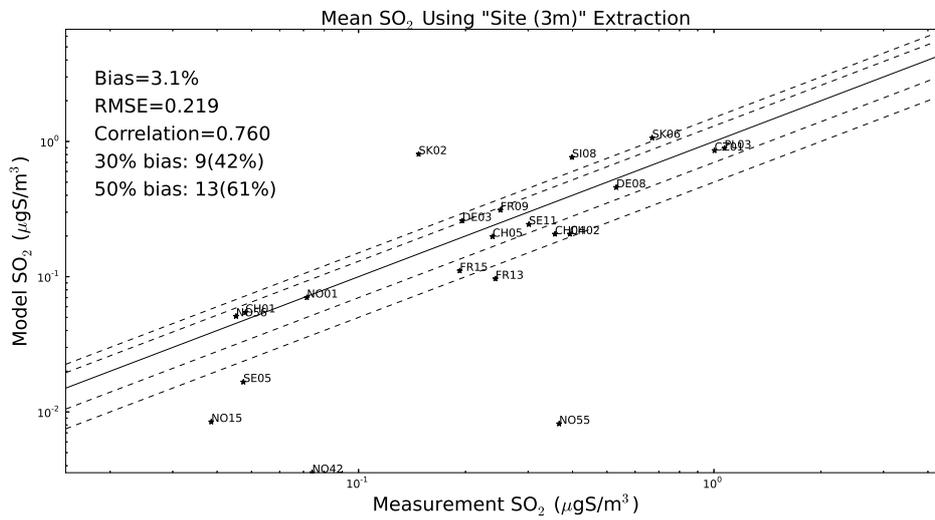


Figure 4.10: Scatterplot with the 'site' extraction method.

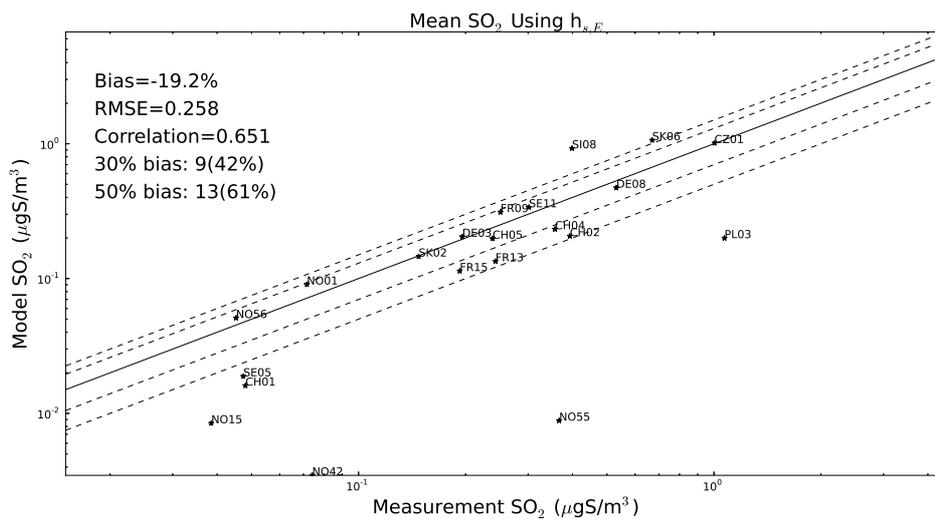


Figure 4.11: Scatterplot with the $h_{s,E}$ extraction method.

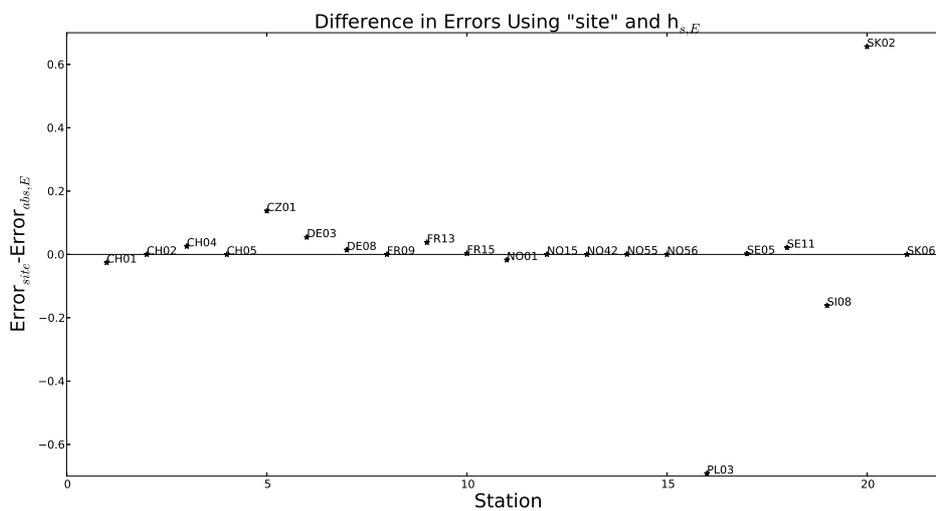


Figure 4.12: The difference between the errors when using 'site' extraction method and 'absolute altitude with EMEP topography'.

4.4 SO₄

The different extraction methods shows almost no improvements for SO₄. All the methods shows the same behaviour, there are very little differences between the methods. The site extraction shows best results in all the statistical variables except the h_{s,E} extraction method

Here the RMSE was 0.172 instead of 0.176 as in the site method, see figure 4.13, 4.14 and table 4.4 for for all the statistical parameters of the different extraction methods.

Table 4.4: The statistical parameters of the extraction methods for SO₄.

	bias	RMSE	Correlation	30% Bias	50% Bias
h _s	-55.9%	0.203	0.921	1(10%)	2(20%)
h _{r,E}	-52.5%	0.177	0.966	2(20%)	3(30%)
h _{s,3}	-52.6%	0.179	0.959	2(20%)	3(30%)
h _{r,5}	-52.7%	0.181	0.953	2(20%)	3(30%)
h _{r,10}	-52.6%	0.179	0.959	2(10%)	3(30%)

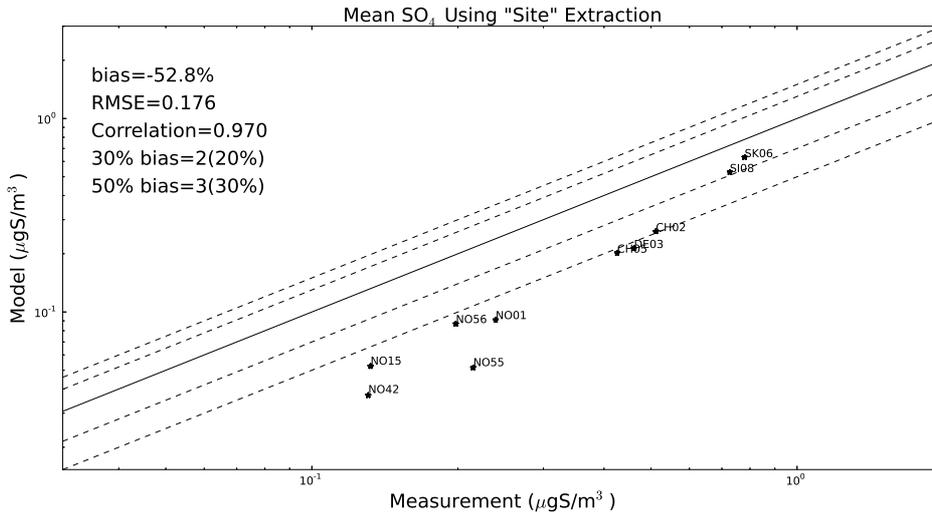


Figure 4.13: Scatterplot with the lowest 'site' extraction method.

4.5 Deposition downscaling

The results regarding the downscaling of land use and temperature and pressure will be presented below.

4.5.1 Land use

The amount of water makes a big difference in canopy resistance. For water SAI= LAI = 0 which means that the only part remaining in equation (2.8) is $1/R_{gs}$ which for temperatures above 0°C is equal to $1/\hat{R}_{gs} = 1/2000$ (for water and ice, see table A.6) which gives a high surface resistance and a low deposition velocity.

The difference in grass and forest are plotted against altitude for the chosen stations in figure 4.15. For most of the stations, the amount of forest is larger for the EMEP land use than for the

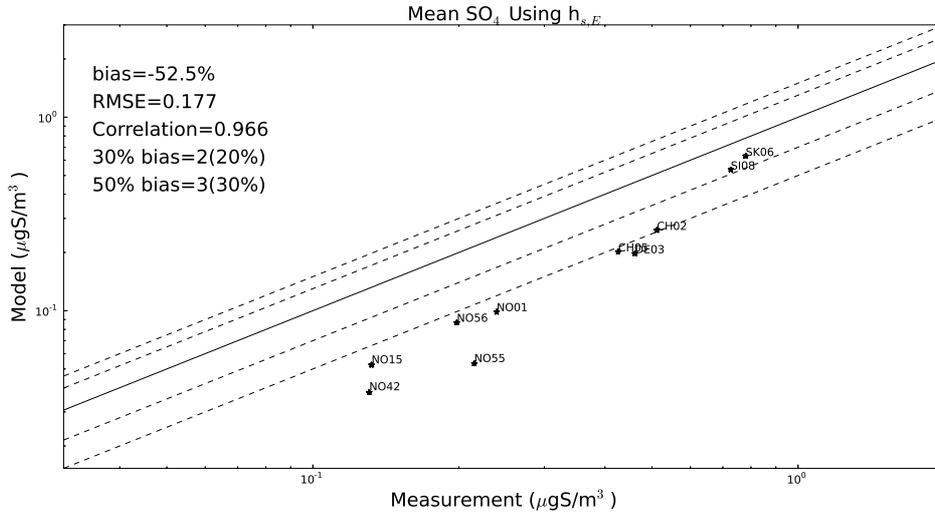


Figure 4.14: Scatterplot with the and $h_{s,E}$ extraction method.

1x1 km data set with radius 5 km. The difference is calculated by subtracting the percentage of EMEP land use with the percentage of that land class for the high resolution dataset.

$$\text{Difference} = \text{Landuse}_{1x1km} - \text{Landuse}_{EMEP} \quad (4.1)$$

The difference in the amount of water is plotted against the error for site extraction and $h_{r,3}$ in figure 4.16. The error is the difference in yearly mean value for measurement and model.

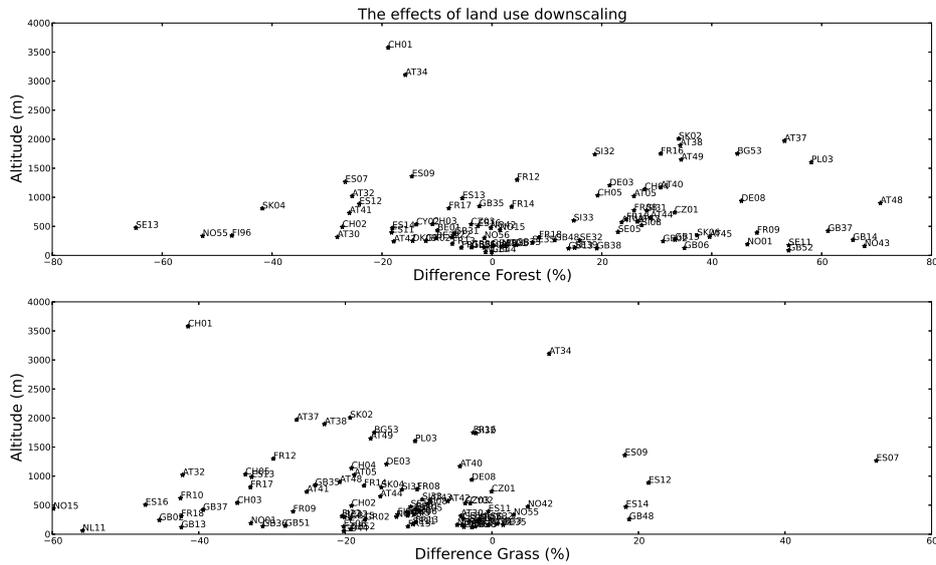


Figure 4.15: The difference in forest and grass land use plotted against absolute altitude.

