

# ICOS Atmospheric Stations: Spatial Characterization of CO<sub>2</sub> Footprint Areas and Evaluating the Uncertainties of Modelled CO<sub>2</sub> Concentrations

**Ida Storm**

---

2020  
Department of  
Physical Geography and Ecosystem Science  
Centre for Geographical Information Systems  
Lund University  
Sölvegatan 12  
S-223 62 Lund  
Sweden



Ida Storm (2020). ICOS Atmospheric Stations: Spatial Characterization of CO<sub>2</sub> Footprint Areas and Evaluating the Uncertainties of Modelled CO<sub>2</sub> Concentration. Master degree thesis, 30 credits in Master in Geographical Information Science Department of Physical Geography and Ecosystem Science, Lund University

ICOS Atmospheric Stations: Spatial Characterization of CO<sub>2</sub> Footprint  
Areas and Evaluating the Uncertainties of Modelled CO<sub>2</sub> Concentration

---

Ida Storm

Master thesis, 30 credits, in Geographical Information Sciences

Supervisors:

Ute Karstens and Harry Lankreijer

Department of Physical Geography and Ecosystem Science,  
Lund University

## Abstract

The main purpose of this thesis is to present and analyze spatial characteristics of 31 different European atmospheric measurement stations in the ICOS (Integrated Carbon Observation System) station network. The characterization includes quantification of where air arriving at the stations can be expected to have come from, as well as a breakdown of what these areas cover with regards to land cover, point source emissions and population. A dataset regarding emissions of radiocarbon at nuclear power plants has also been processed because possible transports of radiocarbon from these facilities to the stations need to be accounted for when quantifying fossil fuel emissions based on measured ratio between  $^{14}\text{C}$  and  $^{12}\text{C}$ . “Where the air arriving can be expected to have come from” for a specific date and time is synonymous with a station’s footprint. For a general characterization based on annual values, an average footprint based on all three-hourly footprints in the year is used. Each individual footprint has already been combined with data on anthropogenic emissions and a model that quantifies the biospheric component to estimate the  $\text{CO}_2$  concentration at the stations. The known sources that make up the total  $\text{CO}_2$  signal also allows for a breakdown into different categories of contribution. Averages of these are also part of the characterization. The annual averages for the different characterizations vary greatly between the stations: station Pallas’ anthropogenic contribution is only 2.2% of the estimated contribution at the station with the highest value, Heidelberg. Furthermore, there are large variabilities between the footprints at the individual stations that are used to generate annual averages. Individual footprint values are used in closer analysis exemplified for selected stations. However, years 2016 and 2017 show similar annual values for the different stations which indicate stability of the annual characterizations. Dominant land cover classes in the model domain, including ocean, cropland, pastures and different types of forests, are found within the annual footprints of all stations. This is no surprise because of the large spatial extent of most footprints: on average 48% of the sensitivity is to the area within 300 km of the station, again with large differences between stations and between individual footprints of the stations. The large footprint extent is also why the spatial characterization is extra important: it is not possible to know what is within the footprint areas without it. Knowing what is in the area in close proximity to the stations is not enough.

## Table of contents

Abstract .....	iv
Table of contents .....	v
List of abbreviations .....	vii
List of tables.....	viii
List of figures.....	viii
<b>1. Introduction.....</b>	<b>1</b>
1.1 Problem statements and goals.....	1
1.2 Research questions.....	3
1.3 Thesis structure .....	3
<b>2. Background .....</b>	<b>5</b>
2.1 Transport models .....	5
2.2 Biospheric models.....	6
2.3 Anthropogenic contribution.....	7
2.4 Related studies .....	9
<b>3. Methods.....</b>	<b>11</b>
3.1 Study area.....	11
3.2 Data .....	11
3.2.1 Stations.....	11
3.2.2 Land cover .....	14
3.2.3 Population .....	16
3.2.4 Point source.....	17
3.2.5 Radiocarbon .....	17
3.3 Data processing to match footprint.....	18
3.3.1 Point data .....	18
3.3.2 Polygon data.....	19
3.4 Data analysis .....	20
3.4.1 Jupyter Notebooks .....	20
3.4.2 Sensitivity area evaluation .....	20
3.4.3 Breakdown modelled concentration .....	22
3.4.4 Ancillary data.....	22
3.4.4 Modelled concentration compared to measured concentration ..	22
<b>4. Results .....</b>	<b>25</b>
4.1 Sensitivity area evaluation .....	25
4.2 Breakdown modelled concentration .....	28

4.3 Ancillary data.....	32
4.3.1 Land cover .....	32
4.3.2 Population .....	36
4.3.3 Point source.....	38
4.3.4 Radiocarbon .....	40
4.4 Modelled concentration compared to measured concentration .....	43
<b>5. Discussion.....</b>	<b>51</b>
<b>6. Conclusion .....</b>	<b>57</b>
<b>References.....</b>	<b>59</b>
<b>Appendices.....</b>	<b>63</b>
1. Jupyter Notebook structure .....	63

## Abbreviations

CDIAC	Carbon Dioxide Information Analysis Center
CORINE	Coordination of Information on the Environment
CO <sub>2</sub>	Carbon dioxide
ECMWF	European Centre for Medium-Range Weather Forecasts
EDGAR	Emissions Database for Global Atmospheric Research
E-PRTR	European Pollutant Release and Transfer Register
EPSG	European Petroleum Survey Group
ESRI	Environmental Systems Research Institute
EU	European Union
EVI	Enhanced Vegetation Index
FAOSTAT	Food and Agriculture Organization Corporate Statistical Database
GEE	Gross Ecosystem Exchange
GPP	Gross Primary Production
GIS	Geographic Information System
HYSPLIT	The Hybrid Single Particle Lagrangian Integrated Trajectory Model
ICOS	Integrated Carbon Observation System
ICOS RI	ICOS Research Infrastructure
ICOS CP	ICOS Carbon Portal
IPCC	Intergovernmental Panel on Climate Change
MODIS	Moderate Resolution Imaging Spectroradiometer
NetCDF	Network Common Data Form
NOAA MB	National Oceanic and Atmospheric Administration
LSWI	Land Surface Water Index
NEE	Net Ecosystem Exchange
PBL	Planetary Boundary Layer
PPM	Parts Per Million
RADD	Radioactive Discharges Database
STILT	Stochastic Time-Inverted Lagrangian Transport model
TROTREP	Tropospheric Ozone and Precursors - Trends, budgets and policy
UNFCCC	United Nations Framework Convention on Climate Change
VPRM	Vegetation Photosynthesis and Respiration Model

## List of tables

Table 1	<i>Aggregation of IPCC categories into source categories “industry”, “energy”, “transport” and “others”.</i>	9
Table 2	<i>Stations considered for the purpose of this thesis including technical specifications.</i>	12
Table 3	<i>CORINE land cover classes (44) and their corresponding aggregated classes (19).</i>	14
Table 4	<i>Land cover breakdown for the STILT domain compared to the average breakdown of the 31 different stations.</i>	33
Table 5	<i>Expected average shift in radiocarbon from nuclear power plants and fuel reprocessing stations at the different stations year 2016.</i>	41
Table 6	<i>Average model data differences for year 2017.</i>	43

## List of figures

Figure 1	<i>Example of footprint and time series showing modelled and measured CO<sub>2</sub> concentrations.</i>	6
Figure 2	<i>Station locations with station names.</i>	13
Figure 3	<i>Station locations in relation to the extent of the domain.</i>	13
Figure 4	<i>The STILT domain compared to the CORINE land cover classes.</i>	16
Figure 5	<i>Map showing the spatial distribution of radiocarbon emissions from nuclear power plants and fuel reprocessing stations within the STILT domain year 2016.</i>	18
Figure 6	<i>Model to process the land cover data created with “model builder” in ArcMap.</i>	19
Figure 7	<i>Method illustration regarding what cells are included when considering sensitivity within specific distances.</i>	21
Figure 8	<i>Average estimated sensitivity within 100, 200 and 300 km of the stations as well as the whole STILT domain.</i>	26
Figure 9	<i>Time series showing each footprint’s sensitivity to the whole domain for Hyltemossa.</i>	26
Figure 10	<i>Average sensitivity by month, years 2016 and 2017.</i>	27
Figure 11	<i>Average sensitivity by hour, years 2016 and 2017.</i>	27
Figure 12	<i>Percent sensitivity to the outermost rows and columns of the domain for year 2017.</i>	28
Figure 13	<i>Average anthropogenic contribution and biospheric component year 2017.</i>	29
Figure 14	<i>Average anthropogenic contribution broken down into source categories year 2017.</i>	29
Figure 15	<i>Anthropogenic and biospheric contributions winter (Dec 2016-Feb 2017) vs. summer (Jun 2017-Aug 2017).</i>	30
Figure 16	<i>Monthly variation in total anthropogenic emissions for Europe years 2016-2020.</i>	30
Figure 17	<i>Anthropogenic and biospheric contribution winter vs. summer with the anthropogenic contribution broken down into the different source - and fuel categories for Hyltemossa</i>	31

Figure 18	<i>Anthropogenic contribution compared to sensitivity winter vs. summer.</i>	32
Figure 19	<i>Distribution of the four major land cover classes within the year 2017 average footprints for all stations.</i>	34
Figure 20	<i>Percent coniferous forests within the year 2017 average footprints.</i>	35
Figure 21	<i>Land cover breakdown for Hyltemossa year 2017.</i>	35
Figure 22	<i>Spatial distribution of the sensitivity to the land cover class "Cropland: staple except rice" at Hyltemossa.</i>	36
Figure 23	<i>Land cover breakdown for the major land cover classes at Hyltemossa years 2015-2018.</i>	36
Figure 24	<i>Map of population density within the STILT domain.</i>	37
Figure 25	<i>Station sensitivity to population year 2017.</i>	37
Figure 26	<i>Spatial distribution of sensitivity to population at Svartberget and Heidelberg year 2017.</i>	38
Figure 27	<i>Point source emissions year 2017 within the STILT domain.</i>	39
Figure 28	<i>Expected average shift in radiocarbon from nuclear power plants and fuel reprocessing stations year 2016.</i>	39
Figure 29	<i>Sensitivity to point source emissions compared to population year 2017.</i>	40
Figure 30	<i>Map showing Heidelberg's sensitivity to radiocarbon emissions year 2016.</i>	42
Figure 31	<i>Time series showing influence from radiocarbon emissions on the <sup>14</sup>C to <sup>12</sup>C ratio from individual facilities at Heidelberg January year 2016.</i>	42
Figure 32	<i>Time series showing influence from radiocarbon emissions on the <sup>14</sup>C to <sup>12</sup>C ratio at Heidelberg year 2016.</i>	43
Figure 33	<i>Time series with modelled and measured CO<sub>2</sub> with resulting model data differences in a separate graph, Trainou year 2017</i>	44
Figure 34	<i>Average model data differences by month for year 2017</i>	45
Figure 35	<i>Count model data differences one standard deviation below measured concentrations by month year 2017.</i>	46
Figure 36	<i>Count model data differences one standard deviation above measured concentrations by month year 2017.</i>	46
Figure 37	<i>Average model data difference by hour for year 2017.</i>	47
Figure 38	<i>Count model data differences one standard deviation below measured concentrations by hour year 2017.</i>	47
Figure 39	<i>Count model data differences one standard deviation above measured concentrations by hour year 2017</i>	48
Figure 40	<i>Areas associated with high overestimates at Heidelberg in a relative footprint map, and a separate map showing point source emission locations and quantities in close proximity to Heidelberg. Year 2017.</i>	49



# 1. Introduction

A warning climate due to an increasing amount of greenhouse gases in our atmosphere is nowadays generally accepted as a fact. There are several different gases that act as greenhouse gases, and many of them occur naturally in the atmosphere (IPCC, 2018). However, the main problem is the anthropogenic emissions which have been increasing the concentrations in the atmosphere to unprecedented levels. Since the year 2000 the estimated rise in global CO<sub>2</sub> concentration is 20 ppm per decade and the increasing concentration can be translated into an increase in global average temperature: since 1970 the average rate is about 1.7 °C per century (IPCC, 2018, p. 54). In Europe as a whole the rise in emissions has been stopped and even slightly reversed with the year 2012 emissions being 12% lower than year 1970. However, the corresponding number for the global CO<sub>2</sub> emissions is a 180% increase (Janssens-Maenhout et al., 2017, p. 41). Anthropogenic emissions of CO<sub>2</sub> released into the atmosphere become part of the carbon cycle, which also includes the terrestrial ecosystems and oceans. The exchanges of carbon between the different carbon reservoirs are called fluxes. The terrestrial ecosystems absorb CO<sub>2</sub> through photosynthesis and release it through respiration. Photosynthesis and respiration are also part in the exchange with the oceans, but CO<sub>2</sub> also dissolves into water by reacting with water molecules.

## 1.1 Problem statement and goals

To reach the goal of keeping the global average temperature to well below 2 °C above pre-industrial levels – in accordance with the Paris Agreement signed by 194 countries year 2016 – global emissions need to be reduced fast. However, our ability to reach the goal will also depend on the global carbon reservoirs and if they will continue to absorb  $3.0 \pm 0.8 \text{ PgC yr}^{-1}$  of the  $9.4 \pm 0.5 \text{ PgC yr}^{-1}$  anthropogenic emissions to the atmosphere (Lansø, A. S. et al. 2018). It is unsure how this might change with for instance the changing climate and ocean acidification. ICOS, a European research infrastructure, has a network of measurement stations and operates with the mission to “understand the carbon cycle and to provide necessary information on greenhouse gases” (ICOS Handbook, 2019, p. 12). There are atmospheric, ecosystem and oceanic station types, each with different focuses: ecosystem sites are mainly concerned with the biospheric greenhouse gas fluxes, ocean sites with ocean-atmosphere exchange fluxes, and at the atmospheric sites atmospheric greenhouse gas concentrations are measured. Currently there are 130 stations in twelve countries. For each station type there is a corresponding thematic centre where the data are coordinated and processed before the data reach the ICOS Carbon Portal hosted by Lund University. The Carbon Portal offers access to the collected data to both researchers and the public, as well as data products that regard the science of the carbon cycle and quantifying emission sinks and sources in Europe (ICOS Handbook, 2019). An example is the “STILT footprint tool” at the Carbon Portal webpage (“STILT result viewer”, n.d) where CO<sub>2</sub> concentrations are estimated based on what fluxes – anthropogenic and biospheric – the air passed on its way to the measurement sensors at the ICOS stations (see *Fig. 1*). The quantified air transport to a location given a specific point in time is referred to as a “footprint” and is a grid with cell values corresponding to the cell area’s sensitivity to potential fluxes (see section 2.1). The process of estimating CO<sub>2</sub> concentrations in the footprint tool is referred to as the “STILT model system” throughout this thesis.

In this thesis, footprints are the foundation for a characterization of 31 different atmospheric stations in the ICOS network. With measurement sensors located on tall towers this station type generally has large footprints and hence large areas of potential surface fluxes influencing the CO<sub>2</sub> concentrations. The characterization offers additional information about the footprint areas, such as how much of the sensitivity are to the area close to the station as opposed to far away, and quantifications of ancillary data – including land cover and population – within the footprint areas by multiplying footprints with the data and summarizing the resulting cells values. By generating annual average values, the stations have been ranked among themselves which is valuable information when for instance selecting stations to work with given specific purposes. It has also proven valuable to characterize individual footprints, representing a single point in time, to understand why modelled CO<sub>2</sub> concentration values are different from measured concentration values. The differences stem from uncertainties in the transport model and our ability to correctly quantify and spatially resolve surface fluxes and these cannot be fully disentangled – but feedback from characterizing footprints can potentially help improve the model.

It is possible to subset the influence anthropogenic emissions can be expected to have on the concentration at a given station, either for a single point in time or on average, and further into subcategories. The result will depend on the footprint, as well as the anthropogenic emission dataset which is mainly based on country reports (see section 2.3). With uncertainties in modelled CO<sub>2</sub> concentrations stemming from several different sources, it is not possible to use the difference between modelled and measured concentrations to verify the reported anthropogenic emission values. An alternative, or rather additional, approach to estimate specifically the fossil fuel component within a footprint area - without relying on reported emissions - is by using measurements of the radioactive carbon isotope <sup>14</sup>C (radiocarbon). At ICOS's Central Radiocarbon Laboratory in Heidelberg they are currently working on optimizing the sampling strategy to be integrated as a part of standard ICOS observations (ICOS strategy, 2019). Radiocarbon is created naturally in the atmosphere when cosmic rays induce reactions of neutrons with atmospheric nitrogen (Levin et al., 2003, p. 1) which means that the ratio between <sup>14</sup>C and the much more common isotope <sup>12</sup>C is relatively stable in the clean background air as well as in living biological matter such as animals and plants. Radioactivity decreases over time and for radiocarbon the half-life is 5370 years which means that fossil fuels, which are millions of years old, have no radioactive isotopes left. In turn, burning them emits CO<sub>2</sub> that offset the atmospheric ratio between the carbon isotopes. The size of the offset can in turn be used to quantify the fossil fuel emissions within the footprint area. However, better estimates can be made if anthropogenic emissions of radiocarbon from nuclear power plants and fuel reprocessing stations are accounted for. These facilities are treated as an ancillary data layer and resulting annual averages indicate at which stations their influence should be considered.

All the analyzes have been run in something called Jupyter Notebooks (see section 3.4.1) which will be distributed through the ICOS Carbon Portal as data products. It is possible to define different date ranges and to only run analyzes for selected stations. Furthermore, new locations within the STILT model domain can be defined in the “STILT footprint tool” and computations be made for a specified date range. Resulting footprints with corresponding modelled concentrations will be viewable on the webpage, as well as automatically available to use within the Notebooks which

means potential station locations can be characterized and compared to existing stations. The information can be useful in the process of deciding where to located new ICOS stations.

## 1.2 Research questions

The main objective is to spatially characterize the different ICOS atmospheric stations, but also to look at the STILT model system's ability to predict CO<sub>2</sub> concentrations at the station. Specific questions include:

- How do the stations differ in terms of sensitivity to the model domain, both among themselves and between individual footprints at the same station? In turn, generally how sensitive are the individual stations to areas close to the stations, as opposed to far away?
- What are the main land cover classes the stations are sensitive to? How does the land cover breakdown look like for the stations compared to the land cover breakdown of the whole domain?
- How well does the STILT model system generally predict CO<sub>2</sub> concentrations at the different stations, and can any hourly and/or monthly patterns be detected in the model performance?

## 1.3 Thesis structure

The thesis begins with a background section that includes relevant research, as well as technical specifications about transport models (2.1), biospheric models (2.2) and anthropogenic emissions (2.3) with specific focus on the STILT transport model, the VPRM biospheric model and the EDGAR anthropogenic emissions dataset. These combined make up the STILT model system used to generate modelled CO<sub>2</sub> values.

In the methods section, the ancillary data that have been used are described. Also, the 31 ICOS stations are described in terms of name, location, station type and measuring height. Section 3.3 describes how the ancillary data was prepared to be combined with footprints within the Notebooks and in section 3.4 the general methods used to process the data within the different Notebooks are described.

Section 3.4, as well as the whole result section, are structured around four Notebooks where all the analyses have taken place: "Sensitivity area evaluation", "Breakdown of modelled concentration", "Ancillary data" and "modelled concentration compared to measured concentration". In the results, the ancillary data section is further subdivided into "Land cover", "Population", "Point source" and "Radiocarbon".

The final sections are discussion, 5, and conclusion, 6, followed by references and appendices.



## 2. Background

### 2.1 Transport models

Put simply, a transport model uses as input meteorological data to simulate how the air travels to a receptor, such as a measurement sensor, located at a specific location and elevation within the model domain. Transports are simulated for a specific date and time, and the air is traced for a specified time window. Different transport models vary both by input data used and the exact process of how movement is simulated, but are generally either lagrangian or eulerian. In lagrangian transport models, individual particles' trajectories are considered, whereas eulerian models are using a fixed grid on which transport and diffusion of concentration plumes is calculated. The footprints used throughout this thesis have been derived using a lagrangian transport model called Stochastic Time-Inverted Lagrangian Transport (STILT) model (Lin et al. 2003). A large number of hypothetical air particles of equal mass are traced to any location within the transport model domain (Lin et al., 2003, p. 7). The direction and speed of the particles change is based on two factors: mean wind derived from the input data as well as a random – stochastic – component simulated by a Markov chain process (Lin et al., 2003, p. 2). The random component is why the individual particle trajectories will differ despite being released from the same location and the reason why running the tool for the same point in time more than once will yield footprints that are slightly different. Different data can be input to the model and for the STILT model system data from ECMWF (European Centre for Medium-Range Weather Forecasts, 2006) were used. This is in accordance with the paper by Vardag et al. (2015). The spatial resolution of the meteorological fields is 25 km x 25 km with 49 vertical levels, and the temporal resolution is three hours.

The particles are, in the case of the STILT model system, tracked for ten days. They can be tracked either forwards or backwards in time and establishing the influence region for a specific location, such as an atmospheric ICOS station, is more computationally efficient using the backward mode: only the trajectories of particles that will “end up” at the station receptor need to be simulated. With a forward simulation, particles would have to be released from all over the STILT domain and only the trajectories of the particles ending up at the station would be considered for establishing the influence region quantified as a footprint (Lin et al., 2003 p. 2). The forward and backward simulations have been shown to generate similar results (Lin et al., 2003, p. 15). Once all the particle trajectories are established, the residence times within the lower half of the planetary boundary layer (PBL) – which is dependent on current meteorological conditions – are used to derive a footprint with cell values representing the sensitivity to potential surface fluxes within the cell in ppm / ( $\mu\text{mol} / \text{m}^2\text{s}$ ). The equation describing this process is found in the paper by Lin et al. (2003, p. 3, equation 8). The output footprint grid is  $1/12^\circ$  latitude  $\times$   $1/8^\circ$  longitude corresponding to about 10 km  $\times$  10 km (Vardag et al., 2015, p.12708) but with the length in longitude direction decreases moving north from the equator. The quantification of the surface influence in each cell is typically done every 0.25 h to 1 h (Vardag et al., 2015, p.12708), and the output footprints are used in the STILT model system to calculate the modelled concentrations for a single point in time. The resulting modelled concentration values are downloaded for use in the analysis where modelled concentration values are quantified. The footprint representations in the Carbon Portal's footprint tool, see *Fig. 1*, are the result of aggregating these into one

footprint for the whole ten-day time-window. This type of time-step aggregated footprints is used in all other analysis. However, this will have no influence on the result of multiplying the footprints with the ancillary data layers because they have no temporal resolution: yearly quantities of point source and radiocarbon emissions are distributed evenly over the whole year and in terms of population and land cover the values represent certain years.

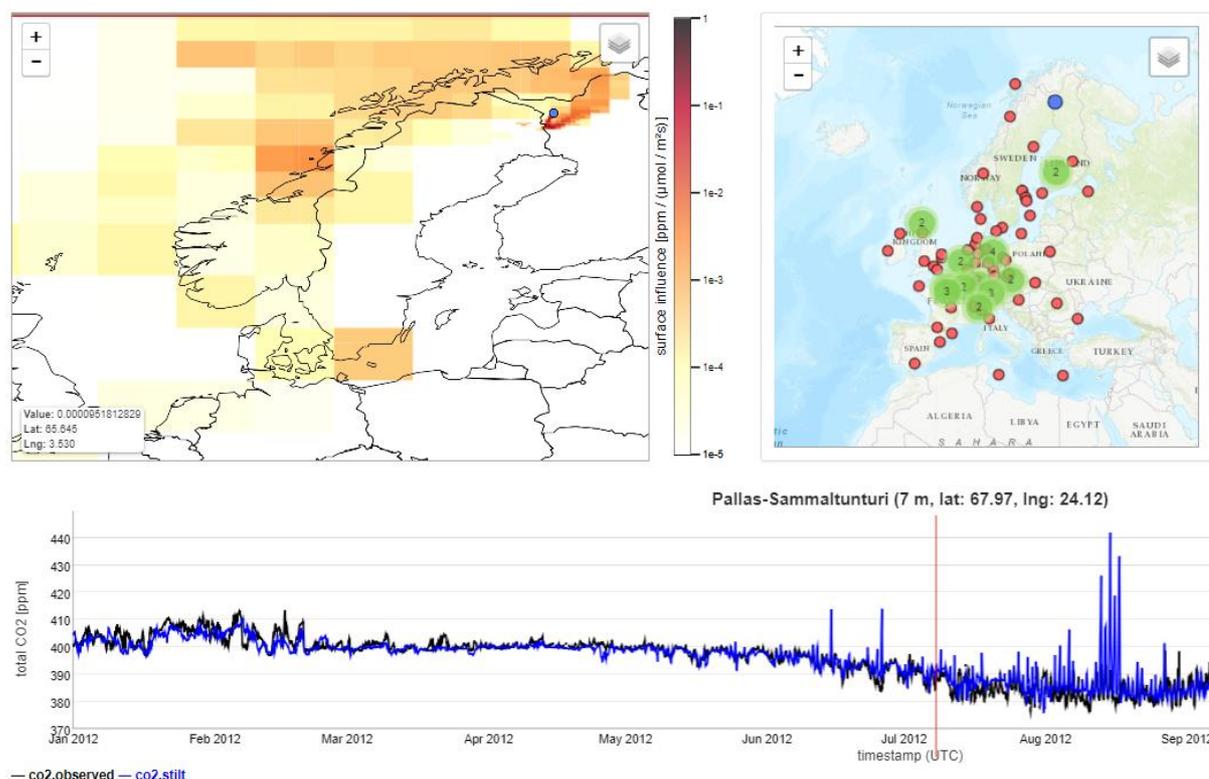


Figure 1. 2012-07-07 21:00 footprint for Pallas, Finland. The blue line represents the modelled  $\text{CO}_2$  concentrations and the black line the measured  $\text{CO}_2$  concentrations. Areas with small similar values are aggregated to the same color and make the cells appear to have different resolution.