When the land use changes from grass to forest, the first part of the non-stomatal conductance in (2.12) will increase since SAI is higher for forest areas (approximate 5 and $3 \text{ m}^2 \text{ m}^{-2}$). The second part consists of two resistances, for temperatures above 0°C , $R_{gs}^{O3} = \hat{R}_{gs}^{O3}$ will be approximate 1000

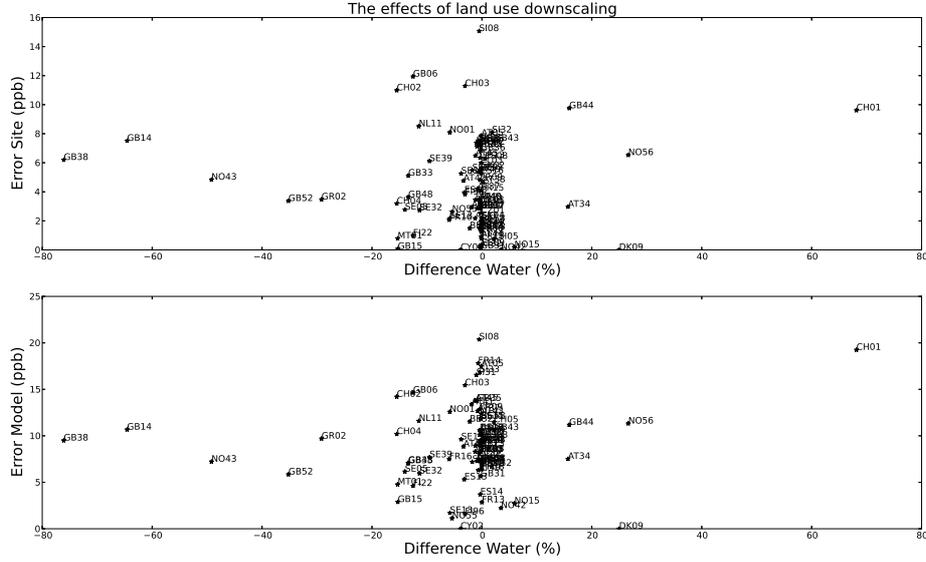


Figure 4.16: The difference in amount of water plotted against the error for site extraction and 'relative altitude 10 km with EMEP topography.

for grass and 200 for forest (see table A.6). R_{inc} will be higher for forest due to a higher vegetation height and SAI. If assuming a frictional velocity $u_* = 2ms^{-1}$, $h = 17m$ and $SAI = 5$ the in canopy resistance for forest will be $R_{inc} \approx 14 \cdot 5 \cdot 17/2 \approx 600$ and for grass assuming $u_* = 2ms^{-1}$, $h = 2m$ and $SAI = 3$ will give $R_{inc} \approx 14 \cdot 3 \cdot 2/2 \approx 40$. The result for the second term will therefore for forest be approximately $1/(600 + 200)$ and for grass $1/(40 + 1000)$, which leads to a higher conductance for forest,

$$G_{ns,forest} \approx \frac{5}{2500} + \frac{1}{600 + 200} = 0.00325ms^{-1} \quad (4.2)$$

$$G_{ns,grass} \approx \frac{3}{2500} + \frac{1}{40 + 1000} \approx 0.0022ms^{-1}. \quad (4.3)$$

The first part of the canopy resistance in (2.8) is $SAI \cdot g_{sto}$. The total non-stomatal conductance depends on g_{max} which for normal temperatures and pressure is $\approx g_{max}^m/41000$. The factors are there to correct for the time of year, day and for which conditions the stomatal conductance is high or low. If assuming optimal conditions, i.e. all factors are equal to 1, then the first term of the canopy resistance for forest and grass, using values from table A.1, would be

$$SAI_{forest} \cdot g_{sto,forest} \approx 5 \cdot 160/41000 \approx 0.02ms^{-1} \quad (4.4)$$

$$SAI_{grass} \cdot g_{sto,grass} \approx 3 \cdot 250/41000 \approx 0.018ms^{-1} \quad (4.5)$$

The total canopy conductance will for forest and grass be 0.02325 and $0.0202 ms^{-1}$, assuming optimal conditions with approximate values for LAI, R_{inc} and R_{gs} .

4.5.2 Temperature and Pressure

The results from the temperature and pressure variation is shown in figure 4.17 for grassland and figure 4.18 for coniferous forest. In figure 4.19(a) and 4.19(b) the temperature and pressure downscaling for the different stations can be seen. The used temperature and pressure model data comes from the dataset corresponding to January 2008. The changing of R_c for water is not

presented in a separate figure, because the resistance will be $1/2000$ for all temperatures above 0° and increase with the exponential factor of R_{low} when going below further down in temperature.

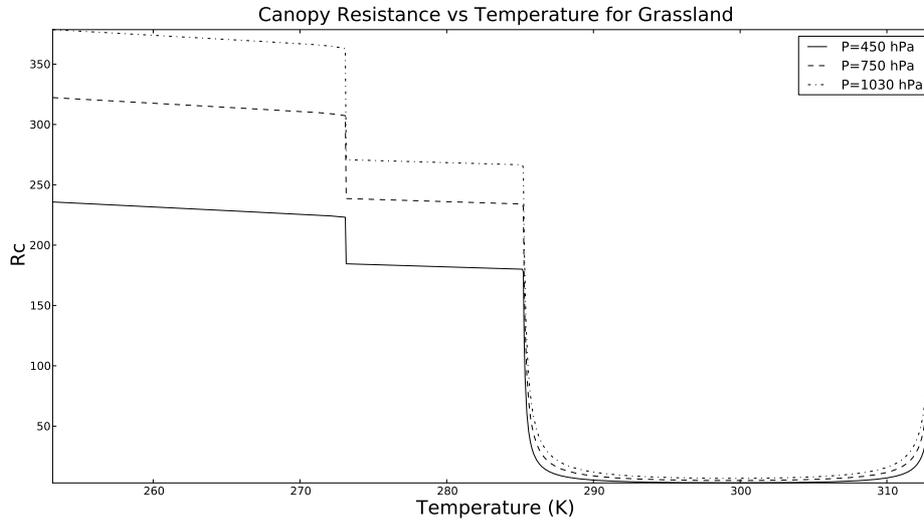


Figure 4.17: The variation of the canopy resistance R_c for O_3 and grassland when temperature and pressure is changing.

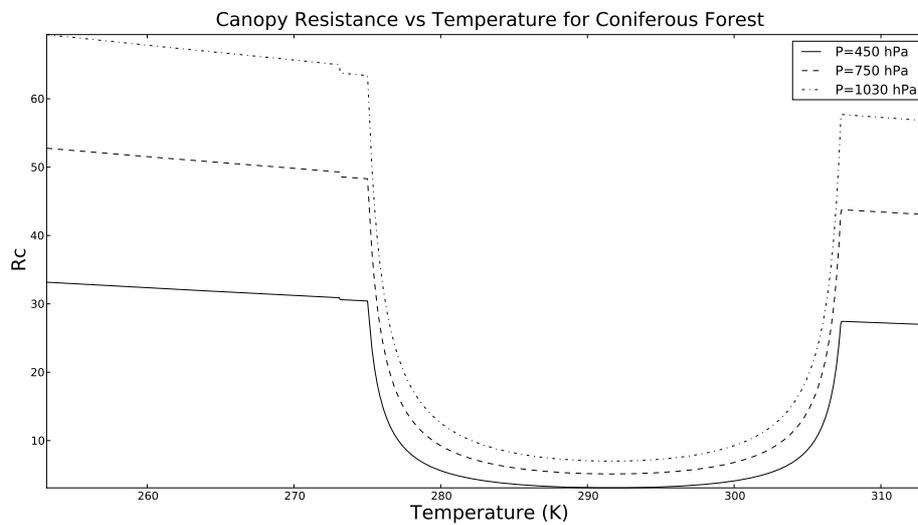
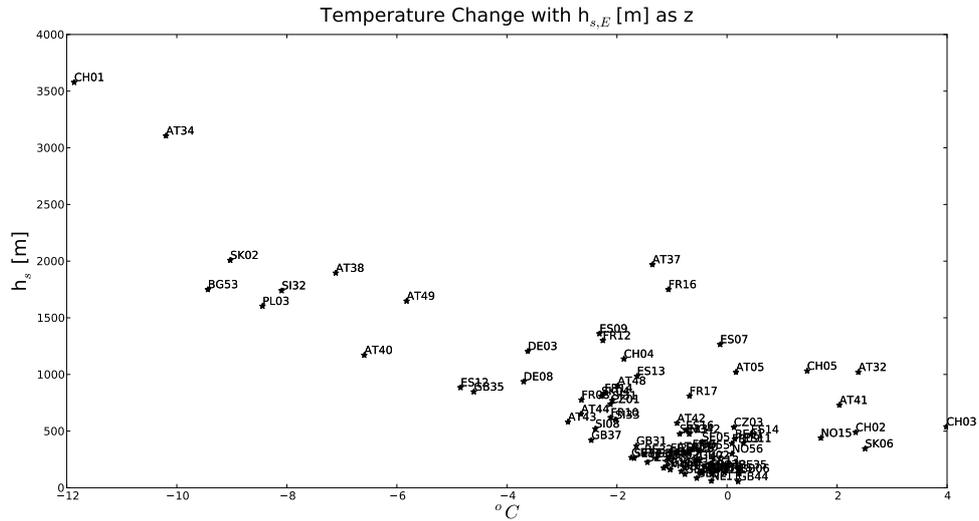
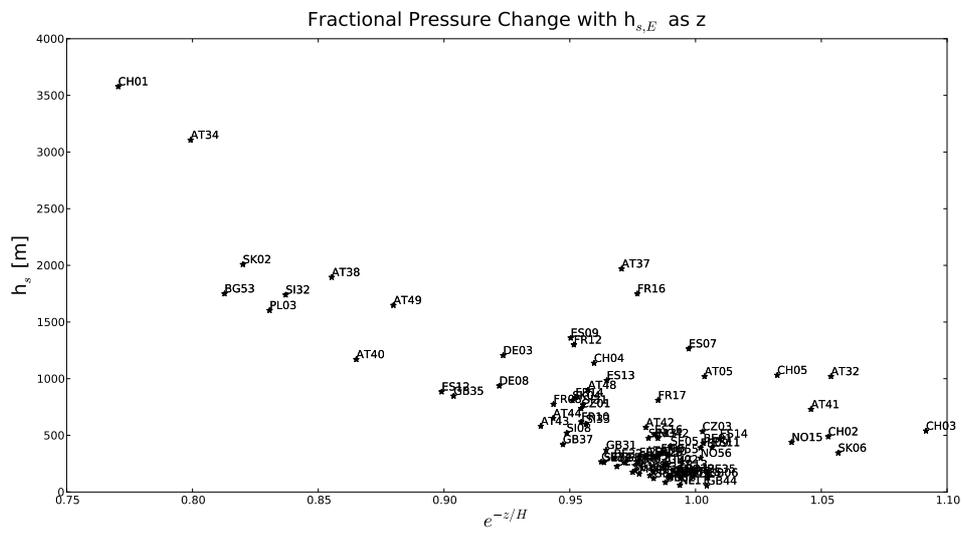


Figure 4.18: The variation of the canopy resistance R_c for O_3 and coniferous forest when temperature and pressure is changing.



(a) The temperature change



(b) The fractional pressure change

Figure 4.19: The change in temperature and pressure after downscaling with a fixed lapse rate of $6.5^{\circ}C\ km^{-1}$ and a scale height of 7km.

5

Discussion

5.1 The results

The sub-grid modelling of O_3 shows interesting results both when it comes to improvements and to new ideas for further development of this post-processing of data. The improvements comes with the higher elevated stations with their diurnal variations. But together with these improvements there are also some worsening results for some of the stations that already had an underestimated diurnal variation. To be able to handle these worsening cases the sub-grid modelling needs to be refined. Suggested changes and developments of the method can be seen further down in this section.

When looking at the stations with an higher relative altitude of 200 m two stations can be seen giving worse results after the downscaling, GB48 and SI08. When investigating the geographical region these stations are situated in, both stations are surrounded with higher elevated ground. This is a problem since the current model can not see that kind of surrounding. Suggested improvements are also given in the discussion section below about the sub-grid model.

The results from looking at the summer and winter average diurnal O_3 values the summer seems to be better coincided with the yearly mean. This is hard to deduce the origin of, but the EMEP MSC-W model might be better adapted for summer conditions than for winter or this extraction methodology is not robust enough.

For NO_2 , the best method were not based solely on best results for RMSE and stations within 30 and 50% bias. Relative altitude 10 km actually showed better results for the bias. But the difference for the stations which was outside these limits was small which was why the results for the high station CH01 was considered more important. Also, a relative altitude based on a search radius of 5 km is motivated by (Loibl et al. 1994) and 3 km by (Klingberg et al. 2012), but no literature suggesting a radius of 10 km have been found.

The station showing largest improvements for SO_2 was SK02. Two stations at high altitudes that shows worse results are CH01 (3578 m) and PL03 (1604 m). SK02, PL03 and CH01 all have a large relative height (1357 m, 1270 m and 2097 m for 5 km). PL03 and DE03 are the main reasons why the results are getting worse when using different extraction methods, when the results for SK02 improves, the result for PL03 is worsening. Other outliers are e.g. NO42 and NO55 which are positioned at high latitudes, $78^{\circ}54'N$ and $69^{\circ}28'N$, where the concentrations are very low. However, when looking at time series, the measurements shows events with large SO_2 concentration which the model do not capture, hence the underestimation in the model.

The stations that shows any difference between the extraction methods are SI08 and SK06, two stations with a low relative altitude and situated in valley like environments, see the site description on the EMEP webpage. The reason for the non existing change among the other stations is due to the small vertical variations of SO_4 when just extracting one or two layers above the site measurement, see figure 2.6(b) for the vertical variation of CH01.

The EMEP MSC-W model seems to have generally too low SO_4 concentrations and therefore this sub-grid model is not able to improve the results. What might be done is to add a downscaled

wet deposition, since this is the major sink for SO_4 , into the sub-grid modelling. But that lies beyond the scope for this thesis. SO_4 and SO_2 both have underestimations in the model and both are strongly dependent on local emissions of SO_2 and other sulphur containing gases that works as precursors for the two pollutants. If the EMEP MSC-W model do not have correct emission data, this might be a valid cause for the underestimation.

5.2 The sub-grid model

When looking at the results for O_3 there is one station, SI08, in particular that has a much larger diurnal variation than the other, even if this station has a absolute altitude of 520 m and a larger relative altitude than 45 m. When investigating the properties of this station it apparently lies in a valley with mountain tops of over 1000 m absolute altitude. This station actually points out a strong weakness in this definition of relative altitude, because it can not distinguish valleys at higher altitudes in a region with fast varying terrain. Because of this sensitivity to fast variations, a more robust method would be to use the mean altitude of a region with some search radius instead of the minimum altitude, as in Klingberg et al. (2012). It is hard to say how these mean relative altitude would look like, but most likely they would be smaller, since the mean altitude of an area is larger than the minimum value. How this mean dependent relative altitude would be implemented should also be considered differently. Because the results indicate a rather large height difference when extracting at new model layer and a smaller relative altitude would not achieve that if one looks at the difference between it and the EMEP topography. But if solely considering the mean dependent relative altitude and extract a model layer corresponding to that, the results might get better.

The reason for testing the relative and absolute altitude with and without the EMEP topography is to see how strong the impact of the EMEP topography have on the model results is. The 'best method' of each of the four pollutants became very different and it is therefore not straight forward to say anything about the importance of the model topography. The difference may either depend on a weak dependence of the model topography or the fact that these four new extraction methods, especially the relative altitude methods, are not robust enough.