## 2.2 Biospheric models

Biospheric fluxes mean the exchange between the terrestrial ecosystem and the atmosphere through respiration and photosynthesis. Photosynthesis is associated with uptake of  $\text{CO}_2$  from the atmosphere, whereas respiration is a source of  $\text{CO}_2$ . Their combined value is called net ecosystem exchange (NEE). There are different biospheric models designed to simulate these fluxes, and the one used in the STILT model system is the VPRM model introduced in the paper by Mahadevan et al. (2008). This model setup is also discussed in the paper by Lin et al. (2011), where an error analysis framework has been set up to attribute the model uncertainties to the different model parameters. The model is a simple, data-driven diagnostic biosphere flux model with minimum number of parameters. It has, however, shown to have relatively good performance with 60-80% of the observed hourly variability expected to be correctly simulated (Lin et al., 2011, p. 3). Two of the main parameters for estimating the NEE are MODIS satellite image derived indexes Enhanced Vegetation Index (EVI) and Land Surface Water Index (LSWI).

Put simply, the EVI index reflects the “greenness” of the surface, and is used along with shortwave radiation and light-use efficiency to estimate the gross ecosystem exchange (GEE) – reduction of CO<sub>2</sub> from the atmosphere because of photosynthesis (negative value). GEE is the terminology used in Madhadevan et al. (2008) and reflects the same process as gross primary production (GPP), but GPP is from the perspective of biospheric uptake (positive value). Light-use efficiency is calculated from temperature, leaf phenology and canopy water content. The two latter are derived from the LSWI index. Temperature is also used to calculate respiration in a function where increased temperature means increased soil and vegetation respiration (Mahadevan et al., 2008, p. 4). The data on shortwave radiation and temperature for the STILT model setup are from ECMWF.

Remaining parameters to calculate NEE are attributed based on land cover type. These parameters have been optimized using measured NEE at sites representing the different land cover types within the STILT model domain. Land cover is assigned based on SYNMAP – a 1x1 km global grid – which also have information on associated plant functional type mixtures which makes it appropriate for parametrization (Jung et al., 2006). The model operates on the spatial resolution of 1/4x1/6 degree, but the different land cover types within each grid cell are treated separately in the process of deriving the total NEE for the cell (Lin et al., 2011, p.6). The output from multiplying the VPRM model results with the footprints in the STILT model system is one value for the total uptake of CO<sub>2</sub> (GEE) and one for the total source of CO<sub>2</sub> (respiration) within a footprint.

### 2.3 Anthropogenic contribution

In order to estimate how much influence anthropogenic emissions within a footprint have on concentrations at the receptors the emissions need to be quantified in space, and preferably also in time. There are several different global datasets and in the STILT model system EDGARv4.3.2 for the base year 2010 is used (Janssens-Maenhout et al., 2019). The Carbon Portal version also includes information on the fuel mix, meaning contribution from the different fuel types oil, gas, coal and bio fuel per emission category (U. Karstens, pers. comm.) which also allows for temporal extrapolation into the year 2018 based on fuel consumption statistics on the national level (BP statistical review of World Energy, 2019). The dataset is the result of information on CO<sub>2</sub> emissions from many different sources with the main one being national inventory reports. However, only Annex I countries – including most of the developed countries – are required to submit reports on a yearly basis (Janssens-Maenhout et al., 2019, p. 960). Europe has, compared to the rest of the world, very good coverage with a couple of exceptions (Janssens-Maenhout et al., 2019, p. 961). The dataset also accounts for abatement measures, meaning any practices to remove or hinder CO<sub>2</sub> to get to the atmosphere (Janssens-Maenhout et al., 2019, p. 962).

Disaggregation of country totals spatially was done by the EDGAR team using proxies including a point source emission database, road networks and information on housing (Janssens-Maenhout et al., 2019, p. 973). For some emission categories, the emissions were rather distributed evenly in the whole country. The result is a 0.1°x0.1° raster grid with one value for each of the emission categories in each grid

cell. More details on the process of disaggregating the data in space are found in the EDGAR gridding manual (Janssens-Maenhout et al., 2013). In a study by Andres et al. (2011) the disaggregation into space was pointed out as the largest uncertainty associated with generating global emission maps based on country totals. Additional uncertainties stem from establishing the country totals in the first place: in a comparison between four different emission inventories, including EDGAR, the annual country totals differed in general around 10% (Peylin et al., 2011). Furthermore, grid cells can cross country borders but can only be associated with one country which also add to the uncertainties.

Once the yearly country totals had been distributed to a  $0.1^{\circ} \times 0.1^{\circ}$  grid, the emissions can be assumed to have been emitted at the same rate during the whole year, or rather to have changed on a daily and/or seasonal basis. In the development of the EDGAR database, the latter was true and the approach of Steinbach et al. (2011) was followed. Sector specific time profiles established in the TROTREP/POET Project (Denier van der Gon et al., 2011) were used. Each sector was assigned an emission factor ranging between zero to just over two where an emission factor of zero means no emission during that specific time of the year or hour and an emission factor of two means twice as much emissions compared to the yearly average. The time profiles used to assign emission factors have generally been derived from long time series of data which in turn were used to establish sinusoidal curves representing how emission quantities changes with time. Example of sectors with large variability during the different hours of the day are “combustion” and “road transport” (Denier van der Gon et al., 2011, p. 6). Combustion peaks during mornings and later in the evenings, which is explained by the fact that residential heating mainly happens during these hours. Road transport peaks a little later in the morning compared to combustion, and a little earlier in the late afternoon/evening in accordance with traffic rush hours. In terms of seasonal cycles, there is more combustion in the winters (Denier van der Gon et al., 2011, p. 2), mainly because of the colder temperatures in Europe during these months. Road transport is assumed to be the same the whole year. The same emission factors are used for all European countries and an example where this can be expected to cause inaccuracies is the combustion sector with very pronounced seasonal variation in winter-cold countries, but warmer countries such as Greece actually have more combustion in the summer probably because of the use of air-conditioner (Denier van der Gon et al., 2011, p. 16).

The comparison between four different emission inventories in the study by Peylin et al. (2011) also included calculating the resulting fossil fuel component by combining the emission datasets with footprints derived from different transport models. The differences in results from using different emission inventories in the same model were compared to the difference in result when also using different transport models: the transport model caused a 2-3 higher difference in resulting fossil fuel component. The STILT transport model was not included in the comparison, but the differences in result depending on what model was used indicate high uncertainties associated with footprints derived from transport models.

The emissions are provided for the different IPCC categories used when creating the national reports for the UNFCCC (Janssens-Maenhout et al., 2019, Table 1). At the Carbon Portal, they have been aggregated into the more general emission categories “Energy”, “Transport”, “Industry” and “Others” in accordance with *Table 1*.

Combined they make up anthropogenic emission's influence on concentrations at the stations and can be visualized together with the footprints on the Carbon Portal webpage. Also, each subcategory's average influence is part of the characterization of the stations.

*Table 1. EDGAR anthropogenic emission categories aggregated to the categories "Energy", "Transport", "Industry" and "Others" used at the Carbon Portal.*

<b>Carbon Portal Category</b>	<b>IPCC category (2006)</b>	<b>Meaning</b>
<b>Energy</b>	1a1a	Power industry
	1a1bcr	Refinery and transportation
<b>Transport</b>	1a3b	Road transport
	1a3ce	Non-road transport
	1a3a and 1c1	Aviation
	1a3d and 1c2	Shipping
<b>Industry</b>	1a2 and 6cd	Manufacturing
	2a	Mineral industry
	2bcefg	Chemical processes
	3	Solvents and products use
<b>Others</b>	1b2abc	Flaring and venting
	7a	Fossil fuel fires
	1a4	Energy for buildings including heating
	4f	Agricultural waste burning

## 2.4 Related studies

There are similarities between the methodologies that have been used to output the results of this thesis and other studies. In the paper by Henne et al. (2010) the authors have identified a problem in that there are many different ways, many times subjective, in which measurement stations are categorized. They suggest a non-subjective categorization into six categories based on how "representative" the measurements are: for a station to be categorized as representative the CO<sub>2</sub> concentrations in its close proximity cannot not be fluctuating. This true when the air arriving at the station has not been subjected to anthropogenic emissions and biospheric fluxes. Parameters used to establish each station's representativeness include the value referred to as a footprint "total average sensitivity" in this thesis – the result of summarizing all footprint cells – as well as the average and standard deviation of population multiplied by the footprints in a year. The station characterization values produced for the purpose of this thesis could be used to generate similar results, but keeping the results separately make it possible for each user to categorize stations based on their specific need.

In the paper by Aalto et al. (2015), the measured concentration values at the ICOS station Pallas (PAL) were considered to derive and analyze values that represent the clean background air. The process included filtering out unsuitable measurements: values were disregarded if they were more than one standard deviation higher or lower than the monthly means, as well as measurements where the wind direction was associated with transport over regions associated with more anthropogenic emissions.

If the STILT model system was rather used, it would be possible to quantify each footprint area in terms of influence from anthropogenic emissions and biospheric fluxes and use those values to subset the measurements. Sub-setting footprints for separate analysis has been practiced for the purpose of this thesis.

The process of sub-setting footprints based on their characteristics is present in the paper by Uglietti et al. (2010), as well: using a relative footprint map, high measured concentrations of CO<sub>2</sub> at Jungfraujoch were associated with air transport over Northeastern Europe. In this type of map, relative significance of each cell in the average of the subset footprints is compared to the average of all footprints. The use of relative footprint maps is exemplified in this thesis for Heidelberg (HEI) station by sub-setting footprints associated with model overestimates of CO<sub>2</sub> (see *Fig. 43*), but other characteristics can be used as well.

### 3. Methods

The work within ArcMap (ArcGIS Desktop, 10.6. ESRI, Redlands, USA) has been limited to preparation of the ancillary data to fit the footprints' cell dimensions of  $1/8^\circ$  longitude x  $1/12^\circ$  latitude in a grid with 400 columns and 480 rows making up the STILT domain. Once prepared, the data was exported to NetCDF format which can be used within the Jupyter Notebooks in combination with the footprints. The bulk of the analysis, with resulting figures and graphs, come from these Notebooks. Only the most important and GIS-relevant methods are described in section 3.4.

#### 3.1 Study area

The study area is limited to the extent of the STILT model domain covering the greater part of Europe, i.e.  $15^\circ\text{W}$ - $35^\circ\text{E}$  and  $33^\circ\text{N}$ - $73^\circ\text{N}$  (see *Fig. 3*). The extents of the ancillary datasets vary and are found in the subsections to section 3.2.

#### 3.2 Data

##### 3.2.1 Stations

The official ICOS network of measurement stations currently includes 36 atmospheric stations ("ICOS stations", n.a). For the purpose of this study, 29 ICOS official atmospheric stations have been processed as well as HEI and Freinsheim (FRE) which are not official ICOS stations: HEI is specifically for testing equipment in ICOS and FRE is a test site for ICOS measurement strategies. *Table 2* has information about the different stations including station name and code, country, location, site type and intake height. Intake height represents the elevation above ground where the measurements take place, which is also the elevation that is specified for the generation of modelled  $\text{CO}_2$  concentrations in the STILT model system. 16 of the stations have measured concentrations for year 2017 and come from two different sources: ICOS (ICOS RI, 2019) or GLOBALVIEWplus (Cooperative Global Atmospheric Data Integration Project, 2019). Which of the two is found in the column "dataset measured concentrations".

The locations of the stations are indicated in the maps below. In the *Fig. 2* map, the points representing the stations are labeled with their three-letter station codes (see *Table 2*). The purpose of the map in *Fig. 3* is to locate the stations in relation to the extent of the STILT domain.