The stations that were used had a relative altitude higher than 45 m when the search radius was 10 km. The motivation behind this choice was that for lower situated stations there would be no significant difference when extracting a bit further up. One can also argue that no selection rule had to be used since when using the interpolation there would be a difference to all of the stations that has a positive relative altitude and the statistical significance would also be higher for all the pollutants.

When looking at the results for O_3 one can clearly see that many of the higher stations still lie above the measurement diurnal variation and many lower stations lie below the measurement diurnal variation. This trend maybe could have been avoided if the selection of used station was at an even higher relative altitude. Another idea that could improve the sub-grid modelling is if the search radius was adapted to some frequency analysis of the topography. In other words, if a site is located in region with fast varying terrain the search radius could be smaller and vice versa and to make the methodology more robust one could use the mean altitude in the search region or the difference in maximum and minimum altitude instead of only the minimum value.

This study is done solely on background concentration of the different species and regarded only topographical effects. The map of O_3 might therefore be somewhat misleading, since urban and coastal areas also would have had higher mean concentrations, just as the higher elevated sites. This O_3 map is also overestimated since the model data generally is too high, see section 3.2.1. The origin for this overestimation is hard to deduce, but some important contributions might be the used chemical scheme, emission data and the deposition calculations in the EMEP MSC-W model.

5.3 Downscaling dry deposition

The motivations for scaling down the dry deposition was to see whether this could be used as a further development for this sub-grid modelling. Either the generally high average O_3 values or the large diurnal variations of the highly elevated stations could be adjusted.

One of the topographic effects on land use for the used stations is according to figure 4.15 that grass regions are replaced by forest regions. It is shown that the deposition in forest areas in most cases is larger than in grass areas. If a downscaled land use would have been used for validation, this might lower the mean concentrations leading to a better agreement between measurements and model. However, how big this effect would have been is not possible to say based on this thesis. A hypothesis was that there would be a correlation between the error in mean concentration and the difference in water between EMEP land use and 1x1 km land use dataset. As shown in figure 4.16, the difference in water for the chosen stations was not large. The main reason for that is probably that since we choose stations with a relative altitude above 45 m, there were not many coastal stations. Most of the stations which have a difference more than 20% have a quite large error. GB38, GB14, NO43 and GR02 are coastal sites. If a downscaling of land use would be done, it should be a good idea to correct for coastal stations. Coastal stations have a lower diurnal variation and therefore higher mean values. If this would have been taken into account, there should be a better correlation between measurement and model for such stations.

The station CH01 has biggest change in temperature and pressure. When comparing these results with the downscaling of land use for this station which shows a water (and ice) increase of more than 60% a drastic change in the canopy resistance can be expected since water has a very large R_c through the non-stomatal conductance G_{ns} . A larger R_c implies a smaller deposition velocity which in principle could compensate for the large diurnal variation of O_3 .

Another interesting station is AT34 which shows a temperature decrease of about 10°C and a pressure decrease of about 20%. When comparing these results with the land use downscaling, the forest decreases with nearly 20% , grassland increases with about 10% and the water increases with around 15%. This implies that the the larger R_c for these temperature and pressure changes for water and grassland and the smaller amount of the smaller forest R_c would give a larger total R_c and therefore the deposition velocity. This could also as in the case for CH01 be used as a correction to the large diurnal variation in the EMEP model.

6

Conclusion

This sub-grid modelling showed interesting results, both in terms of some improvements and further development. As a central question, whether the EMEP topography plays any part of the modelled data, O_3 , SO_2 and SO_4 showed that it seems to be important when extracting the new levels of concentration. But for NO_2 this was not the case. The reason for this is not clear, maybe the statistical significance is too low for all the pollutants except O_3 , which indicates that the results only from O_3 is valid. But are there enough stations to make the O_3 results certain? This difference in results may likely also depend on the robustness of the used relative altitude definition. In regions with fast varying terrain and where there are other higher elevated sites in the search region, this principle fails. Therefore as a next step to improve the methodology might be to implement an adaptive search radius dependent on how fast the terrain varies in in a certain region. One could also try to use the relative altitude definition that was used in the paper of Klingberg et al. (2012), where the average altitude in a search region was used instead of the minimum altitude.

As a further development of this methodology the dry deposition could be downscaled. The downscaling of dry deposition of O_3 in terms of the land use shows that much of the grass regions gets replaced by forests which could lead to a higher dry deposition rate and therefore a lower mean value.

The downscaling of the temperature and pressure correlated with the land use shows that some higher elevated stations might get a smaller diurnal variations, since the water (ice) and grassland amount increases and forest decreases. This could lead to a smaller deposition velocity and therefore smaller variations during the day.

Bibliography

- Michael Eisinger and Alfred Wegener. Tropospheric sulfur dioxide observed by the ers-2 gome instrument. *Geophysical Research Letters*, 25:4177–4180, 1998. URL http://ife.physik.uni-bremen.de/doas/paper/grl_98_eisinger.pdf.
- L.D. Emberson, D. Simpson, J.-P. Tuovinen, M.R. Ashmore, and H.M. Cambridge. Towards a model of ozone deposition and stomatal uptake over europe. Research note no. 42, The Norwegian Meteorological Institute, 2000.
- B. Finlayson-Pitts and J. Jr. Pitts. *Chemistry of the Upper and Lower Atmosphere*. London: Academic Press, 2000.
- ORNAL DAAC (Distributed Active Archive Center for Biogeochemical Dynamics). URL http://webmap.ornl.gov/wcsdown/wcsdown.jsp?dg_id=10003_1.
- D. Fowler, M. R. Ashmore, R. A. Cox, P. A. Clark, R. G. Derwent, R. M. Harrison, C. N. Hewitt, Ø. Hov, Irwin J. G., D.P.H. Laxen, Maynard R.L., S.A. Penkett, Weston K.J., Williams M.L., and Woods P.J. Ozone in the united kingdom 1993 - third porg report. Technical report, 1993.
- J. Klingberg, P. E. Karlsson, G. Pihl Karlsson, Y. Hu, D. Chen, and H. Pleijel. Variation in ozone exposure in the landscape of southern Sweden with consideration of topography and coastal climate. *Atmospheric Environment*, 47:252–260, FEB 2012. ISSN 1352-2310. doi: {10.1016/j.atmosenv.2011.11.004}.
- J. Liang and M.Z. Jacobson. A study of sulphur dioxide oxidation pathways over a range of liquid water contents, values and temperature. 2009.
- W. Loibl, W. Winiwarter, A. Kopsca, J. Zueger, and R. Baumann. Estimating the Spatial-Distribution of Ozone Concentrations in Complex Terrain. *Atmospheric Environment*, 28(16): 2557–2566, SEP 1994. ISSN 1352-2310. doi: {10.1016/1352-2310(94)90430-8}.
- J. Marshall and A.R. Plumb. *Atmosphere, Ocean and Climate Dynamics - An Introductory Text*, page 24. Elsevier Academic Press, 2008.
- Meteorological Synthesising Centre West MSC-W. *User's Guide, The EMEP MSC-W Model*, 2011.
- M. Ménégoz, D. Salas y Melia, M. Legrand, H. Teyssédre, M. Michou, V.-H. Peuch, M. Martet, B. Josse, and I. Dombrowski-Etchevers. Equilibrium of sinks and sources of sulphate over europe: comparison between a six-year simulation and emep observations. *Atmospheric Chemistry and Physics*, 9:4505–4519, 2009.
- D. Simpson, J.-P. Tuovinen, L.D. Emberson, and M.R. Ashmore. Characteristics of an ozone deposition module ii: Sensitivity analysis. *Water, Air and Soil Pollution*, 143:123–137, 2003. doi: 10.1023/A:1022890603066. URL <http://www.springerlink.com/content/t0h185n143u81400/fulltext.pdf>.

- D. Simpson, A. Benedictow, H. Berge, R. Bergström, L.D. Emberson, H. Fagerli, G.D Hayman, M. Gauss, J.E. Jonson, M.E. Jenkin, C. Richter A. Nyíri, V.S. Semeena, S. Tsyro, J.-P. Tuovinen, Á. Valdebenito, and P. Wind. The emep msc-w chemical transport model - part 1: Model description. *Atmospheric Chemistry and Physics Discussions*, 12:3781–3874, 2012. doi: 10.5194/acpd-12-3781-2012.
- M. Speidel, R. Nau, F. Arnold, H. Schlager, and A. Stohl. Sulfur dioxide measurements in the lower, middle and upper troposphere: Deployment of an aircraft-based chemical ionization mass spectrometer with permanent in-sight calibration. *Atmospheric Environment*, 41:2427–2437, 2006. doi: 10.1016/j.atmosenv.2006.07.047. URL <http://zardoz.nilu.no/~andreas/publications/114.pdf>.
- K. Torseth, W. Aas, K. Breivik, A.M. Fjåraa, M. Fiebig, A.G. Hjellebreke, C. Lund Myhre, S. Solberg, and K.E. Yttri. Introduction to the european monitoring and evaluation programme (emep) and observed atmospheric composition change during 1972-2009. *Atmospheric Chemistry and Physics Discussions*, 12:1733–1820, 2012. doi: 10.5194/acpd-12-1733-2012.
- South Dakota USGS, United States Geological Survey's EROS Data Center in Sioux Falls. *GTOPO30 Documentation*. URL <http://www1.gsi.go.jp/geowww/globalmap-gsi/gtopo30/README.html>.
- J.M. Wallace and P.V. Hobbs. *Atmospheric Science - An Introductory Survey*. Canada : Academic press, 2nd edition, 2006.
- World Health Organization Regional Office Europe WHO. Health aspects of air pollution with particulate matter, ozone and nitrogen dioxide. Technical report, World Health Organization Regional Office Europe, 2003.

A

Tables and figures

Table A.1: What the different grid cells approximately corresponds to in meter.

Latitude [°]	Ground distance E/W [m]
30	804
40	712
50	598
60	465
70	318
73	272
78	193

A.1 Stations and land classes

Land Class	g_{max}^m mmol O ₃ m ⁻² PLA s ⁻¹	f_{min}	Phenology			Light α	Temperature			VPD		SWP	
			f_{phen_a}	f_{phen_b}	f_{phen_c}		T_{min} °C	T_{opt} °C	T_{max} °C	VPD_{max} kPa	VPD_{min} kPa	SWP_{max} MPa	SWP_{min} MPa
CF	160	0.1	0.2	130	130	0.0083	1	18	36	0.6	3.3	-0.76	-1.2
DF	134	0.13	0.3	50	50	0.006	6	20	34	0.93	3.4	-0.55	-1.3
NF	180	0.13	0.3	110	150	0.013	4	20	37	0.4	1.6	-0.4	-1
BF	200	0.03	0.3	110	150	0.009	4	20	37	1.8	2.8	-1.1	-2.8
TC	300	0.01	0.1	0	45	0.009	12	26	40	0.9	2.8	-0.3	-1.1
MC	156	0.019	0.1	0	45	0.0048	0	25	51	1.0	2.5	-0.11	-0.8
RC	360	0.02	0.2	20	45	0.0023	8	24	50	0.31	2.7	-0.44	-1.0
SNL	60	0.01	0.1	0	45	0.009	1	18	36	88.8	99.9	na	na
GR	270	0.01	1.0	0	0	0.009	12	26	40	1.3	3.0	-0.49	-1.5
MS	213	0.014	0.2	130	130	0.012	4	20	37	1.3	3.0	-1.1	-3.1

Notes: For definition of land-classes, see Table 5.1. For sources of data see Emberson et al. (2000b), except: DF: New values from UNECE (2003). Note that g_{age} , g_{light} and g_{temp} from Emberson et al. (2000b) have been renamed to f_{phen} , f_{light} and f_{temp} here, in order to distinguish these factors from true conductances. Conversion of g_{max}^m as given here to units of m s⁻¹ is done using equation [8.14]. PLA = projected leaf area (m²).

Figure A.1: Image of the table showing the different land class parameters for stomatal conductance of O₃ from the Simpson et al. (2012).

Table A.2: The station list and their altitude information given in [m] for absolute height and relative height for the different search radius.

Station	Absolute h	Relative h(3km)	Relative h(5km)	Relative h(10km)
AT05	1020.0	345.0	349.0	431.0
AT30	315.0	101.0	101.0	112.0
AT32	1020.0	450.0	450.0	493.0
AT34	3106.0	1592.0	1884.0	2174.0
AT37	1970.0	1310.0	1369.0	1398.0
AT38	1895.0	1422.0	1422.0	1433.0
AT40	1170.0	538.0	760.0	861.0
AT41	730.0	345.0	350.0	350.0
AT42	570.0	53.0	79.0	113.0
AT43	581.0	299.0	315.0	386.0
AT44	651.0	281.0	311.0	331.0
AT45	320.0	137.0	137.0	137.0
AT47	240.0	86.0	87.0	131.0
AT48	899.0	501.0	546.0	546.0
AT49	1648.0	577.0	941.0	1013.0
BE01	430.0	55.0	130.0	190.0
BE32	295.0	273.0	273.0	283.0
BE35	160.0	70.0	70.0	81.0
BG53	1750.0	506.0	658.0	946.0
CH01	3578.0	1326.0	2097.0	2833.0
CH02	489.0	47.0	50.0	60.0
CH03	539.0	75.0	99.0	155.0
CH04	1137.0	704.0	708.0	708.0
CH05	1031.0	618.0	618.0	634.0
CZ01	737.0	176.0	232.0	406.0
CZ03	534.0	71.0	112.0	213.0
DE03	1205.0	682.0	850.0	978.0
DE08	937.0	292.0	413.0	540.0
ES07	1265.0	363.0	485.0	619.0
ES08	134.0	134.0	134.0	134.0
ES09	1360.0	107.0	193.0	315.0
ES11	393.0	47.0	58.0	97.0
ES12	885.0	321.0	373.0	473.0
ES13	985.0	180.0	242.0	287.0
ES14	470.0	61.0	143.0	246.0
ES16	506.0	98.0	105.0	110.0
FI22	310.0	158.0	158.0	166.0
FI96	340.0	73.0	80.0	91.0
FR08	775.0	300.0	345.0	455.0
FR09	390.0	257.0	261.0	270.0
FR10	620.0	84.0	113.0	260.0
FR12	1300.0	358.0	660.0	1040.0
FR13	200.0	10.0	50.0	69.0
FR14	836.0	415.0	450.0	496.0
FR15	133.0	40.0	79.0	83.0
FR16	1750.0	250.0	279.0	345.0
FR17	810.0	111.0	178.0	283.0
FR18	309.0	103.0	119.0	194.0

Table A.3: cont. of the station list and their altitude information given in [m] for absolute height and relative height for the different search radius.