Table 2. Stations considered for the purpose of this thesis including technical specifications. Site types: tall tower (t), ground (g), mountain (m), costal/continental (c/c), non-forested island on the sea (nfis).

station code	name	country	lat	lon	site type	intake height	dataset measured concentrations
BIR	Birkenes	Norway	58.39	8.25	g	2	No available data
CES200	Cabauw	Netherlands	51.97	4.93	t	200	No available data
CMN	Monte Cimone	Italy	44.18	10.7	m	12	No available data
FRE	Freinsheim	Germany	49.5	8.35	g	30	No available data
GAT344	Gartow	Germany	53.07	11.44	t	344	No available data
HEI	Heidelberg	Germany	49.42	8.67	g	30	GLOBALVIEWplus
HEL	Helgoland	Germany	54.18	7.9	t	15	GLOBALVIEWplus
HPB131	Hohenpeissenberg	Germany	47.8	11.01	t	131	ICOS
HTM150	Hyltemossa	Sweden	56.1	13.42	t	150	ICOS
IPR100	Ispra	Italy	45.81	8.63	t	100	No available data
JFJ	Jungfrauoch	Switzerland	46.55	7.99	m	10	ICOS
JUE	Forschungszentrum Jülich	Germany	50.91	6.41	t	100	No available data
KIT200	Karlsruhe	Germany	49.09	8.43	t	200	No available data
KRE250	Kresin	Czech Republic	49.57	15.08	t	250	ICOS
LIN099	Lindenberg	Germany	52.21	14.12	t	99	GLOBALVIEWplus
LMP	Lampedusa	Italy	35.52	12.62	g	8	No available data
LUT	Lutjewad	Netherlands	53.4	6.35	c/c	60	GLOBALVIEWplus
NOR100	Norunda	Sweden	60.09	17.48	t	101	ICOS
OPE120	OPE	France	48.56	5.5	t	120	ICOS
OXK163	Ochsenkopf	Germany	50.03	11.81	m	163	No available data
PAL	Pallas	Finland	67.97	24.12	g	5	ICOS
PRS	Plateau Rosa	Italy	45.93	7.7	m	10	GLOBALVIEWplus
PUI	Puijo	Finland	62.91	27.66	t	84	No available data
PUY	Puy de Dome	France	45.77	2.97	m	15	ICOS
SAC100	Saclay	France	48.72	2.14	t	100	No available data
SMR125	SMEAR/Hyytiälä	Finland	61.85	24.3	t	125	ICOS
STK200	Steinkimmen	Germany	53.04	8.46	t	200	No available data
SVB150	Svartberget	Sweden	64.26	19.77	t	150	ICOS
TOH147	Torfhaus	Germany	51.8	10.53	t	147	No available data
TRN180	Trainou	France	47.97	2.11	t	180	GLOBALVIEWplus
UTO	Utö	Finland	59.78	21.38	nfis	60	No available data

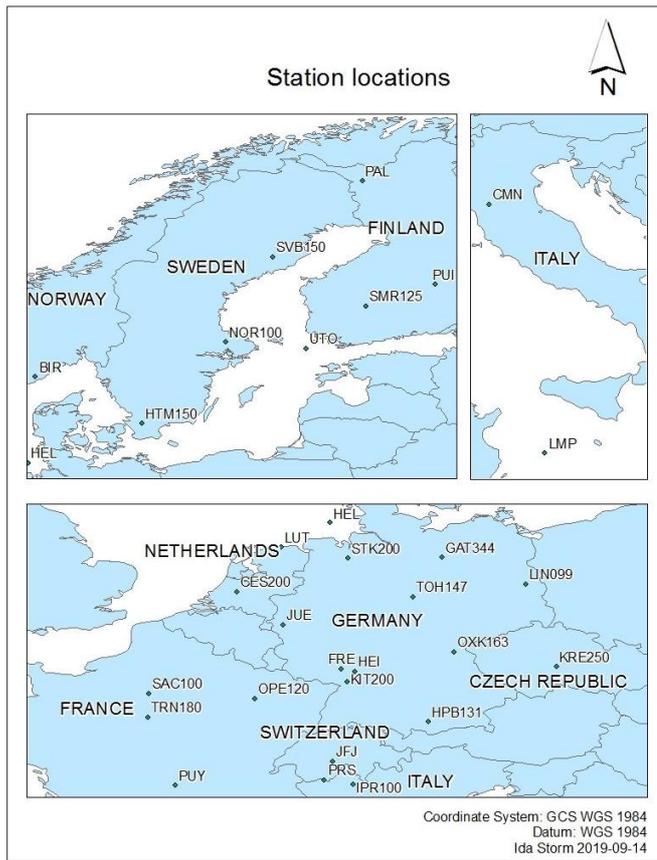


Figure 2. Station locations of the 31 stations characterized in this thesis.

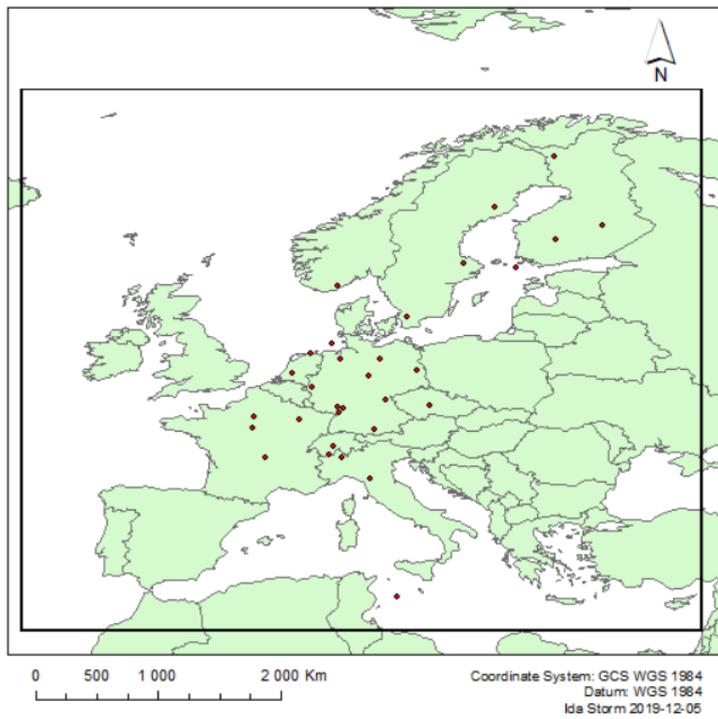


Figure 3. Station locations in relation to the extent of the STILT model domain.

### 3.2.2 Land cover

Spatial resolution: 10x10 meters

Minimum mapping unit: 0.25 km<sup>2</sup>, linear elements minimum width of 0.1 km

Spatial extent: 55.6°W-61.0°E and 72°N-24°S (see *Fig. 4*)

Data year: 2018

The land cover data represents year 2018 and comes from a European program coordinated by the European Environment Agency called CORINE land cover (CLC) vector version. The dataset is based on images from Earth Observation satellites Sentinel-2A and 2B. The images are used to generate polygons that are assigned one of 44 different land cover classes (see *Table 3*). For an area with a specific land cover class to be recognized, it needs to cover an area of at least 0.25 km<sup>2</sup> and linear elements, such as rivers, must be at least 100 meters wide. The data were processed on country level, by teams in the different countries (CLC2018 Technical Guidelines, 2017).

The 44 different land cover classes have been aggregated into 19 different classes as specified in *Table 3*. However, the 44 original classes have been processed separately in ArcMap, and imported into Jupyter Notebooks before aggregation. This means that the individual user can choose to get the breakdown of each of the 44 CORINE classes, or any other aggregation of classes.

*Table 3. CORINE land cover classes and their corresponding aggregated classes.*

CORINE class	Aggregated class	CORINE class	Aggregated class
Continuous urban fabric	Urban	Vineyards	Cropland: fruits, berries, grapes and olives
Discontinuous urban fabric		Fruit trees and berry plantations	
Industrial or commercial units		Olive groves	
Road and rail networks and associated land		Pastures	Pastures
Port areas		Broad-leaved forest	Broad-leaved forest
Airports		Coniferous forest	Coniferous forest
Mineral extraction sites		Mixed forest	Mixed forest
Dump sites		Agro-forestry areas	Transitional woodlands and shrub
Construction sites		Sclerophyllous vegetation	
Green urban areas		Transitional woodland-shrub	
Sport and leisure facilities		Beaches, dunes, sands	Natural bare areas
Industrial or commercial units		Bare rocks	
Mineral extraction sites	Sparsely vegetated areas		
Construction sites	Industrial	Burnt areas	

Road and rail networks and associated land	Road and railroads	Glaciers and perpetual snow	Glaciers and perpetual snow
Port areas	Ports and airports	Inland marshes	Wet areas
Airports		Peat bogs	
Dump sites	Dump sites	Salt marshes	
Non-irrigated arable land	Cropland: staple except rice	Salines	
Permanently irrigated land		Intertidal flats	Inland water bodies
Annual crops associated with permanent crops		Water courses	
Complex cultivation patterns		Water bodies	
Land principally occupied by agriculture, with significant areas of natural vegetation		Coastal lagoons	
Rice fields	Rice fields	Estuaries	
Natural grasslands	Natural grasslands	Ocean	Ocean
Moors and heathland			

The CORINE land cover extent does not cover the whole STILT domain, which leaves areas with no land cover data. These areas include Belarus, most of Ukraine, and the most westerly part of Russia. In the south, parts of northern Africa are missing (see *Fig. 4*). Another aspect of the CORINE dataset is that only the water within the countries exclusive economic zones (370 km from the coastal baseline) are included. The remaining part of the oceans have been complemented by data from Natural Earth. Remaining areas beyond the land cover extent are quantified and referred to as “no data”.

To include another land cover dataset to cover the missing areas was considered, but the no-data class has remained small for most of the stations except Lampedusa (LMP) and Puijo (PUI) which are the stations located closest to the edges of the CORINE domain (see *Fig. 3*). Their averages for year 2017 were 24.1% for LMP and 14.1% for PUI. For all other stations, the yearly average of the no-data class varied between 0.9 to 8.9% with an average for all stations of 3.8%. For year 2016 the corresponding values were 0.9 to 8.8% with an average for all stations of 4.4%.



*Figure 4. The STILT domain compared to the CORINE land cover. Oceans have been complemented by data from Natural Earth, but area within the STILT domain beyond the land cover extent are referred to as “No CORINE data” in the legend and are quantified as “no data” in the land cover breakdowns (see section 4.3.1).*

### 3.2.3 Population data

Spatial resolution: 1x1 km

Spatial extent: All European countries except Russia, Bosnia and Herzegovina, Montenegro and Serbia

Data year: 2011

The GEOSTAT population grid is the result of a common effort of the different countries it covers. The process of aggregating population counts to 1x1 km cells looks different depending on the population data source of the different countries. For some, population counts were associated with buildings which in turn could be connected to point locations. Next, the points that fell in each grid cell had their values summarized. In other cases, the population data were associated with census enumeration areas. These were disaggregated to the cells based on ancillary data that indicate the spatial distribution of population within the census areas such as night time lights, road networks and buildings (GEOSTAT 1A – Representing Census data in a European population grid, 2011). Since the census areas are usually small this make for good data precision – sufficient for the purpose of this thesis where the population data has been further aggregated into a  $1/12^\circ$  latitude  $\times$   $1/8^\circ$  longitude grid.

The GEOSTAT population grid consist of a reference grid with the population data for each grid cell in a separate table. The reference grid's cells each have a unique value which was used as the "key" in the process of joining the population data ("join by attribute" in ArcMap).

#### 3.2.4 Point source data

Spatial resolution: Point location for each facility

Spatial extent: All EU members and Iceland, Liechtenstein, Norway, Switzerland and Serbia

Data year: 2017

All European Union members are required to annually submit information on emissions from industries with the location of the individual facilities and their emissions to the database E-PRTR. There are 91 different substances that need to be reported on, including emissions into air, water and land. For the purpose of this thesis, CO<sub>2</sub> emissions into air was considered.

The data is downloaded as a database which has three tables necessary to establish the point source emissions of CO<sub>2</sub> for year 2017. There is one table with the facility names and their spatial locations specified by latitude and longitude values. The quantities of emissions are stored in a separate table but share a common field with the table containing the facilities which facilitates a join between the two ("join by attribute", ArcMap). Before the join, only the emission quantities of CO<sub>2</sub> were selected ("select by attribute" > "export selection", ArcMap). The third table includes information on which year's report the emissions are associated with, and was joined to the table with facilities and emission quantities. In the resulting table, year 2017 was selected and the final table was exported ("select by attribute" > "export selection", ArcMap). The latitude and longitude values in the final table were used to add the data to the map ("add XY data", ArcMap). It resulted in 2172 facilities, by which 2149 are within the extent of the STILT domain and are therefore included in the processing described section 3.3.

#### 3.2.5 Radiocarbon (<sup>14</sup>CO<sub>2</sub>)

Spatial resolution: Point location for each facility

Spatial extent: All nuclear power plants and fuel reprocessing stations with radiocarbon emissions within the STILT domain (67 locations)

Data year: 2016

Data on radiocarbon (<sup>14</sup>CO<sub>2</sub>) emissions from nuclear power plants and fuel reprocessing stations come from the values used in the paper by Zazzeri et al. (2018). The <sup>14</sup>CO<sub>2</sub> emissions values have been estimated based on the amount of electrical power produced during the year because measurements of <sup>14</sup>CO<sub>2</sub> emissions are generally not available (Zazzeri et al., 2018, p. 1067). For all facilities, except those with pressurized water reactors (PWRs), all the <sup>14</sup>C are emitted as <sup>14</sup>CO<sub>2</sub>. For PWRs, it is estimated to be 32% of the <sup>14</sup>C emissions. 2016 is the final year of data in Zazzeri's supplemental table and the spatial distribution of the facilities located within the STILT domain and their associated emission quantities are shown in Fig. 5. For more recent years, <sup>14</sup>C values reported to the European Commission RAdioactive

Discharges Database (RADD) – again normally calculated as a function of power production – can be used but would first need to be adjusted in case of emissions from PWRs.