Station	Absolute h	Relative h(3km)	Relative h(5km)	Relative h(10km)
GB02	243.0	-4.0	36.0	95.0
GB06	126.0	58.0	58.0	66.0
GB13	119.0	50.0	58.0	76.0
GB14	267.0	115.0	189.0	217.0
GB15	270.0	71.0	80.0	190.0
GB31	370.0	220.0	222.0	239.0
GB33	180.0	37.0	139.0	180.0
GB35	847.0	371.0	630.0	712.0
GB36	137.0	61.0	73.0	82.0
GB37	420.0	173.0	257.0	272.0
GB38	120.0	120.0	120.0	120.0
GB43	160.0	160.0	160.0	160.0
GB44	55.0	55.0	55.0	55.0
GB48	260.0	76.0	111.0	258.0
GB51	145.0	80.0	81.0	85.0
GB52	85.0	85.0	85.0	85.0
GR02	250.0	249.0	250.0	250.0
MT01	160.0	160.0	160.0	160.0
NL11	60.0	60.0	60.0	60.0
NO01	190.0	130.0	150.0	190.0
NO15	439.0	111.0	119.0	409.0
NO42	474.0	474.0	474.0	474.0
NO43	160.0	40.0	120.0	159.0
NO55	333.0	193.0	193.0	193.0
NO56	300.0	125.0	125.0	160.0
PL03	1603.0	896.0	1098.0	1270.0
SE05	404.0	105.0	131.0	136.0
SE11	175.0	87.0	114.0	155.0
SE13	475.0	178.0	181.0	200.0
SE32	261.0	89.0	115.0	144.0
SE35	225.0	42.0	56.0	85.0
SE39	132.0	39.0	52.0	101.0
SI08	520.0	87.0	284.0	304.0
SI31	770.0	393.0	411.0	472.0
SI32	1740.0	1188.0	1392.0	1427.0
SI33	600.0	344.0	382.0	415.0
SK02	2008.0	1102.0	1357.0	1542.0
SK04	808.0	129.0	160.0	220.0
SK06	345.0	-13.0	65.0	92.0

Table A.4: List of country codes for the stations.

Country code	Country
AT	Austria
BE	Belgium
BG	Bulgaria
CH	Schweiz
CZ	Czech Republic
DE	Germany
ES	Spain
FI	Finland
FR	France
GB	Great Britian
GR	Greece
MT	Malta
NL	Netherlands
NO	Norway
PL	Poland
SE	Sweden
SI	Slovenia
SK	Slovakia

Table A.5: Table from Simpson et al. (2012) showing the different land classes and their parameters concerning height h , albedo α , growing season and LAI-parameters.

Code	Landcover	h [m]	α [%]	Growing Season (SGS-EGS)	LAI min	LAI max	L_S	L_E
CF	Temp./boreal coniferous forests	20 [†]	12	All year	3.4	4.5	192	96
DF	Temp./boreal deciduous forests	20 [†]	16	90-270	3.5	5.0	56	92
NF	Mediterranean needleleaf forests	15	12	All year	3.5	3.5	192	96
BF	Mediterranean broadleaf forests	15	16	All year	3.5	3.5	192	96
TC	Temperate Crops	1	20	105-197*	0.0	3.5	70	22
MC	Mediterranean Crops	2	20	105-197*	0.0	3.0	70	44
RC	Root Crops	1	20	130-250	0.0	4.2	35	65
SNL	Seminatural/Moorland	0.5	14	All year	2.0	3.0	192	96
GR	Grassland	0.5	20	All year	2.0	3.5	140	135
MS	Mediterranean scrub	3	20	All year	2.5	2.5	1	1
WE	Wetlands	0.5	14	All year	na	na	na	na
TU	Tundra	0.5	15	All year	na	na	na	na
DE	Desert	0	25	All year	na	na	na	na
W	Water	0	8	All year	na	na	na	na
I	Ice	0	70	All year	na	na	na	na
U	Urban	10	18	All year	na	na	na	na

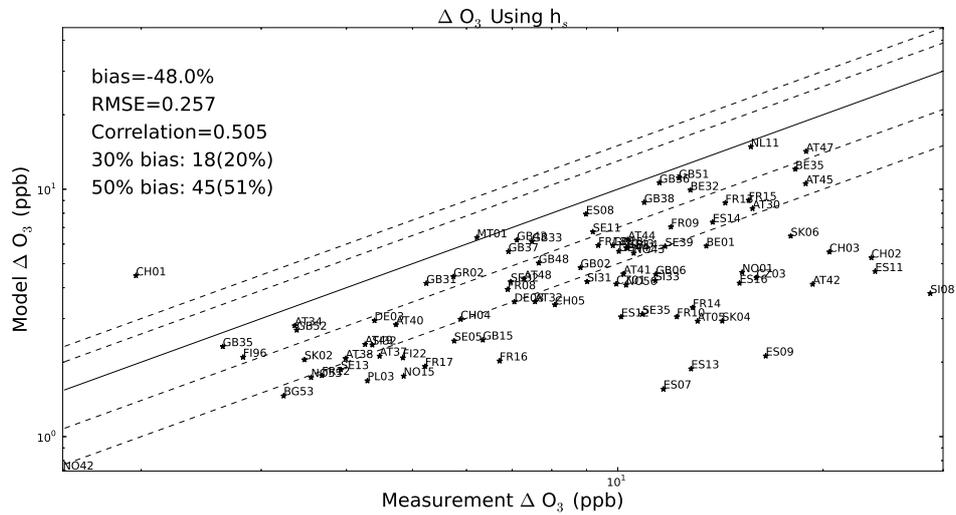
[†] For boreal forests north of 60°N, height is reduced by 5% per degree extra latitude, down to a minimum of 6 m for 74°N and above.

* For these crops growing seasons vary with location. Currently we use a simple latitude-based function, although this will likely be replaced in future. Default values here apply to 50°N. SGS and EGS occur earlier at the rate of 3 days per degree latitude on moving south, or increase on moving north.

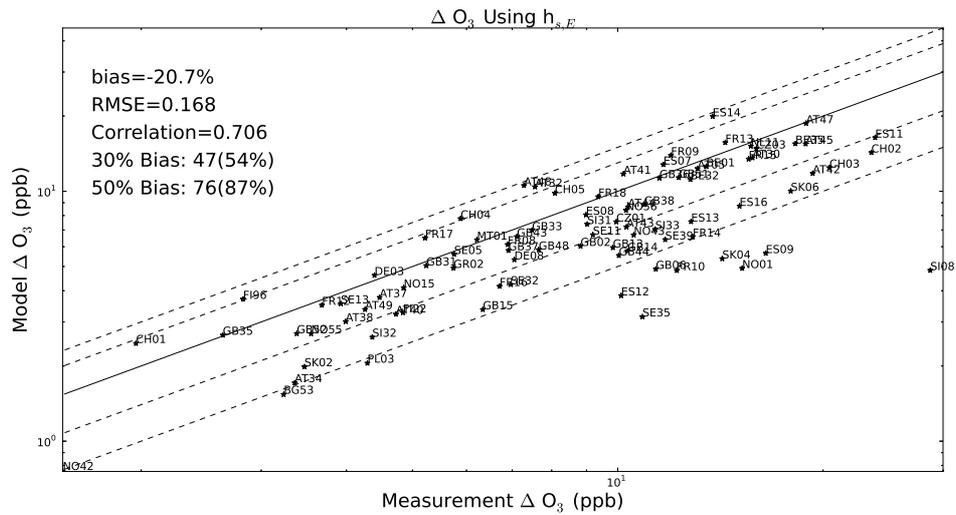
Table A.6: Values for $\hat{R}_{gs}^{O_3}$ from (Simpson et al. 2003)

Land class	$\hat{R}_{gs}^{O_3}$ [m]
Forests, Mediterranean scrub	200
Crops	200
Moorland	400
Grasslands	1000
Wetlands	400
Desert	400
Water	2000
Ice	2000
Urban	400

A.2 Results O₃

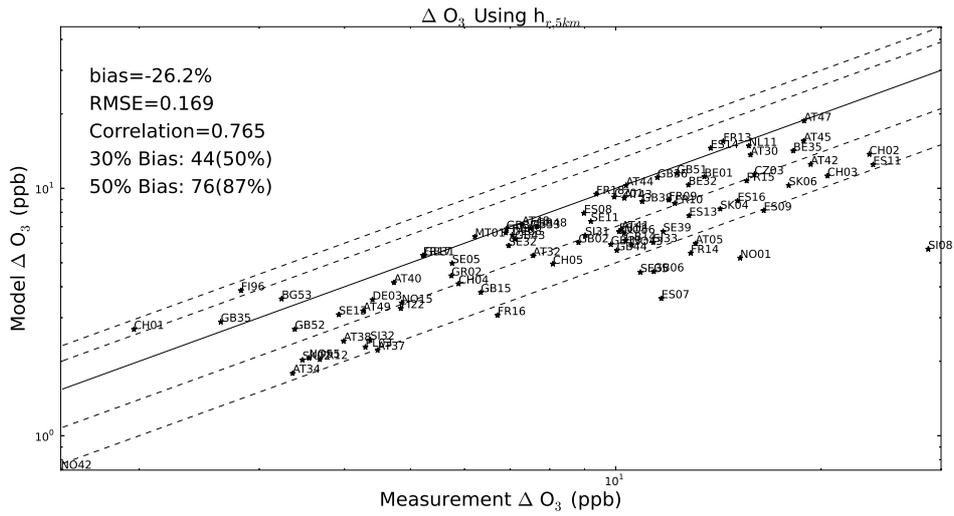


(a) h_s

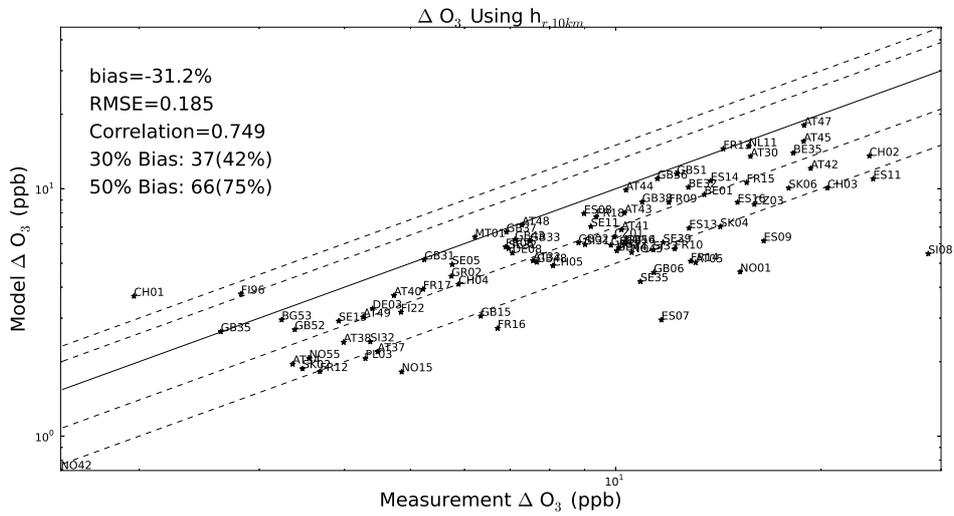


(b) $h_{s,E}$

Figure A.2: Scatterplots of the two methods h_s and $h_{s,E}$ for O₃.



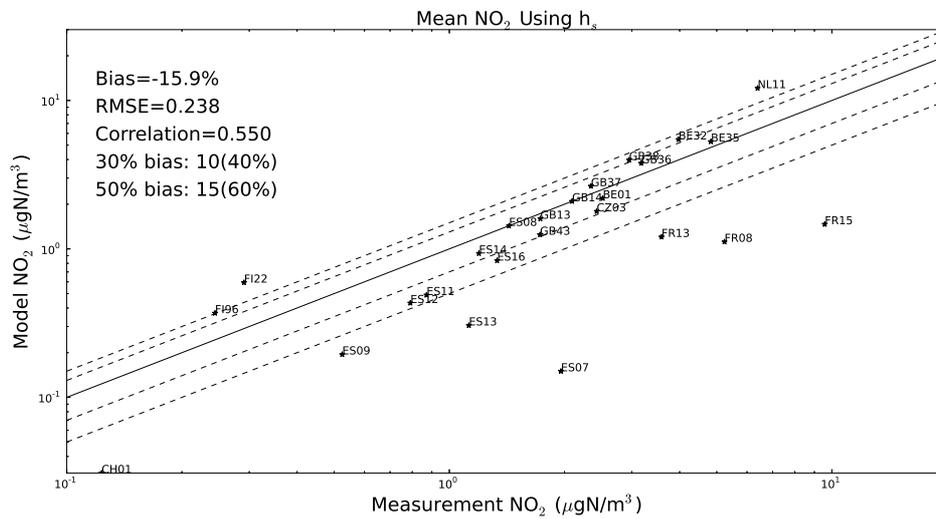
(a) Relative altitude 5 km



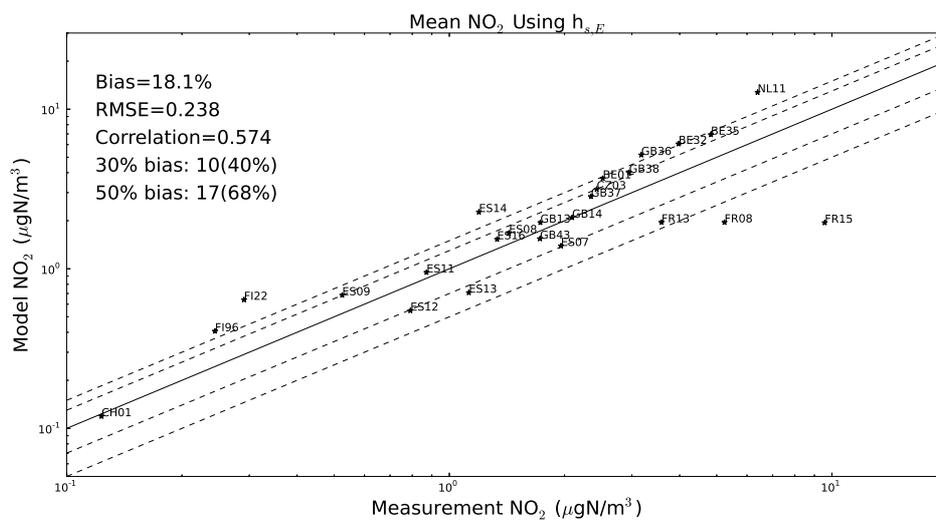
(b) Relative altitude 10 km

Figure A.3: Scatterplots of the two h_r methods with $R=5$ and 10 km for O_3 .

A.3 Results NO₂

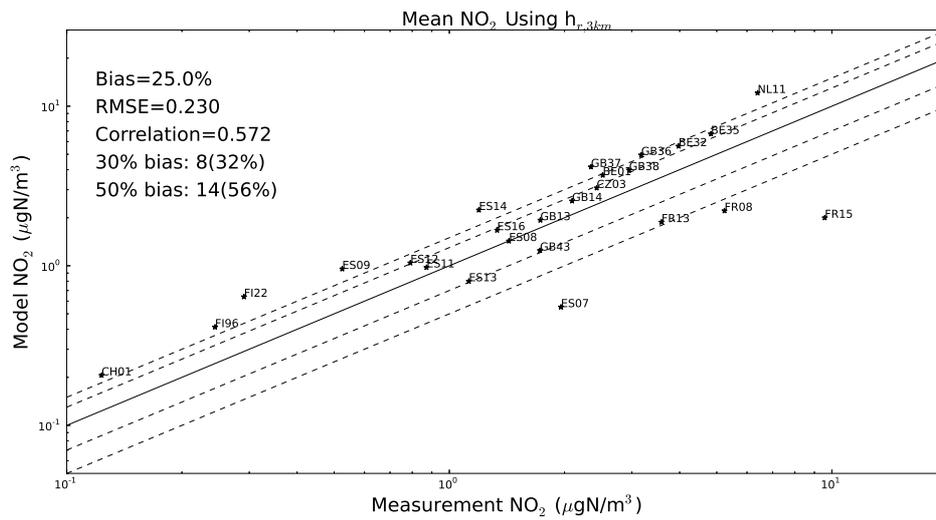


(a) h_s

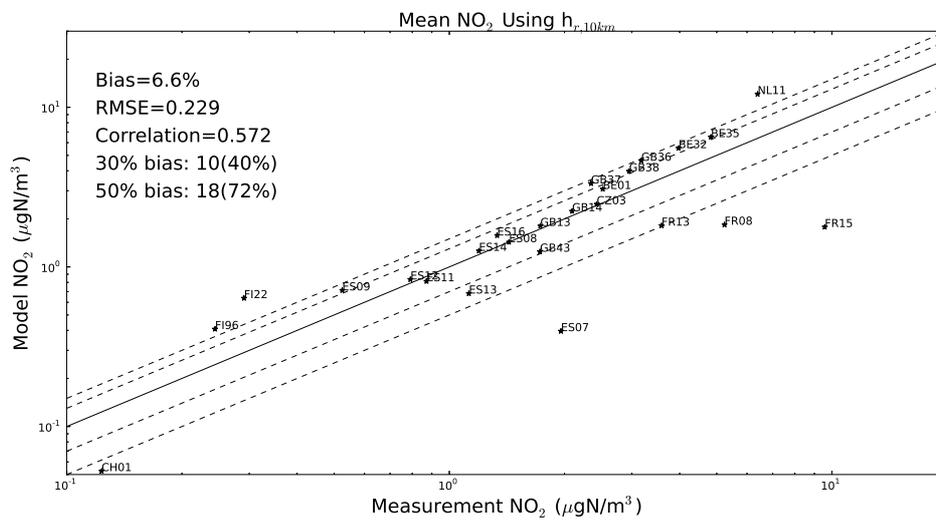


(b) $h_{s,E}$

Figure A.4: Scatterplots of the two methods h_s and $h_{s,E}$ for NO₂.



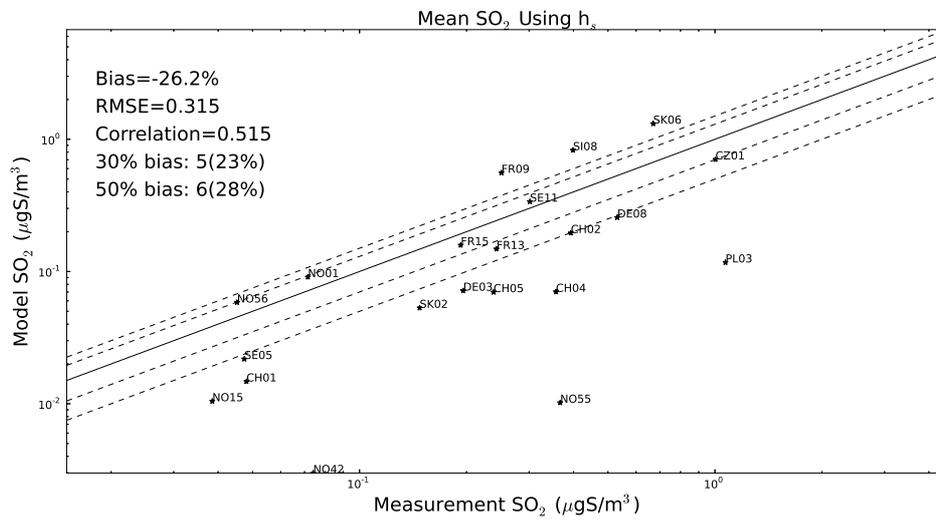
(a) $h_{r,3}$



(b) $h_{r,10}$

Figure A.5: Scatterplots of the two h_r methods with $R=3$ and 10 km for NO_2 .

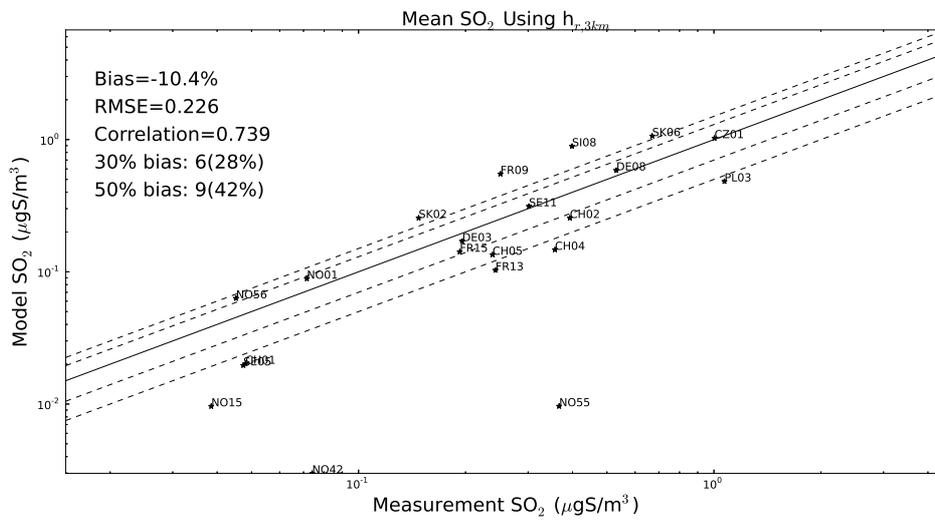
A.4 Results SO₂



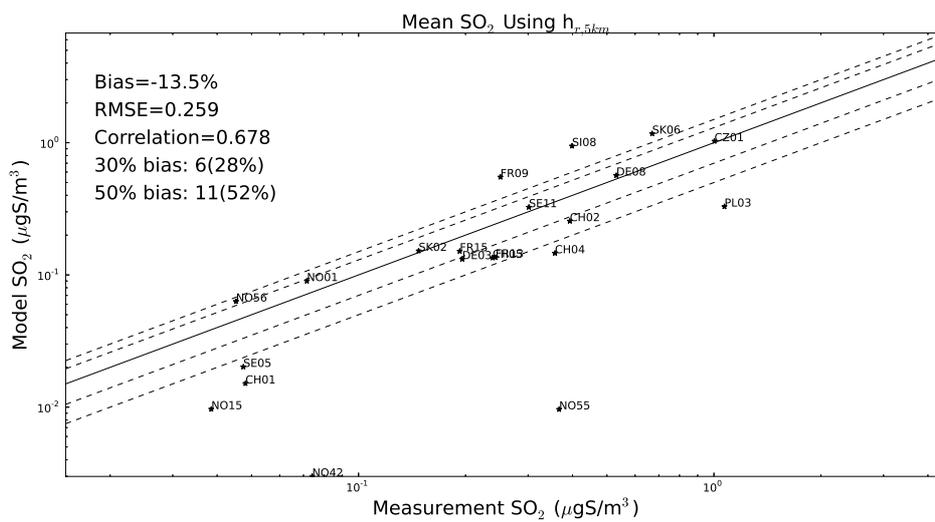
(a) h_s

Figure A.6: Scatterplots of the h_s method for SO₂.

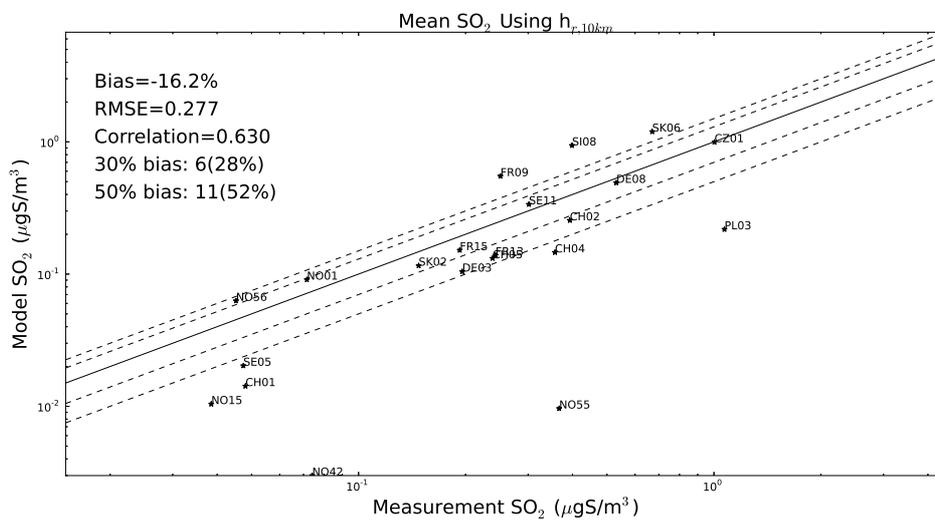
A.5 Results SO₄



(a) $h_{r,3}$



(b) $h_{r,5}$



(c) $h_{r,10}$

Figure A.7: Scatterplots of the three different h_r methods for SO₂.

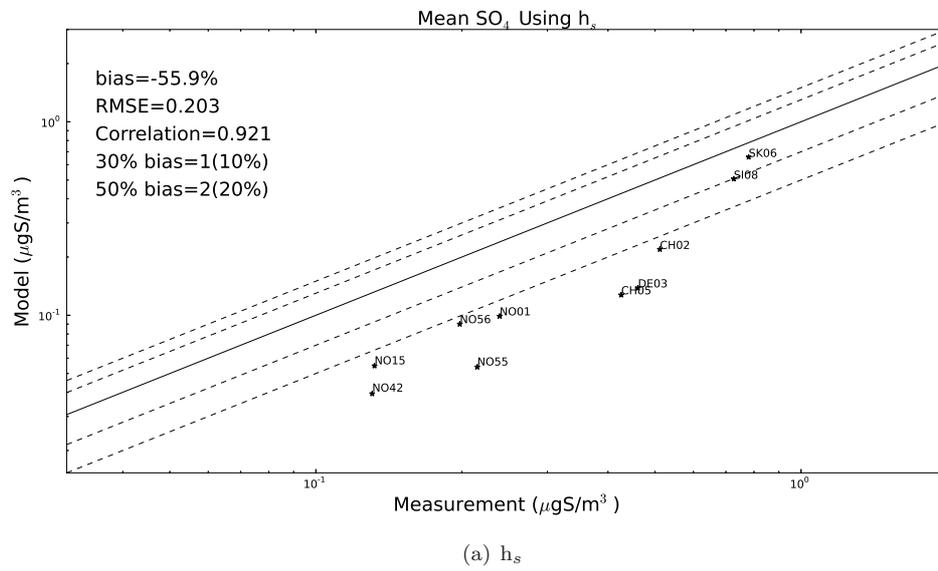
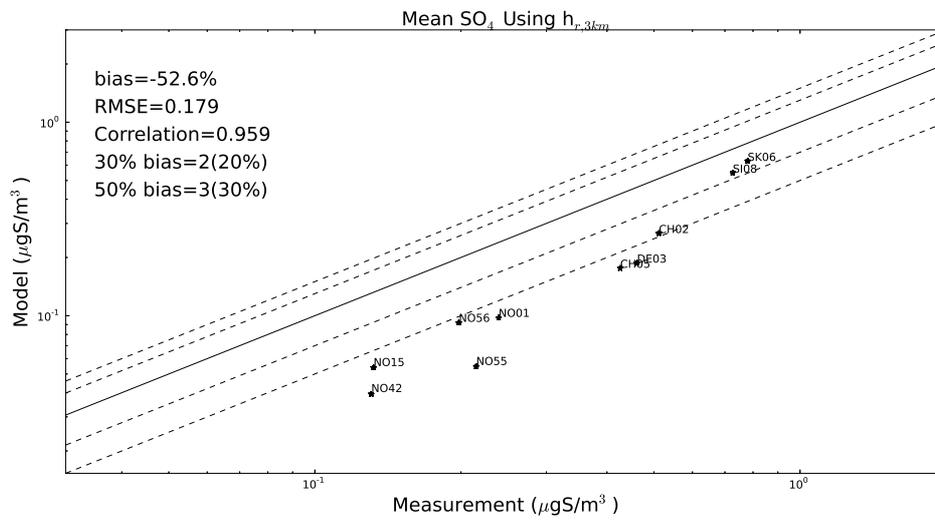
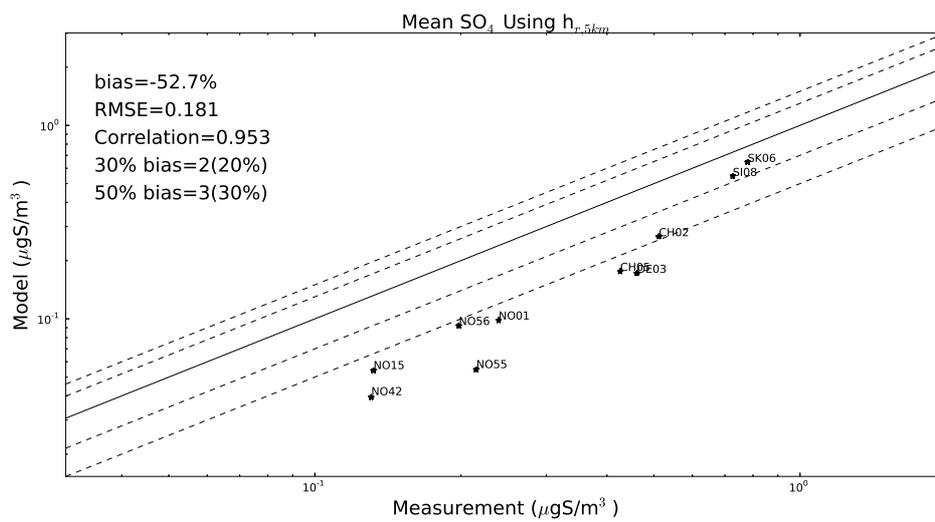