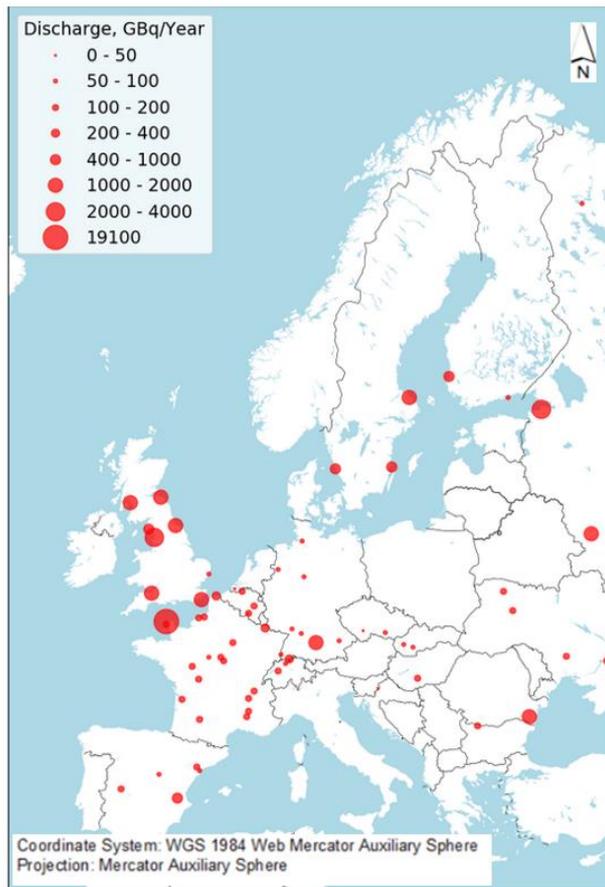


Figure 5. Radiocarbon ( $^{14}\text{CO}_2$ ) emissions within the STILT domain from nuclear power plants and fuel reprocessing stations year 2016

### 3.3 Data processing to match the footprint data

All ancillary data layers need to be quantified to the same resolution and extent as the footprints and stored in the same format, NetCDF. The process looks slightly different depending on if the ancillary data is saved as points, such as for the population and point source emissions, or polygons, which is true for the land cover. Both processes include the use of a polygon shapefile representing the STILT domain which was created using the ArcMap tool “create fishnet” with the STILT domain’s extent and cell size as input parameters.

#### 3.3.1 Point data

The point data values were aggregated to the different fishnet cells with a “spatial join” (ArcMap), choosing “sum” as the statistical method. The output is another fishnet with the STILT domain’s extent and cell size, but with the needed sum-values for each of the cells. The ArcMap tool used to export the data as a NetCDF file, “feature to NetCDF”, needs point data as input. Therefore, the fishnet was exported to a point layer (“feature to point”, ArcMap) where each output point represents a

fishnet cell's midpoint before the export to NetCDF could be completed. The resulting files were uploaded to the ICOS CP-server, which meant the data layers were ready to be used within the Notebooks.

### 3.3.2 Polygon data

The land cover data consist of polygons and polygons crossing cell borders of the STILT domain's cells complicates the process: first, one land cover at a time was selected using "select by attribute" (ArcMap). Next, the selected polygons were dissolved into one multi-polygon using "dissolve" (ArcMap). Thirdly, the multi-polygon was divided into one – or zero – polygon for each of the STILT cells. From these polygons, one area value for each STILT cell, for the specific land cover class, was calculated. The area was calculated given the projection Lambert azimuthal equal-area Europe (ESPG:3035). The choice of projection – how 3D locations are translated to a 2D surface – will affect map properties such as area and shape. In an equal-area projection preserving the area of the polygons are prioritized. A model was built using ArcMap's "model builder" to expedite the different steps and enable the data processing to run without supervision (Fig. 6). The output area values were joined ("join by attribute", ArcMap) to the midpoint representations ("feature to point", ArcMap) of the fishnet. The common field for the join was the fishnet cell-id, which had been assigned to each polygon-cell-representation during the intersection with the fishnet. Before the midpoint shapefile was converted to a final NetCDF file, all land cover classes were joined to the midpoint layer. For the ocean class, the values joined to the midpoint representations were complemented with data from Natural Earth (see section 3.2.2). This was archived by using the tool "erase" to crop out and disregard the area representing oceans within the CORINE land cover data from the Natural Earth ocean. The remaining Natural Earth oceans were treated as any other land cover class. However, once joined to the midpoint representation, the area values were added to those of the CORINE ocean land cover class to create one ocean value per cell.

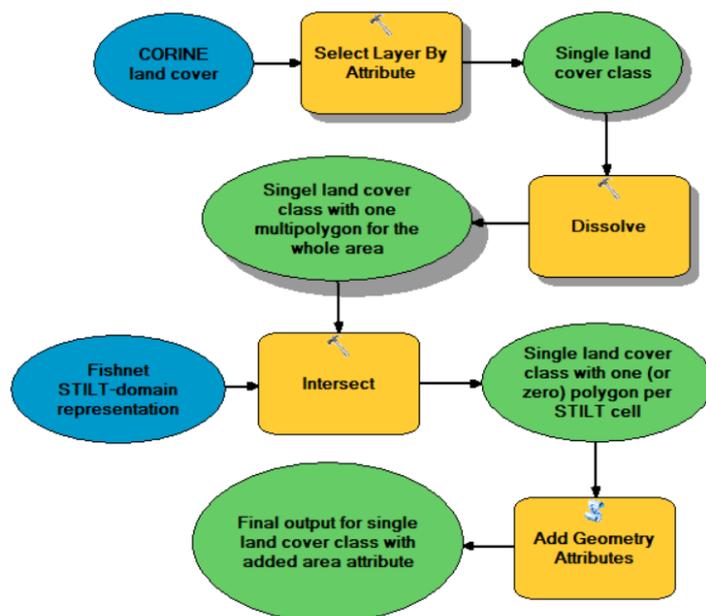


Figure 6. Model to process each land cover class, created with "model builder" in ArcMap.

## 3.4 Data analysis

### 3.4.1 Jupyter Notebooks

The bulk of the analyzes have taken place within Jupyter Notebooks. This is also the format of the output tools that will be available for users though the Carbon Portal. These are currently only available as HTML files which have been published on LUP Student Papers. Links to these are found in Appendix 1. A Notebook consists of cells with code where data is processed followed by the output from “running” the cells when applicable (see Appendix 1). The language used in all code cells is Python 3. The code cells are complemented by cells with markdown language which consist of text and can also have links to different parts of the Notebook.

Modelled and measured concentrations of CO<sub>2</sub> are accessed from the Carbon Portal’s server on demand within the Notebooks using date stamps which make it possible to access specific date-ranges, with the option to also filter out data from certain times of the day.

### 3.4.2 Sensitivity area evaluation

Footprints derived from the STILT transport model are used throughout this thesis and the “sensitivity area evaluation” Notebook makes it possible to examine their spatial characteristics. First, the user is able to define distances within which all the footprint cells are summarized. This gives a general idea of how extensive the area of sensitivity to potential surface fluxes is. For this, the first step was to locate all the stations in what have been referred to as the “fishnet STILT-cell representation” (see section 3.3). A spatial join (“join by location”, ArcMap) between the stations’ point locations and the fishnet was used to assign one latitude and one longitude value for each of the 31 stations. This data, along with the names of the stations, were put in an Excel file and uploaded to the server to be accessed within the Notebook. When a specific station is processed in the Notebook, the station name is matched with the station name in the Excel file, and in turn the corresponding latitude and longitude values are accessed. Distances between two locations can be calculated if the latitude and longitude values are known. There are different ways to do this, and this Notebook uses equirectangular approximation:

$$x = \Delta\lambda * \cos\varphi_m$$

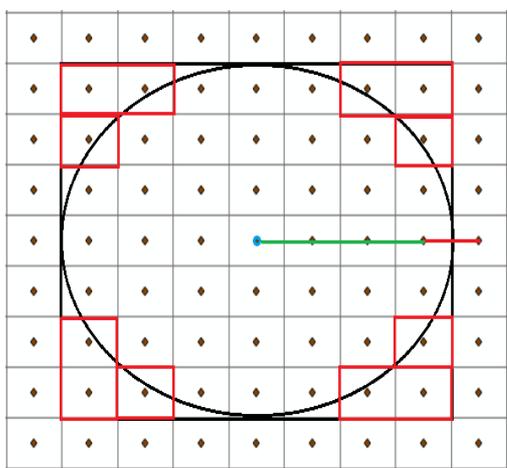
$$y = \Delta\varphi$$

$$d = R * \sqrt{x^2 + y^2}$$

Where  $\varphi$  is latitude in degrees,  
 $\varphi_m$  is the average latitude value,  $((\varphi_1+\varphi_2)/2)$  in decimal degrees,  
 $\lambda$  is longitude in decimal degrees,  
 $R=6378.8$  is the approximate radius of the earth in km

It would be possible to use equirectangular approximation to calculate the distances between each station and the midpoints of all 192000 cells in the STILT domain, but it would be inefficient. Rather, two loops were set up to first define a square to limit

the number of cells to include in the next step. One of the loops calculates the distances to further and further away cells in longitude direction, and the other loop to the cells in latitude direction (see Fig. 7). The user defined distance is used as a threshold and when it is surpassed the loops stop. The row and the column values of the final cells are saved and used to define a square with cells that are potentially within the user defined distance. Distances to all the cells within the square are in turn calculated and if a distance is below the threshold the cell value is added to the sum. The final value when all the cells have been considered is the referred to as “the absolute sensitivity” within the user defined distance. These values are also compared to the sum of all footprint cells to generate a percentage of sensitivity within the distance. For a given distance, the number of cells to loop over in the latitude direction are the same for all stations. However, the length of one longitude degree changes with decreasing width moving north from the equator and means that the number of cells in longitude direction need to be calculated for each of the stations.



*Figure 7. For illustration of the method. The point in the middle represents the station cell location within the STILT domain. First, the square is defined by finding the number of cells in longitude (green line) and latitude directions that are within the user defined distance. Next, the distances to each of the cells within the square are calculated. If the distance to a cell is within the user defined distance, the cell value is added to the sum.*

The results can be presented in a bar graph showing the average influence within three user-defined distances at one time, sorted by the total average sensitivity if many stations have been selected. Another possibility is to visualize each of the summed values within a certain distance, or sensitivity to the whole domain, in a time series. Each value can also be sorted by hour and by month to see if that reveals any patterns in differing averages.

One final aspect considered regarding the footprint themselves is sensitivity to the outermost cells of the STILT domain. This can be used as an indicator of how much of the air arriving a station has come from outside the STILT model domain in longitude or latitude extent. Beyond the STILT domain sensitivity to potential fluxes has not been calculated and can mean poorer model estimates. The outermost cells are defined as the columns and rows with the lowest and the highest values, and only the footprint cells of these are summarized. Stations can be sorted based on these values and be presented in a bar graph.

### 3.4.3 Breakdown of modelled concentration

Modelled CO<sub>2</sub> concentration values with breakdown into the sub-categories “industry”, “energy”, “transport”, “industry” and “others” for anthropogenic emissions (see *Table 1*) and respiration and GEE in terms of biospheric fluxes are accessed from the carbon portal server on demand in the Notebook. Averages based on a specific date range can be visualized in bar graphs and it is possible to compare the values of two different date ranges in the same graph. Furthermore, averages for the stations can also be compared in bar graphs with the option to rank from smallest to largest contribution from anthropogenic emissions or biospheric fluxes.

### 3.4.4 Ancillary data

The ancillary data all have the same type of processing within the Notebook: each specific date and time has a footprint which is multiplied by the corresponding cells' values of the pre-processed ancillary data (see section 3.3). The resulting cell values are in turn summarized. How the result of this should be interpreted vary for the different ancillary datasets: for point source emissions the original unit of CO<sub>2</sub>/year/footprint cell was translated into  $\mu\text{mol}/\text{m}^2\text{s}$  and multiplication with the footprints – with the unit  $\text{ppm}/(\mu\text{mol}/\text{m}^2\text{s})$  – results in an expected shift in CO<sub>2</sub> concentration at the station in ppm. For radiocarbon, the emission quantities radioactivity in Bq/year were translated into Bq/m<sup>2</sup>s and the result from the multiplication with the footprint was divided by the radioactivity in the air given the current modelled concentration of CO<sub>2</sub> at the station to arrive at an expected shift in <sup>14</sup>C. For population, the results are relevant for compared among the stations and the footprints at an individual station, but is not translated into an expected shift in the CO<sub>2</sub> concentrations at the stations. The same is true for the land cover data, but the fact that each cell is entirely covered by the different land cover classes (except for “no data”, which also is quantified) makes it possible to also translate the summed values of each land cover type into a percent value.

The output for the point source and population data in the Notebook are bar graphs with the station values calculated for a defined date range, again with the option to rank from smallest to largest. There is also an option to see the individual cell's contribution in maps, which is also possible for the individual land cover classes. Stations can also be ranked among themselves for each of the individual land cover classes in the same way as the point source and population data. Each station can also get a bar graph with one axis showing the percent breakdown of each of the land cover classes and the other showing the summed, “absolute”, values.

### 3.4.5 Modelled concentration compared to measured concentration

Model data differences – the resulting values from subtracting measured concentration values from corresponding modelled concentration values – are closer analyzed in this Notebook. A negative model data difference means that the model has underestimated the CO<sub>2</sub> concentration for that time-step with associated footprint, and a positive value means an overestimate. The model data differences can be visualized in a line graph, with an average model data difference and the standard deviation of the model data differences for a given station and date range. Furthermore, the values can be sorted by hour and by month to examine if there are

any temporal trends in model performance. The sorted data can be displayed in different ways: by showing the average model data difference for the given hour or month, or by showing a count of how many times the model data differences are above or below a certain threshold such as the standard deviation of the model data differences. The loop that considers if the model data difference is above or below a threshold can also be used to subset the footprints and run separate analyzes in any of the Notebooks on the subset - or use it to generate relative footprint maps (see section 2.4).



## 4. Results

The result section presents the most interesting outputs from running the Notebooks. Generally, annual averages for all stations are presented in bar graphs to give an overview of the output. These include contribution to the modelled concentrations divided into a biospheric component and an anthropogenic contribution, sensitivity within certain distances of the stations, sensitivity to population, contribution to the CO<sub>2</sub> concentrations from point source emissions and shift in <sup>14</sup>C due to radiocarbon (<sup>14</sup>CO<sub>2</sub>) emissions. It is more difficult to present the land cover data for all stations given the large number of land cover classes. Sensitivity to coniferous forest is used to exemplify what can be done for all land cover classes. Closer analyzes of the results are presented for individual stations such as time series showing how the total sensitivity values change over time and similar representations showing influence from radiocarbon emissions.

### 4.1 Sensitivity area evaluation

There are large differences between stations in terms of how much of the air simulated to arrive at the station stations have come from within the STILT domain. The height of the bars in *Fig. 8* represent each station's relative sensitivity to the model domain year 2017. The different colors of the bars indicate how big percentage of the sensitivity is to the area within 100 (black), 200 (dark grey) and 300 (lighter grey) kilometers of the stations. In general, atmospheric stations' sensitivities extend far: the average sensitivity for all stations to the area within 300 km is 48%, ranging from lowest average for Kresin (KRE) with 18% and the highest average for PAL with 77%. Stations with relatively high total average sensitivity also generally have high sensitivity close to the stations, especially within the first 100 km: stations LMP, PAL, CMN, IPR, HEI and FRE – the six stations with the highest total average sensitivity year 2017 – have over 50% of its sensitivity to this area. It should be noted again that there is a difference between the percent and absolute sensitivity values: the percentages within 100, 200, or 300 kilometers are calculated based on the relative sensitivity compared to the total average sensitivity which varies for the stations. Mountain stations generally have both a small total average sensitivity, and relatively more sensitivity further away. The three stations with the lowest total average sensitivity values – Plateau Rosa (PRS), Jungfrauoch (JFJ) and Puy de Dome (PUY) – are all mountain stations. Mountain station Monte Cimone (CMN), with the third largest total average sensitivity value, is an exception.

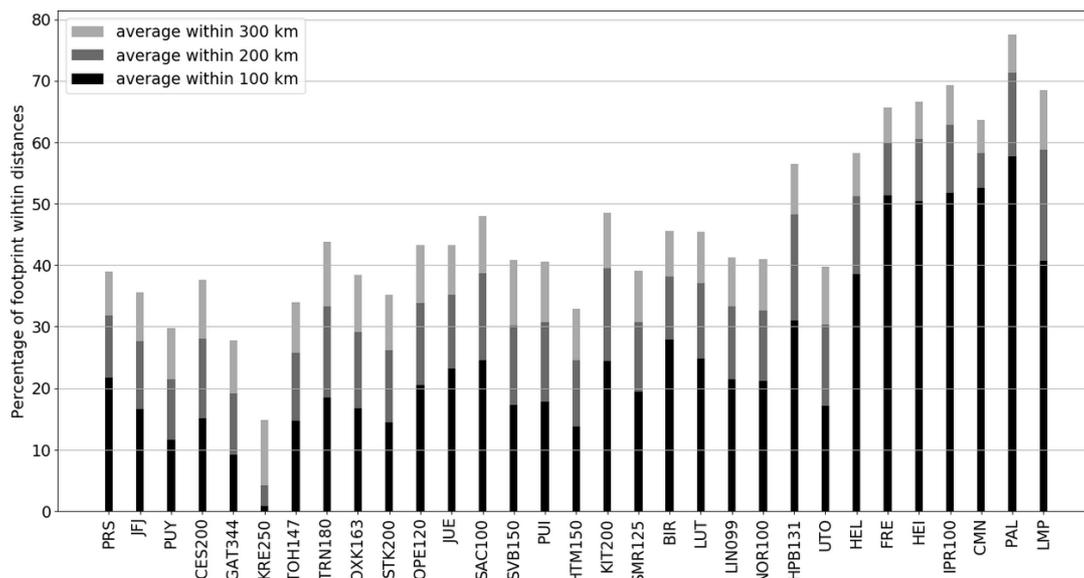


Figure 8. Estimated sensitivity within the model domain year 2017. Sorted lowest to highest total average sensitivity.

Fig. 8 gives a general indication of the spatial characteristics of the footprints at the different stations, but there are also large variabilities between the individual footprints of the stations, both in terms of the total average sensitivity as well as the distribution of the sensitivity within the domain. In Fig. 9 this is exemplified with a time series showing how the total average sensitivity changed over time at Hyltemossa (HTM) year 2017. The total average sensitivity for HTM was  $4.74 \text{ ppm} / (\mu\text{mol} / \text{m}^2\text{s})$  with a standard deviation of 2.49. The difference from one footprint to the next, separated by three hours, can be huge. The average standard deviation for all stations is higher than at HTM with  $3.70 \text{ ppm} / (\mu\text{mol} / \text{m}^2\text{s})$ , ranging from 1.89 at PUY to 12.03 at LMP.

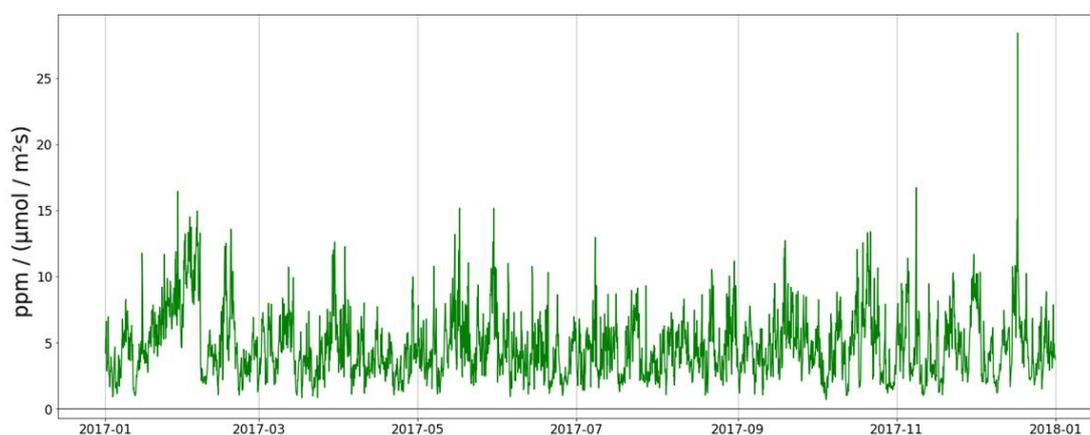


Figure 9. Timeline showing the total sensitivity value for the individual footprints at Hyltemossa year 2017. Average was  $4.74 \text{ ppm} / (\mu\text{mol} / \text{m}^2\text{s})$  with a standard deviation of 2.49.

It is possible to examine trends in the data by aggregating the total average sensitivity by month and by hour. The results from combining the aggregated values for all stations are shown in Fig. 10 and Fig. 11. Two different years, 2016 and 2017, were included to see if the years show similar results. It should be noted that, especially in

terms of monthly variation, there were sometimes opposite trends: with large sensitivity values in winter compared to the summer for PAL and Ispra (IPR), and the opposite being true for LMP, the total averages displayed in the *Fig. 10* are somewhat averaged out. Still, January and February had significantly higher averages whereas April through August had lower. Sensitivity to the area within 100 km of the stations were also considered and generally agrees well with the trends of the total average sensitivity values.

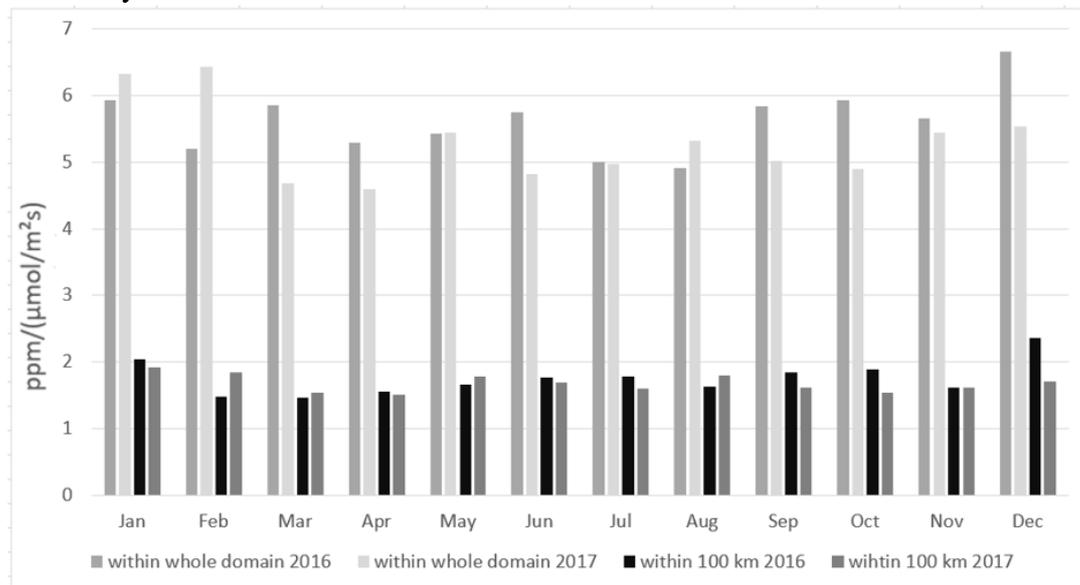


Figure 10. Average sensitivity by month, years 2016 and 2017.

Hourly variations in sensitivity were more pronounced than monthly with higher values during night time (21:00, 0:00, 3:00, 6:00) and peaks at 0:00 and 3:00. From 6:00 to 15:00 the sensitivity decreased and started increasing again at 18:00 (see *Fig. 11*). The sensitivity in was in general higher in year 2016 compared to year 2017 which is due to differences in transports to the stations between the two years. Again, the sensitivities to the area within 100 km were considered and seem to follow the trends of the average sensitivity.

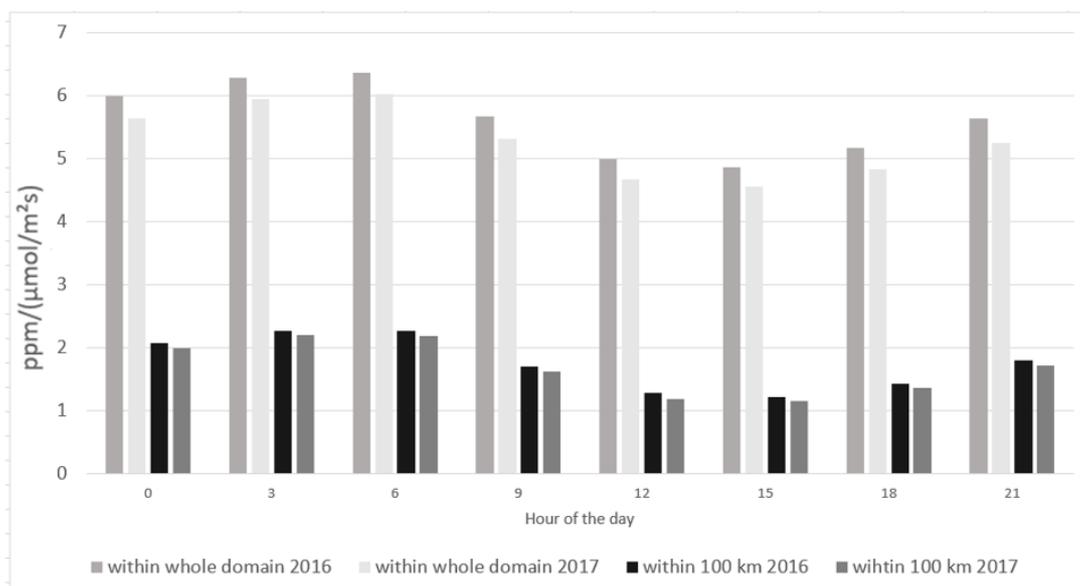


Figure 11. Average sensitivity by hour, years 2016 and 2017.