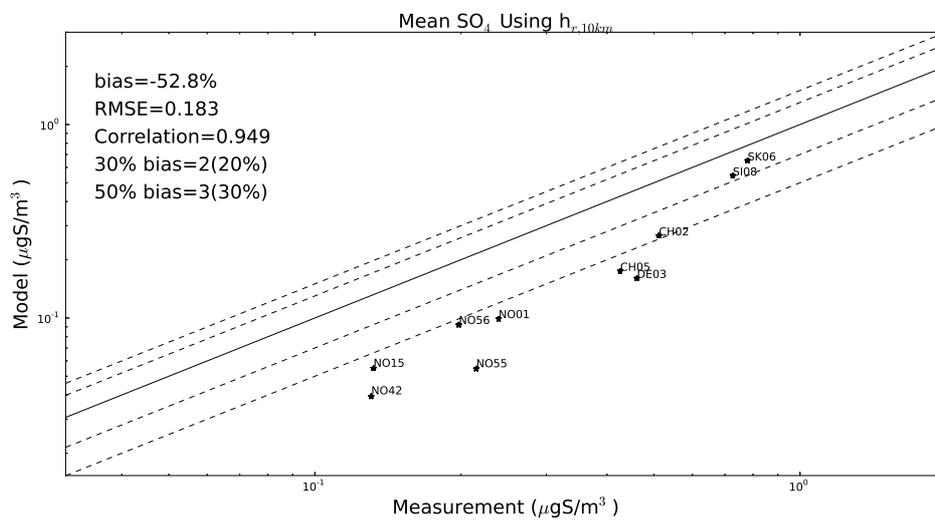
Figure A.8: Scatterplots of the method h_s for SO₄.



(a) $h_{r,3}$



(b) $h_{r,5}$



(c) $h_{r,10}$

Figure A.9: Scatterplots of the three different h_r methods for SO_4 .

B

Source code

The source code for the sub-grid modelling is presented below, but this text is not correct indented. So if someone wants to use this code, please contact the authors.

B.1 main.py

```
# -----  
#  
# This script saves NILU measurements and EMEP model data for selected pollutants.  
# The used stations are specified in sondestat.dat and comes  
# from the EMEP model data output sites and sondes files.  
#  
# -----  
  
# -----  
# ----- Precalculations and predefinitions -----  
# -----  
  
# Defying the species for which the script will run for  
species=['O3','NO2','SO4','SO2']  
species_sonde=2 #0=ozone, 7=SO4, 1=NO2, 2=SO2 controls sondes files  
  
# [0,0]=ozone, [5,2]=SO4, [0,2]=NO2, [5,1]=SO2 controls sites files  
# [i,j]=[line,column] that extracts the right value (DEPENDS ON THE NUMBER OF VARIABLES IN SITES!)  
species_site=[5,1]  
  
# Iterating the months  
month=[0,1,2,3,4,5,6,7,8,9,10,11]  
# Starting day of each month for the year 2008  
# For this data set the dec sondes file only contains up to the 27:th, 3:rd hour  
days=[0,31,60,91,121,152,182,213,244,274,305,335,361]  
  
# What regional uptake the relative height method will use  
hrstr=['3','5','10']  
  
# Getting data for the stations that has both measurement and model data  
execfile('NILUstat.py')  
alt,lat_s,lon_s,sondeindex,ijcoord,statname= NILUstat()  
  
# Relative heights calculations  
execfile('heights.py')  
hr3, hr5, hr10, alt_snm, sondeindex = heights(lat_s,lon_s,ijcoord,sondeindex,alt)
```

```

# The EMEP i and j coordinates for the selected stations
i_s=[]
j_s=[]
for i in range(0,np.size(sondeindex)):
    i_s.append(ijcoord[sondeindex[i],0])
    j_s.append(ijcoord[sondeindex[i],1])

# The measurement data for the selected stations
execfile('NILUdata.py')
if species_sonde==0 or species_sonde==1:
    spec_h, meas_dmean, stats = NILUdata(sondeindex,species_sonde)
if species_sonde==6 or species_sonde==2:
    meas_dmean, stats = NILUdata(sondeindex,species_sonde)
spec_h=0

# Getting the EMEP orography and the corresponding height for all the stations
data_emep=sio.netcdf_file('meteo20090102_cf.nc','r')
emep_height=data_emep.variables['orography']
# An array with emep (j,i) coordinates (! reversed order of (i,j) h_emep(time,j,i)
h_emep=emep_height[:]

# Extracting the EMEP height for the stations within the map, with the (i,j) coordinates
he_s=[]
for jj in range(0,np.size(i_s)):
    he_s.append(float(h_emep[1,j_s[jj],i_s[jj]]))
he_s=2936.350661953+np.asarray(he_s)*0.0921258194197467 # Corrections from the netcdf-file

del data_emep, emep_height, h_emep # To free some memory

# -----
# Function for calculating for extracting EMEP data from different levels.
# -----
def k_level(dh1, dh2, sites_con, sondes_con, sondeindex, site):
    kh=[-2000, 92, 184, 324, 522, 781, 1106, 1503, 1981, 2551, 3230,
        4039, 5013, 6216, 7624, 9043, 10348, 11487, 12702, 14205, 16180]

    kh_mid=[46,138,254,423,652,950,1305,1742,2266,2891,3634,4526,5614,6920,8333]

    k1=[]
    k2=[]
    for m in range(0,np.size(sondeindex)):
        for mm in range(0,np.size(kh)-1):
            if dh1[m]>kh[mm] and dh1[m]<kh[mm+1]:
                k1.append(mm)
    for n in range(0,np.size(sondeindex)):
        for nn in range(0,np.size(kh)-1):
            if dh2[n]>=kh[nn] and dh2[n]<kh[nn+1]:
                k2.append(nn)

    n=np.size(sondeindex) # 89 # Number of stations in sonde file
    model2=np.zeros((np.size(sondeindex),np.size(sondes_con[:,0])/n),dtype=float)
    model21=np.zeros((np.size(sondeindex),np.size(sondes_con[:,0])/n),dtype=float)
    model22=np.zeros((np.size(sondeindex),np.size(sondes_con[:,0])/n),dtype=float)
    model20=np.zeros((np.size(sondeindex),np.size(sondes_con[:,0])/n),dtype=float)
    for l in range(0,np.size(sondes_con[:,0])/n):
        for u in range(0,np.size(sondeindex)):
            #model1[u,l]=sites_con[u,1]

```

```

    model21[u,l]=sondes_con[sondeindex[u]+n*1,k2[u]]
    model22[u,l]=sondes_con[sondeindex[u]+n*1,k2[u]+1]
    if k2[u]==0:
model20[u,l]=sites_con[u,l]
    else:
model20[u,l]=sondes_con[sondeindex[u]+n*1,k2[u]-1]

    # Interpolation between the layers
    if dh2[u]<0:
model2[u,l]=model20[u,l]
    else:
dh=(dh2[u]-kh_mid[k2[u]])
if dh2[u]>kh_mid[k2[u]]:
    # K=dy/dx, ppb/m
    K=model(22[u,l]-model21[u,l])/(kh_mid[k2[u]+1]-kh_mid[k2[u]])

if dh2[u]<kh_mid[k2[u]]:
    if k2[u]>0:
        # K=dy/dx, ppb/m
        K=(model21[u,l]-model20[u,l])/(kh_mid[k2[u]],kh_mid[k2[u]-1])
    else:
        K=(model21[u,l]-model20[u,l])/(kh_mid[k2[u]]-3)
model2[u,l]=model21[u,l]+K*dh

    model1=np.zeros((np.size(sondeindex),np.size(sondes_con[:,0])/n),dtype=float)
    model11=np.zeros((np.size(sondeindex),np.size(sondes_con[:,0])/n),dtype=float)
    model12=np.zeros((np.size(sondeindex),np.size(sondes_con[:,0])/n),dtype=float)
    model10=np.zeros((np.size(sondeindex),np.size(sondes_con[:,0])/n),dtype=float)
    # The sites data for each station
    if site==1:
        model_site=np.zeros((np.size(sondeindex),np.size(sondes_con[:,0])/n),dtype=float)
    for l in range(0,np.size(sondes_con[:,0])/n):
        for u in range(0,np.size(sondeindex)):
            if site==1:
model_site[u,l]=sites_con[u,l]
                model11[u,l]=sondes_con[sondeindex[u]+n*1,k1[u]]
                model12[u,l]=sondes_con[sondeindex[u]+n*1,k1[u]+1]
                if k1[u]==0:
model10[u,l]=sites_con[u,l]
            else:
model10[u,l]=sondes_con[sondeindex[u]+n*1,k1[u]-1]

    # Interpolation between the layers
    if dh1[u]<0:
model11[u,l]=model10[u,l]
    else:
dh_1=(dh1[u]-kh_mid[k1[u]])
if dh1[u]>kh_mid[k1[u]]:
    #K=dy/dx, ppb/m
    K1=(model12[u,l]-model11[u,l])/(kh_mid[k1[u]+1]-kh_mid[k1[u]])
if dh1[u]<kh_mid[k1[u]]:
    if k1[u]>0:
        #K=dy/dx, ppb/m
        K1=(model11[u,l]-model10[u,l])/(kh_mid[k1[u]]-kh_mid[k1[u]-1])
    else:
        K1=(model11[u,l]-model10[u,l])/(kh_mid[k1[u]]-3)
model11[u,l]=model11[u,l]+K1*dh_1

```

```

# # Extracting the exact layer values
#model1=np.zeros((np.size(sondeindex),np.size(sondes_con[:,0])/n),dtype=float)
#model2=np.zeros((np.size(sondeindex),np.size(sondes_con[:,0])/n),dtype=float)

#for l in range(0,np.size(sondes_con[:,0])/n):
#    #for x in range(0,np.size(sondeindex)):
#        #model1[x,l]=sites_con[x,l]
#        #if dh2[x]>45/2:
#model2[x,l]=sondes_con[sondeindex[x]+n*l,k2[x]]
#        #if dh2[x]<45/2:
#model2[x,l]=sites_con[x,l]
#    if site==1:
#        return model1, model2, k1, k2, model_site
#    if site==0:
#        return model1, model2, k1, k2

# -----
# ----- Function for calculating monthly and daily mean-----
# -----

def mean(model1, model2, spec_h, meas_dmean, species_sonde, days, ijk, site, model_site, mon):
    # O3 and NO2 has hourly data which means they can go through this to create the monthly means:
    if species_sonde==0 or species_sonde==1:
        model1_mean=[]
        model2_mean=[]
        err1=[];
        err2=[];
        model1_time_mean=[]
        model2_time_mean=[]
        spec_time_mean=[]
        if site==1:
            modelsite_mean=[]
            modelsite_time_mean=[]
        for x in range(0,np.size(model1[:,0])):
            model1_time=[]
            model2_time=[]
            spec_time=[]
            if site==1:
                modelsite_time=[]
        for t in range(0,np.size(model1[0,:])):
            if spec_h[x,t+days[ijk]*24]>-200:
                model1_time.append(model1[x,t])
                model2_time.append(model2[x,t])
                if site==1:
                    modelsite_time.append(model_site[x,t])
                spec_time.append(spec_h[x,t+days[ijk]*24])
                model1_time_mean.append(np.sum(model1_time)/np.size(model1_time))
                model2_time_mean.append(np.sum(model2_time)/np.size(model2_time))
                if site==1:
                    modelsite_time_mean.append(np.sum(modelsite_time)/np.size(modelsite_time))
                spec_time_mean.append(np.sum(spec_time)/np.size(spec_time))
                err1.append(model1_time_mean[x]-spec_time_mean[x])
                err2.append(model2_time_mean[x]-spec_time_mean[x])

    # S04 and S02, create daily mean for model and measurement data
    if species_sonde==6 or species_sonde==2:
        sulphate_d=meas_dmean.copy()
        err1=[];