One final aspect considered in terms of the sensitivity is how large part of the total average sensitivity is to the edge of the STILT domain. It gives an indication of how much of the air arriving at the stations have arrived from outside the latitude and longitude extent of the domain (see section 3.4.2). The resulting values in *Fig. 12* are generally low with relevance especially for inter-comparison between stations. Stations close to the edge, such as LMP, SVB, PAL and PUI, of the domain had higher values. It should be noted that air arriving at the tall stations from higher vertical levels than the lower half of the PBL cannot be quantified.

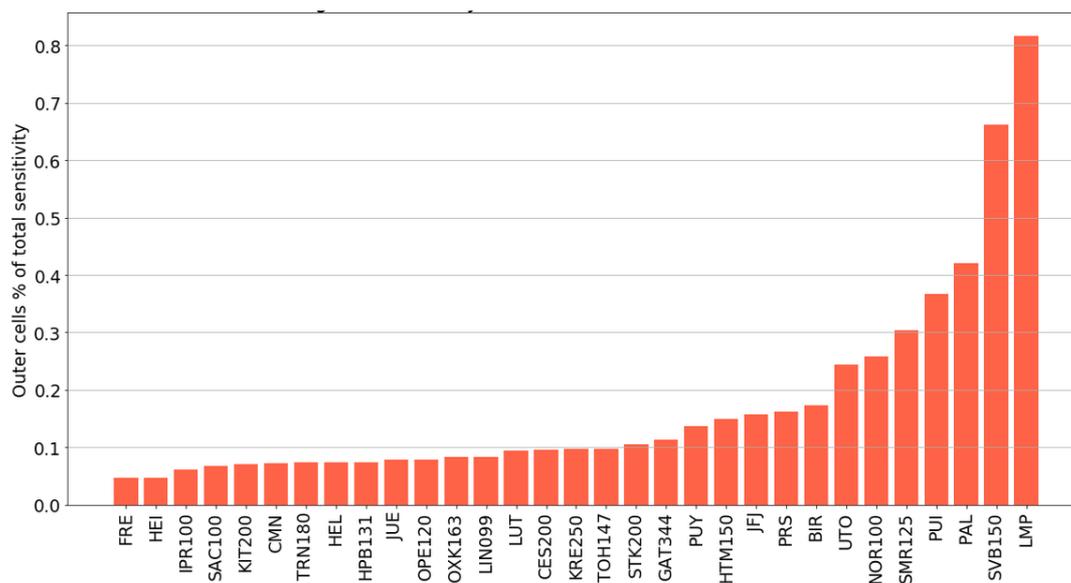


Figure 12. Percent sensitivity to outermost rows and columns of cells for year 2017.

#### 4.2 Breakdown modelled concentration

In *Fig. 13* the station averages in terms of anthropogenic emissions and the biospheric component for year 2017 are shown. The blue bars represent their combined influence on the CO<sub>2</sub> concentrations at the stations. Most stations had an average anthropogenic contribution well under five ppm with the lowest values for the remote stations PAL and Svartberget (SVB) which are located in northern Finland and Sweden respectively. The anthropogenic contribution to the modelled CO<sub>2</sub> concentrations was on average 0.38 ppm at PAL which was only 2.24% of the estimated contribution at Heidelberg (HEI) – the station with the highest average in 2017. The ranking in terms of anthropogenic contribution among the stations varies slightly between years. Comparing years 2016 and 2017, there was a 7.60% lower contribution at LMP, and 25.44 % higher contribution at PAL year 2017. Considering all stations, the contribution was 7.27% higher year 2017. PAL and SVB had the lowest contributions in both years, but in terms of highest contribution station Jülich (JUE) was ahead of HEI year 2016.

Concerning the biospheric component, the photosynthetic uptake generally outweighs the respiration at the annual timescale. All stations considered, the biospheric component was on average a sink of 0.58 ppm year 2017 and 0.64 ppm year 2016. The values ranged from a sink of 2.80 ppm (2.89 ppm year 2016) at Hohenpeissenberg (HPB) to surprisingly being a large source rather than sink at CMN

with 4.64 ppm (5.06 ppm year 2016). This may reflect problems to correctly assess the biospheric component with the VMRM model.

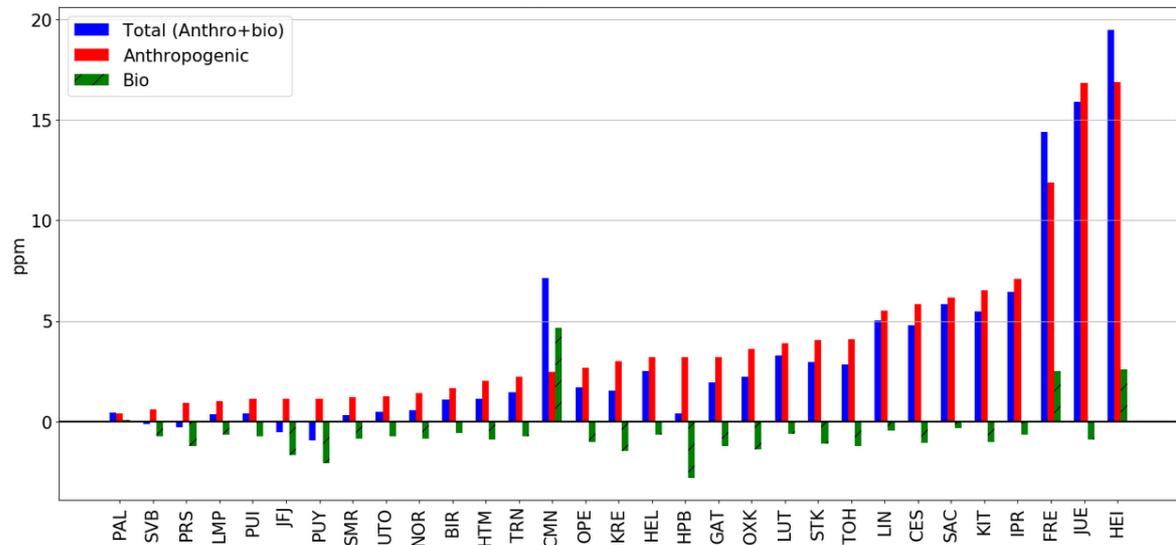


Figure 13. Average anthropogenic contribution and biospheric component for all stations year 2017

The anthropogenic emission can be broken down to different source categories (see Table 1). Their relative contributions for year 2017 are shown in Fig. 14. The category “energy” was the largest for most of the stations year 2017. Exceptions were Trainou (TRN), HPB, Saclay (SAC) and IPR where the “others” category was the largest and at HPB and SAC “energy” had the smallest contribution out of the four categories.

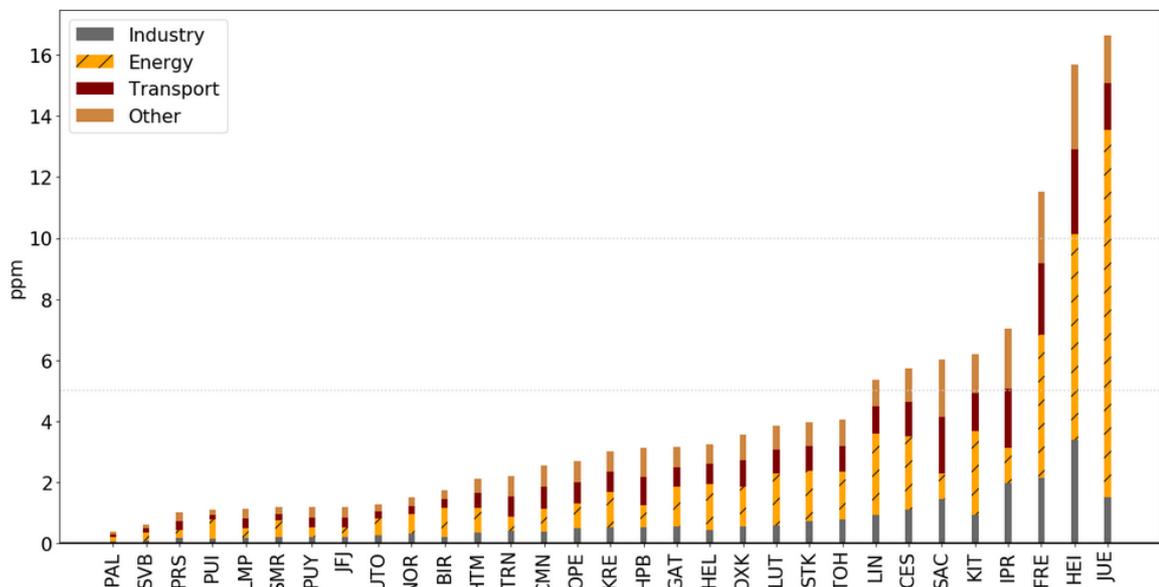


Figure 14. Average anthropogenic contribution broken down into source categories, all stations year 2017

In Fig. 15 the differences between winter (December 2016-February 2017) and summer (June 2017-August 2017) are examined. For all stations except PRS, LMP and JFJ the anthropogenic contributions were larger in the winter. Considering all

stations, the average anthropogenic contribution was just over twice as large in the winter with an average of 6.14 ppm compared to 2.93 ppm in the summer. The biggest percentage difference was at TRN with only 25% anthropogenic contribution in the summer compared to the winter. The trend of more emissions in the winter months is in accordance with the average monthly anthropogenic emissions within the European domain displayed as a time series for years 2006 to 2018 in *Fig. 16*. During the colder winters there are more emissions from for instance residential heating, which is part of the “others” category. The changes in breakdown of the different source categories between summer and winter are exemplified for HTM in *Fig. 17*. HTM is a station in southern Sweden, and whereas there was an increased contribution from all source categories in the winter, the “others” and “energy” categories especially increased. Considering the differences in the biospheric component, it was on average a source of 1.70 ppm in the winter, and a sink of 3.63 ppm in the summer. The biggest difference was seen at HPB where the biospheric component was a source of 1.64 ppm in the winter, and an 8.54 ppm sink in the summer. At CMN, where the biospheric component was a source on average both years 2016 and 2017, the biospheric component was less of a source in the summer (1.91 ppm) compared to the winter (6.94 ppm).

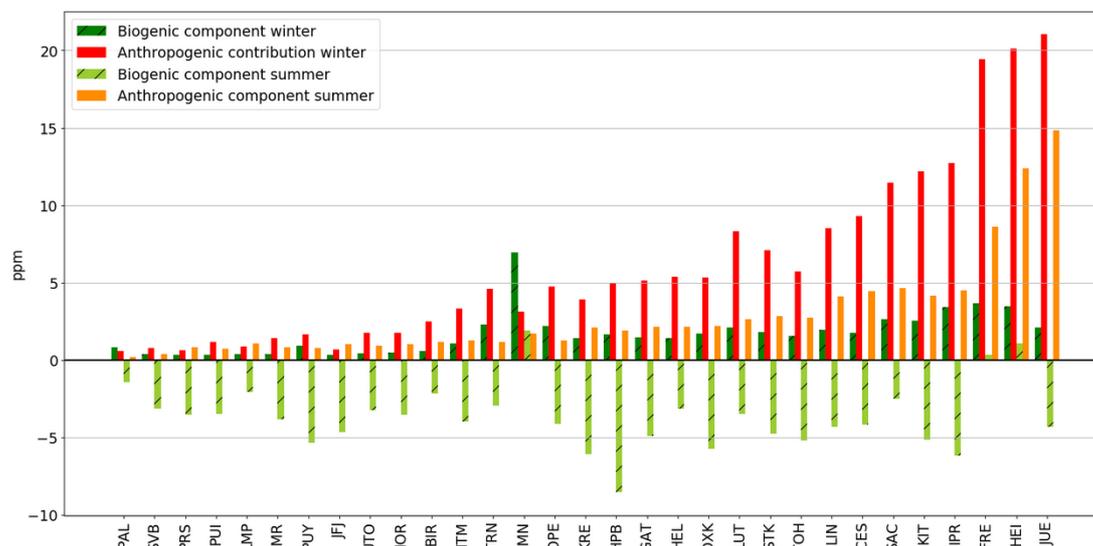


Figure 15. Anthropogenic contribution and the biospheric component winter (Dec 2016-Feb 2017) vs. summer (Jun 2017-Aug 2017). Sorted from lowest to highest anthropogenic contribution year 2017.