```

```

err2=[];
model1_time_mean=np.zeros((np.size(sondeindex)),dtype=float)
model2_time_mean=np.zeros((np.size(sondeindex)),dtype=float)
if site==1:
    modelsite_time_mean=np.zeros((np.size(sondeindex)),dtype=float)
    spec_time_mean=np.zeros((np.size(sondeindex)),dtype=float)
    for xx in range(np.size(model1[:,0])):
        # Creating monthly mean of model data
        model1_time_mean[xx]=np.sum(model1[xx,:])/np.size(model1[xx,:])
        model2_time_mean[xx]=np.sum(model2[xx,:])/np.size(model2[xx,:])
        if site==1:
modelsite_time_mean[xx]=np.sum(model_site[xx,:])/np.size(model_site[xx,:])

        # Creating monthly mean of the measurement data
        sulphate_time=[]
        for t in range(0,days[ijk+1]-days[ijk]):
if sulphate_d[xx,days[ijk]+t]>-200:
    sulphate_time.append(sulphate_d[xx,days[ijk]+t])
        spec_time_mean[xx]=np.sum(sulphate_time)/np.size(sulphate_time)
        err1.append(model1_time_mean[xx]-spec_time_mean[xx])
        err2.append(model2_time_mean[xx]-spec_time_mean[xx])

# Creating daily mean for the model data
model1_dmean=np.zeros((np.size(sondeindex),days[ijk+1]-days[ijk]),dtype=float)
model2_dmean=np.zeros((np.size(sondeindex),days[ijk+1]-days[ijk]),dtype=float)
if site==1:
    modelsite_dmean=np.zeros((np.size(sondeindex),days[ijk+1]-days[ijk]),dtype=float)
for xx in range(0,np.size(model1[:,0])):
    for d in range(0,days[ijk+1]-days[ijk]):
        model1_time=[]
        model2_time=[]
        if site==1:
modelsite_time=[]
            for tt in range(0,24):
if mon==0 and d==0:
    model1_time=model1[xx,0:23].copy()
    model2_time=model2[xx,0:23].copy()
        if site==1:

# -----
# -----Selection of model data according to the five different selection methods -----
# -----

# Iterating over months
for ijk in range(0,np.size(month)):
    mon=month[ijk]

    # The layered model data 45m -->
    execfile('sondes_read.py')
    sondes_con = sonde_data(mon,species_sonde)
    # The site data at 3m
    execfile('sites_read.py')
    sites_con = sites_data(mon,stats,species_site)

    hrr=np.transpose(np.vstack((hr3,hr5,hr10)))
    k_altall=np.zeros(np.shape(hrr),dtype=float)
    k_relall=np.zeros(np.shape(hrr),dtype=float)
    k_abstopoall=np.zeros(np.shape(hrr),dtype=float)

```

```

k_reltopoall=np.zeros(np.shape(hrr),dtype=float)

# Iterating over the three different relative heights methods
for dd in range(0,np.size(hrr[0,:])):
    hr=hrr[:,dd]

    # The model data for the different extracting methods
    # Extracting model data for sites level (3m) and the level
    # corresponding to the relative hight
    mod_alt, mod_rel, k_alt, k_rel, mod_site =
    k_level(alt_snm, hr, sites_con, sondes_con, sondeindex,1)

    # Extracting model data for (abs-emep_topo) and (rel.h.-emep_topo)
    mod_abs_topo, mod_rel_topo, k_abstopo, k_reltopo =
    k_level(alt_snm-he_s, hr-he_s, sites_con, sondes_con, sondeindex,0)

    # The monthly and daily mean of the model and measurement data
    modalt_time_mean,modrel_time_mean,modsite_time_mean, spec_time_mean, modalt_dmean, modrel_dmean,
    modsite_dmean, modalt_err, modrel_err = mean(mod_alt, mod_rel, spec_h, meas_dmean, species_sonde, days,
    modabstopo_time_mean,modreltopo_time_mean, spec_time_mean, modabstopo_dmean, modreltopo_dmean,
    modabstopo_err, modreltopo_err = mean(mod_abs_topo, mod_rel_topo, spec_h, meas_dmean, species_sonde, days,

    # The corresponding sigma layers
    k_altall[:,dd]=np.asarray(k_alt)
    k_relall[:,dd]=np.asarray(k_rel)
    k_abstopoall[:,dd]=np.asarray(k_abstopo)
    k_reltopoall[:,dd]=np.asarray(k_reltopo)

    # ----- SAVING THE DATA -----
    mont=str(mon+1)
    # Save the hourly data for 03
    po=0
    if species_sonde==0 or species_sonde==1:
        spec=species[species_sonde]

        # The sonde file for january has no data for the first hour and
        # December has no data for the 3*24+21 last hours
        if mon==0:
po=1

data3=np.transpose(np.vstack((spec_h[:,1+days[ijk]*24:(days[ijk]*24+(days[ijk+1]-days[ijk])*24)],
mod_alt,mod_rel,mod_site)))

data31=np.transpose(np.vstack((spec_h[:,1+days[ijk]*24:(days[ijk]*24+(days[ijk+1]-days[ijk])*24)],
mod_abs_topo,mod_rel_topo)))
if mon==11:
po=1

data3=np.transpose(np.vstack((spec_h[:,days[ijk]*24:(days[ijk]*24+(days[ijk+1]-days[ijk])*24)],
mod_alt[:,0:624],mod_rel[:,0:624],mod_site[:,0:624])))

data31=np.transpose(np.vstack((spec_h[:,days[ijk]*24:(days[ijk]*24+(days[ijk+1]-days[ijk])*24)],
mod_abs_topo[:,0:624],mod_rel_topo[:,0:624])))
if po!=1:#mon!=0 or mon!=11:
data3=np.transpose(np.vstack((spec_h[:,days[ijk]*24:(days[ijk]*24+(days[ijk+1]-days[ijk])*24)],
mod_alt,mod_rel,mod_site)))

```

```

data31=np.transpose(np.vstack((spec_h[:,days[ijk]*24:(days[ijk]*24+(days[ijk+1]-days[ijk])*24)],
mod_abs_topo,mod_rel_topo)))
    txtname3=['Hourly meas data, Hourly alt data and Hourly Relative height stacked
on each other with all stations (nr of columns/3)']

    txtname31=['Hourly meas data, Hourly abs.-emep_topo data and Hourly rel.-emep_topo
stacked on each other with all stations (nr of columns/3)']

    np.savetxt('Hourly.coldiv3.alt.rel'+hrstr[dd]+'0'+mont+'08.'+spec+'ll.txt',
data3, delimiter=',',fmt='%-4s')
    np.savetxt('Hourly.coldiv3.abst.rel'+hrstr[dd]+'0'+mont+'08.'+spec+'ll.txt',
data31, delimiter=',',fmt='%-4s')
# If S04 save the daily data
if species_sonde==6 or species_sonde==2:
    if species_sonde==6:
spec=species[2]
        if species_sonde==2:
spec=species[3]
            data3=np.transpose(np.vstack((meas_dmean[:,days[ijk]:days[ijk+1]],
modalt_dmean,modrel_dmean,modsite_dmean)))

            data31=np.transpose(np.vstack((meas_dmean[:,days[ijk]:days[ijk+1]],
modabstopo_dmean,modreltopo_dmean)))
            txtname3=['Daily meas data, Daily alts data and Daily Relative height stacked
on each other with all stations (nr of columns/3)']

            txtname31=['Daily meas data, Daily abs.-emep_topo data and
Daily rel.-emep_topo stacked on each other with all stations (nr of columns/3)']

            np.savetxt('Daily.coldiv3.alt.rel'+hrstr[dd]+'0'+mont+'08.'+spec+'ll.txt',data3, delimiter=',',
            np.savetxt('Daily.coldiv3.abst.rel'+hrstr[dd]+'0'+mont+'08.'+spec+'ll.txt',data31, delimiter=',

# Save the monthly mean
txtname2=['mon_meas', 'alt_mean', 'rel_mean', 'site_mean', 'error(meas-alt)', 'error(meas-rel)']
txtname21=['mon_meas', 'alttopo_mean', 'reltopo_mean', 'error(meas-abstopo)', 'error(meas-reltopo)']

data2=np.transpose(np.vstack((spec_time_mean,modalt_time_mean,modrel_time_mean,
modsite_time_mean,np.transpose(modalt_err),np.transpose(modrel_err))))

data21=np.transpose(np.vstack((spec_time_mean,modabstopo_time_mean,modreltopo_time_mean,
np.transpose(modabstopo_err),np.transpose(modreltopo_err)))) # mon_meas[:,0]=monthly mean of January

np.savetxt('stat_data.alt.rel'+hrstr[dd]+'site.0'+mont+'08.'+spec+'.txt',
np.vstack((txtname2,data2)), delimiter=',',fmt='%-4s')

np.savetxt('stat_data.alttopot.reltopo'+hrstr[dd]+'0'+mont+'08.'+spec+'.txt',
np.vstack((txtname21,data21)), delimiter=',',fmt='%-4s')

print 'check'+mont+spec

kalt_3=20-k_altall[:,0];
kalt_5=20-k_altall[:,1];
kalt_10=20-k_altall[:,2];

krel_3=20-k_relall[:,0];
krel_5=20-k_relall[:,1];
krel_10=20-k_relall[:,2];

```

```
kabstopo_3=20-k_abstopoall[:,0];
kabstopo_5=20-k_abstopoall[:,1];
kabstopo_10=20-k_abstopoall[:,2];

kreltopo_3=20-k_reltopoall[:,0];
kreltopo_5=20-k_reltopoall[:,1];
kreltopo_10=20-k_reltopoall[:,2];

txtname1=['sondeindex', 'stats', 'alt_snm','kalt_3','kalt_5','kalt_10', 'krel_3','krel_5',
'krel_10','kabstopo_3','kabstopo_5','kabstopo_10','kreltopo_3','kreltopo_5','kreltopo_10','hr 3','hr5','hr10']

data1=np.transpose(np.vstack((sondeindex,stats,alt_snm,kalt_3,kalt_5,kalt_10,krel_3,krel_5,
krel_10,kabstopo_3,kabstopo_5,kabstopo_10,kreltopo_3,kreltopo_5,kreltopo_10,hrr[:,0],hrr[:,1],hrr[:,2])))

np.savetxt('stat_info_all.08.int.+spec+'.txt',np.vstack((txtname1,data1)), delimiter=',', fmt='%+4s')
```

B.2 NILUstat.py

```

# -*- coding: utf-8 -*-
import numpy as np
import scipy.io as sio

# -----
# This script extracts information altitude, coordinates
# (lat, lon) and (i,j) and names for the selected stations.
# -----

def NILUstat():
    sondestat=np.loadtxt('sondestat.dat',dtype='string')
    stat=np.loadtxt('stations.dat',skiprows=2,dtype='string')

    sondename=sondestat[:,0]
    ijcoord1=sondestat[:,5:8]

    #daily = sio.netcdf_file('/fou/emep/Auxiliary/NILU/data_daily.2008.nc', 'r')
    #hourly = sio.netcdf_file('/fou/emep/Auxiliary/NILU/data_hourly.2008.nc', 'r')
    daily=sio.netcdf_file('/home/moji/KF/MsTh/EmmaMET/data_daily.2008.nc', 'r')
    hourly=sio.netcdf_file('/home/moji/KF/MsTh/EmmaMET/data_hourly.2008.nc', 'r')

    STATION=hourly.variables['station']
    station=STATION[:]

    # Extracts the station index (in sonde file) and station names
    sondeindex=[]
    statname=[]
    for j in range(0,np.size(sondename)):
        for i in range(0,np.size(station[:,1])):
            if station[i,0]+station[i,1]+station[i,2]+station[i,3]==sondename[j]:
                statname.append(station[i,0]+station[i,1]+station[i,2]+station[i,3])
                sondeindex.append(j)

    ijcoord=np.zeros((np.size(sondeindex),3), dtype=int)
    for h in range(0,np.size(sondeindex)):
        ijcoord[h,:]=ijcoord1[sondeindex[h],:]

    alt=[]
    latlon_data=np.zeros((np.size(statname),8), dtype=str)
    lat=np.zeros((np.size(statname),3), dtype=float)
    lon=np.zeros((np.size(statname),3), dtype=float)
    #The altitudes of the stations
    for t in range(0,np.size(statname)):
        for q in range(0,np.size(stat[:,0])):
            if statname[t]==stat[q,0]:
                alt.append(float(stat[q,9]))
    latlon_data[t,:]=stat[q,1:9]
    lat[t,:]=stat[q,1:4]
    lon[t,:]=stat[q,5:8]

    # Rewrite all the coordinates to only degree and
    # rewrite westerly coordinates with a minus sign
    lat_s=[]
    lon_s=[]
    # Here the westerly lonituedata will be saved
    lon_data=np.zeros((np.size(latlon_data[:,1]),3),dtype=float)

```

```
for s in range(0,np.size(latlon_data[:,1])):
    if latlon_data[s,7]=='W':
        lon_data[s,0]=-float(lon[s,0])
        lon_data[s,1]=-float(lon[s,1])
        lon_data[s,2]=-float(lon[s,2])
    elif latlon_data[s,7]=='E':
        lon_data[s,0]=float(lon[s,0])
        lon_data[s,1]=float(lon[s,1])
        lon_data[s,2]=float(lon[s,2])
    la=float(lat[s,0]) + float(lat[s,1])/60 + float(lat[s,2])/(60*60)
    lat_s.append(la)
    lo=lon_data[s,0]+ lon_data[s,1]/60 + lon_data[s,2]/(60*60)
    lon_s.append(lo)

lat_s=np.asarray(lat_s)
lon_s=np.asarray(lon_s)

return alt,lat_s,lon_s,sondeindex,ijcoord,statname
```

B.3 NILUdata.py

```

# -*- coding: utf-8 -*-
import numpy as np
import scipy.io as sio

# -----
# This script extracts the NILU measurement data for the chosen
# pollutants for the selected stations.
# -----

def NILUdata(sondeindex,species):
    #species: 0=ozone, 6=S04, 1=NO2
    # Getting the measurement data from NILU
    daily = sio.netcdf_file('/fou/emep/Auxiliary/NILU/data_daily.2008.nc', 'r')
    hourly = sio.netcdf_file('/fou/emep/Auxiliary/NILU/data_hourly.2008.nc', 'r')
    #daily=sio.netcdf_file('/home/moji/KF/MsTh/EmmaMET/data_daily.2008.nc', 'r')
    #hourly=sio.netcdf_file('/home/moji/KF/MsTh/EmmaMET/data_hourly.2008.nc', 'r')

    # Extracting the measurement stations corresponding to daily data and O3, S04 and NH3
    STATION=daily.variables['station']
    station=STATION[:]
    EPDL626=daily.variables['EPDL626']
    ozone_dmax=EPDL626[:]
    EPDL608=daily.variables['EPDL608']
    ozone_dmean=EPDL608[:]
    EPDL53=daily.variables['EPDL53']
    sulphate_dmean=EPDL53[:]
    EPDL31=daily.variables['EPDL31']
    no2_dmean=EPDL31[:]

    # Extracting the measurement stations corresponding to hourly data and O3 and NH3
    Station_h=hourly.variables['station']
    station_h=Station_h[:]
    O3_h=hourly.variables['EPDL34']
    o3_h=O3_h[:]
    nitrogendioxide_h=hourly.variables['EPDL31']
    no2_h=nitrogendioxide_h[:]

    # Depending on the used species, the molar weight has to be
    # concluded in order to calculate the conversion factor from ug/m3 to ppm
    if species==0:
        M=3*16
        spe_dmean=ozone_dmean
    if species==6:
        M=96
        spe_dmean=sulphate_dmean
    if species==1:
        M=14+16*2
        spe_dmean=no2_dmean
    conv=1/(0.0409*M)

    # Extracting the station names that has model data and then
    # compares them with the measurement station and collect those how has both
    sondestat=np.loadtxt('sondestat.dat',dtype='string')
    statname=[]
    for x in range(0,np.size(sondeindex)):
        statname.append(sondestat[sondeindex[x],0])

```

```

pos=[]; pos_h=[]
stats=[]; stats_h=[]
for j in range(0,np.size(statname)):
    for i in range(0,np.size(station[:,1])):
        if station[i,0]+station[i,1]+station[i,2]+station[i,3]==statname[j]:
pos.append(i)
stats.append(statname[j])
    for i in range(0,np.size(station_h[:,1])):
        if station_h[i,0]+station_h[i,1]+station_h[i,2]+station_h[i,3]==statname[j]:
pos_h.append(i)
stats_h.append(statname[j])

# Extracting the daily mean and hourly data for the corresponding stations
dmean=np.zeros((np.size(pos),np.size(spe_dmean[0,:])), dtype=float)
ozone_h=np.zeros((np.size(pos),np.size(o3_h[0,:])), dtype=float)
NO2_h=np.zeros((np.size(pos),np.size(no2_h[0,:])), dtype=float)
for q in range(0,np.size(pos)):
    dmean[q,:]=spe_dmean[pos[q],:]
for y in range(0,np.size(pos_h)):
    ozone_h[y,:]=o3_h[pos_h[y],:]
    NO2_h[y,:]=no2_h[pos_h[y],:]

if species==0:
    return ozone_h,dmean,stats_h
if species==1:
    return NO2_h,dmean,stats_h
if species==6:
    return dmean, stats

```

B.4 heights.py

```

# -*- coding: utf-8 -*-
#!/usr/bin/env python

import numpy as np
import scipy.io as sio
import matplotlib.pyplot as plt
import matplotlib.cm as cm

def heights(lat_s, lon_s, ij_coord, index, alt_s):

#####
# -Input: latlon-coordinates for the Sonde-NILU stations both on
# (i,j) and (lat,lon), index for the station, altitude for the stations
#
# -Get topography
# -Rewrite to west with a minus sign fo latitude
# -Check if station is within boundary, save those stations
# -Get EMEP topography and the EMEP height for the valid stations
# -Calculate the relative height for the stations (not the complete map)
# (Have to choose the mountain stations only!)
#
# -Output: Relative height, the index corresponding to the mountain stations
# in the map, EMEP grid height fir the mountain stations in the map
#####

# The GTOPO30 data
mapG30=sio.netcdf_file('wcs.nc','r')
z=mapG30.variables['Band1']
alt=z[:]
alt=alt.astype(float)

# Create lat lon data -25 E to 39.999999999974 W and
# 35 S to 79.999999999982 N, x:[0:7800], y:[0,5400], each step 0.00833333
# And set all sea values to zero instead of negative values
lat=[]
lon=[]
lon_start=-25
lat_start=33
step=0.00833333
for i in range(0,np.size(alt[:,0])): # shape(z_alt)=(5400, 7800)=(y,x)
    lat.append(lat_start+step*i)
for j in range(0,np.size(alt[0,:])):
    lon.append(lon_start+step*j)

# Check if station lies within the boundary of the maps
lat_sn=[]
lon_sn=[]
index_n=[]
alt_sn=[]
i_n=[]
j_n=[]
for ii in range(0,np.size(lat_s)):
    if min(lat)<lat_s[ii]<max(lat) and min(lon)<lon_s[ii]<max(lon):
        lat_sn.append(lat_s[ii])
        lon_sn.append(lon_s[ii])
        index_n.append(index[ii])

```

```

    alt_sn.append(alt_s[ii])
    i_n.append(ij_coord[ii,0])
    j_n.append(ij_coord[ii,1])

# Extracting indices for the stations
# (to be used when extracting their relative height from the map)
lo_s=[]
la_s=[]
for m in range(0,np.size(lat_sn)):
    ab=abs(lat-lat_sn[m])
    la_s.append((ab==min(ab)).nonzero()[0][0])
for n in range(0,np.size(lon_sn)):
    ab1=abs(lon-lon_sn[n])
    lo_s.append((ab1==min(ab1)).nonzero()[0][0])

la_s=np.asarray(la_s)
lo_s=np.asarray(lo_s)
alt_sn=np.asarray(alt_sn)
index_n=np.asarray(index_n)

# Deciding how many grid spaces 5km corresponds to.
# 25+-35, 34+-45, 45+-55, 55+-65, 65+-75, 75+-85
cc=np.array([804, 712, 598, 465, 318, 272])
dd=np.array([924, 925, 927, 929, 930, 930])
d_f=np.mean(dd)
d5_f=5000/cc-1
d3_f=3000/cc-1
d10_f=10000/cc-1
d3=[]
d5=[]
d10=[]
for c in range(0,np.size(lat_sn)):
    if float(25) <lat_sn[c]<= float(35):
        i=0
        d3.append(d3_f[i])
        d5.append(d5_f[i])
        d10.append(d10_f[i])
    elif float(35) <lat_sn[c]<=float(45):
        i=1
        d3.append(d3_f[i])
        d5.append(d5_f[i])
        d10.append(d10_f[i])
    elif float(45) <lat_sn[c]<= float(55):
        i=2
        d3.append(d3_f[i])
        d5.append(d5_f[i])
        d10.append(d10_f[i])
    elif float(55) <lat_sn[c]<= float(65):
        i=3
        d3.append(d3_f[i])
        d5.append(d5_f[i])
        d10.append(d10_f[i])
    elif float(65) <lat_sn[c]<= float(75):
        i=4
        d3.append(d3_f[i])
        d5.append(d5_f[i])
        d10.append(d10_f[i])
    elif 75.000000 <lat_sn[c]<= 85.000000:

```

```

    i=5
    d3.append(d3_f[i])
    d5.append(d5_f[i])
    d10.append(d10_f[i])

# Relative altitude calculations
hr5=[]
hr10=[]
hr3=[]
h5=[]
h10=[]
h3=[]
la_s=np.size(alt[:,0])-la_s # Some kind of adjustment
for jj in range(0,np.size(alt_sn)): # Same range as la_s and lo_s
    h5=alt[(la_s[jj]-d5[jj]):(la_s[jj]+d5[jj]+1),(lo_s[jj]-4):(lo_s[jj]+5)]
    for mm in range(0,np.size(h5[:,1])):
        for mn in range(0,np.size(h5[1,:])):
if h5[mm,mn]<0:
    h5[mm,mn]=0
    h10=alt[(la_s[jj]-d10[jj]):(la_s[jj]+d10[jj]+1),(lo_s[jj]-9):(lo_s[jj]+10)]
    for pp in range(0,np.size(h10[:,1])):
        for pq in range(0,np.size(h10[1,:])):
if h10[pp,pq]<0:
    h10[pp,pq]=0
    h3=alt[(la_s[jj]-d3[jj]):(la_s[jj]+d3[jj]+1),(lo_s[jj]-2):(lo_s[jj]+3)]
    for oo in range(0,np.size(h3[:,1])):
        for op in range(0,np.size(h3[1,:])):
if h3[oo,op]<0:
    h3[oo,op]=0
    hr5.append(alt_sn[jj]-h5.min())
    hr10.append(alt_sn[jj]-h10.min())
    hr3.append(alt_sn[jj]-h3.min())
hr5=np.asarray(hr5)
hr10=np.asarray(hr10)
hr3=np.asarray(hr3)

return hr3, hr5, hr10, alt_sn, index_n

```

B.5 sondes_read.py

```

#!/usr/bin/env python
# -*- coding: utf-8 -*-

#-----
# Script that reads the sonde files
#-----

import numpy as np
import scipy.io as sio

def sonde_data(mon,species):
    mon=str(mon+1)
    if int(mon)<=9:
        sond_sc=np.loadtxt('sondes.0'+mon+'08',dtype=np.dtype(np.dtype))
    else:
        sond_sc=np.loadtxt('sondes.'+mon+'08',dtype=np.dtype(np.dtype))

    # Extracting the coordinates for the different stations in the sonde file
    # Remember that the counting of the above created array does not tak up until the last value just up...
    # The number of stations
    a=sond_sc[0]
    # Creating a float out of it
    b=float(a[0])
    # Index to the rows that should be read out for the stations
    ind=np.r_[2:(b+2-1+1)]
    # Index to the columns corresponind to the coordinates: grad, min, sec
    jnd=np.r_[5:8]
    # How many pollutants
    aa=sond_sc[b+2]
    # Creating a float out of it
    bb=float(aa[0])
    # Index to the rows that should be read out for the pollutants
    iind=np.r_[(b+2+1):(b+2+bb)]

    coord=np.zeros((np.size(ind),np.size(jnd)),dtype=float)
    names=[]
    r=[]
    for i in range(0,np.size(ind)):
        r=sond_sc[ind[i]]
        names.append(r[0])
        for j in range(0,np.size(jnd)):
            coord[i,j]=float(r[jnd[j]])

    # If a new sondestat.dat needs to be written
    #text1=['Station','i','j']
    #data1=np.transpose(np.vstack((names,coord[:,0],coord[:,1])))
    #np.savetxt('sondestat.dat',np.vstack((text1,data1)), delimiter=',',          ',fmt='%-4s')

    poll_names=[]
    q=[]
    for ii in range(0,np.size(iind)):
        q=sond_sc[iind[ii]]
        poll_names.append(q[1])

    del sond_sc # To free some memory...

```

```
# The actual sonde data
if int(mon)<=9:
    sond_data=np.loadtxt('sondes.0'+mon+'08',dtype=np.dtype(np.dtype),skiprows=109)
else:
    sond_data=np.loadtxt('sondes.'+mon+'08',dtype=np.dtype(np.dtype),skiprows=109)

conc_allstat=[]
for k in range(0,np.size(sond_data)):
    o=0
    h=sond_data[k]
    for l in range(0,np.size(names)):
        if k==0 and l==0:
            conc_allstat.append(np.hstack((sond_data[k+species],sond_data[k+species+1])))
            elif h[0]==names[l] and o==0:
                o=1
            conc_allstat.append(np.hstack((sond_data[k+species+1],sond_data[k+species+2])))

# Rewrite the ozone data into floats
arr_conc=np.zeros((np.shape(conc_allstat)),dtype=float)
for l in range(0,np.size(conc_allstat)/20):
    he=conc_allstat[l]
    for ll in range(0,np.size(he)):
        arr_conc[l,ll]=float(he[ll])
return arr_conc
```

B.6 sites_read.py

```

#!/usr/bin/env python
# -*- coding: utf-8 -*-

#-----
# Script that reads the sites files
# - Remember to enter the first station in the station list into line 28 and 29.
#-----

import numpy as np
import scipy.io as sio

def sites_data(mon,statname,species):
    mon=str(mon+1)
    if int(mon)<=9:
        site_data=np.loadtxt('sites.0'+mon+'08',dtype=np.dtype(np.dtype), skiprows=246)
    else:
        site_data=np.loadtxt('sites.'+mon+'08',dtype=np.dtype(np.dtype), skiprows=246)

    conc_allstat=[]
    names=[]
    for k in range(0,np.size(site_data)):
        o=0
        h=site_data[k]
        h_name=h[0]
        name=h_name.rsplit('_')
        for l in range(0,np.size(statname)):
            if k==0 and l==0 and statname[l]=='AT05': # The first station in the station list
names.append('AT05')
h1=site_data[species[0]]
conc_allstat.append(h1[species[1]])
            if name[0]==statname[l] and o==0:
names.append(name[0])
o=1
h1=site_data[k+species[0]+1]
conc_allstat.append(h1[species[1]])

    # Rewrite the ozone data into floats
    arr_conc=np.zeros((np.size(statname),np.size(conc_allstat)/np.size(statname)),dtype=float)
    for j in range(0,np.size(arr_conc[:,0])):
        for ii in range(0,np.size(arr_conc[0,:])):
            i=j+ii*np.size(arr_conc[:,0])
            arr_conc[j,ii]=float(conc_allstat[i])
    return arr_conc

```