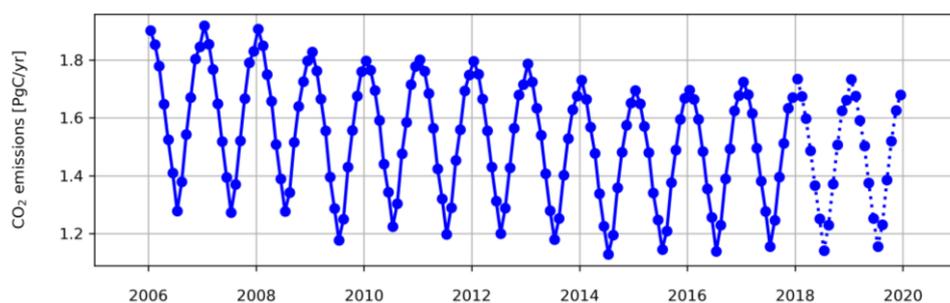


Figure 16. Monthly variation in total anthropogenic emission for Europe years 2016-2020 (projected) in total anthropogenic emission for Europe. Figure was provided by U. Karstens.

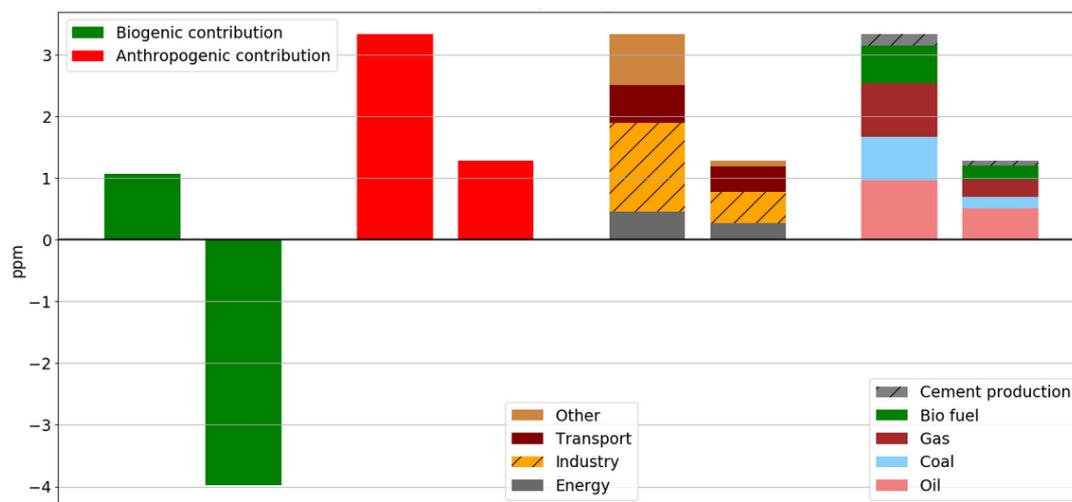


Figure 17. Anthropogenic contribution and the biospheric component winter (Dec 2016-Feb 2017) vs. summer (Jun 2017-Aug 2017) with the anthropogenic contribution broken down into the different source- and fuel categories for Hyltemossa.

There are more anthropogenic emissions in the winter months compared to the summer month, but another difference that influences the anthropogenic contribution is a general increase in sensitivity in the winter (see section 4.1). The relationship between sensitivity and anthropogenic emissions is explored in Fig. 18. Winter values are plotted relative to summer values where 100% means the same average in the winter as in the summer and above 100% means a higher average in the winter. Stations that had significantly lower anthropogenic contribution in the summer compared to the winter had significantly less sensitivity in the winter (25, 48 and 51% sensitivity for LMP, JFJ and PRS respectively). The relationship between the two has been tested for correlation in a regression analysis with the anthropogenic emission values dependent on the sensitivity values. The resulting  $R^2$  value suggest that 79% of the variation in anthropogenic emissions could be explained by the sensitivity values at JFJ, 76% at PRS and 34% at LMP. Hence, for JFJ and PRS the relatively low sensitivity in winter compared to the summer was likely the main reason for the higher anthropogenic contribution values in the summer. Other stations that had lower average sensitivity in the winter compared to the summer was Utö (UTO), PUY and Helgoland (HEL), but they still had significantly higher anthropogenic contribution in the winter.

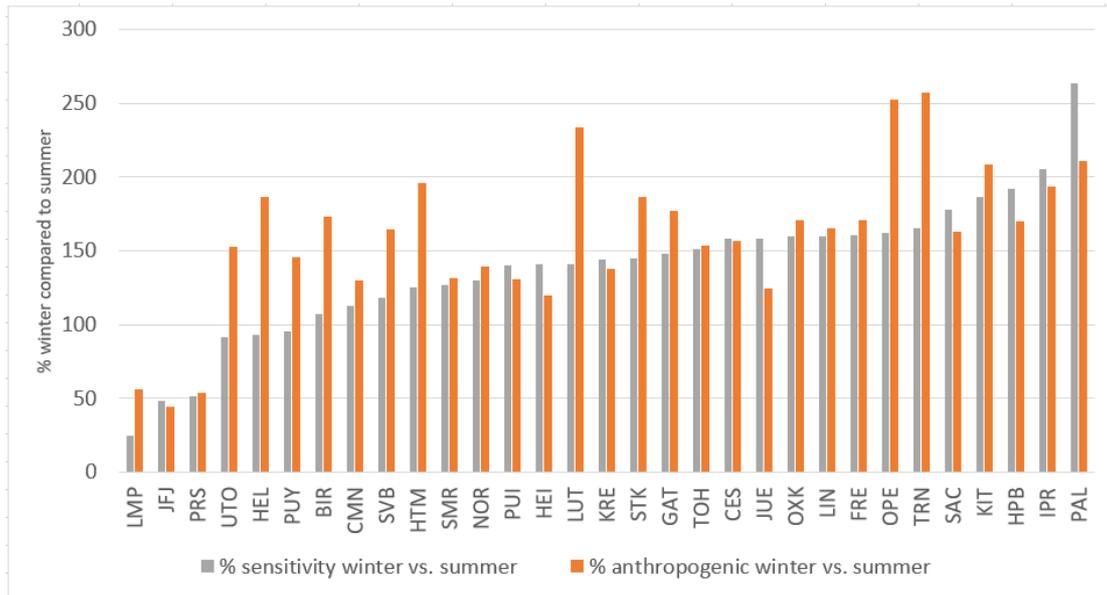


Figure 18. Anthropogenic contribution difference between winter ((Dec 2016-Feb 2017) and summer (Jun 2017-Aug 2017) compared to difference in sensitivity for the same time periods.

### 4.3 Ancillary data

#### 4.3.1 Land cover

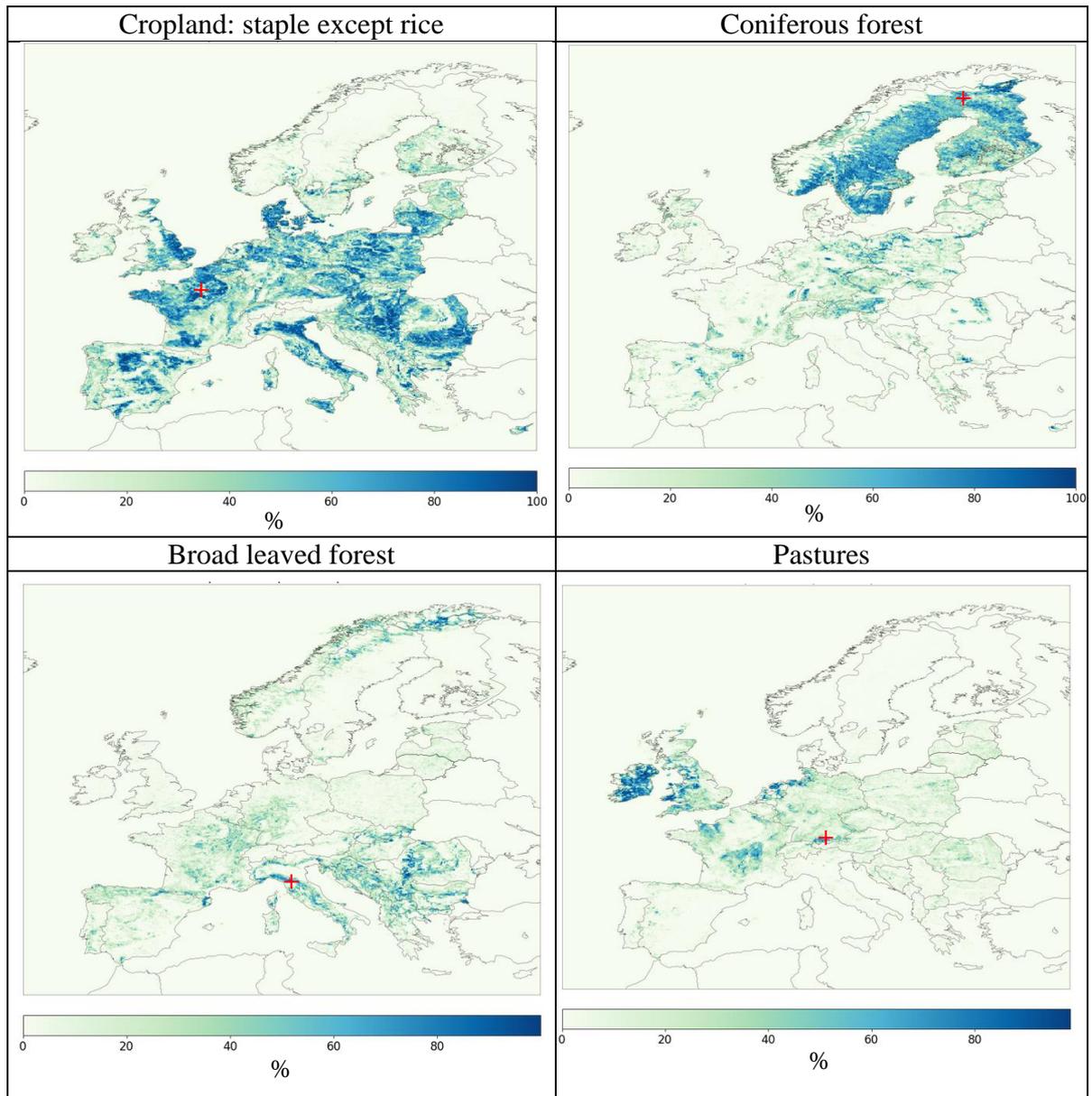
The STILT domain covers an area of almost 14.6 million km<sup>2</sup> which has been broken down into 19 different land cover classes (see section 3.2.2) in *Table 4*. It is compared to the average breakdown of what land cover classes all the stations combined were sensitive to year 2017 to examine how well the station footprint areas represented the domain. 49.1% of the domain is ocean, 10.5% “Cropland: staple except rice” and the 15.9% of the domain fall in the “No data” category (see section 3.2.1). Other significant categories include “Coniferous forest” (5.2%), “Broad leaved forest” (3.8%), “Natural grasslands” (2.8%) and “Pastures” (2.8%). The biggest difference compared to the average stations’ sensitivity were for classes “Ocean” and “No data”, with respectively 17.8 and 12.1% more in the domain. Areas with no data and ocean are mainly found towards the edges of the domain (see section 3.2.2), and most stations are more centrally located (see *Fig. 3*) which explain the big differences. However, the station sensitivity to ocean was still large considering the distance for most stations to oceans. The average for all stations is 31.3%, ranging from 11.4% for PAL to 66.0% for HEL. The extensive footprint areas are the reason for the high percentages (see section 4.1).

Table 4. Land cover breakdown for the STILT domain compared to the average breakdown of the 31 different stations.

Land cover type	STILT domain	Average stations	Diff
Ocean	49.10	31.33	<b>-17.76</b>
Cropland: Staple except rice	10.54	19.49	8.95
Coniferous forest	5.21	11.05	5.84
Broad leaved forest	3.76	7.61	3.85
Pastures	2.78	7.13	4.35
Mixed forest	2.18	3.93	1.75
No data	15.92	3.81	<b>-12.11</b>
Urban	1.26	3.62	2.37
Natural grasslands	2.79	2.79	0.00
Inland water bodies	1.02	2.21	1.19
Transitional woodlands and shrub	2.17	1.96	-0.21
Natural bare areas	1.22	1.78	0.56
Wet areas	0.86	1.20	0.35
Cropland: fruits, berries, grapes and olives	0.79	0.92	0.13
Industrial	0.26	0.74	0.48
Glaciers and perpetual snow	0.04	0.17	0.14
Ports and airports	0.03	0.08	0.05
Rice fields	0.05	0.08	0.04
Roads and railroads	0.03	0.06	0.03
Dump sites	0.01	0.02	0.01

Other than oceans, which have relatively low exchange of CO<sub>2</sub> with the atmosphere, the land cover category that the stations were the most sensitive to was “Cropland: staple except rice” with an average of 19.5% considering all stations year 2017. Averages for the different stations ranged between 2.0% for PAL and 32.7% for SAC. PAL is located in northern Finland and is apparently not in close proximity to much cropland. This is confirmed by the map in *Fig. 20* which shows the spatial distribution of “Cropland: staple except rice”. The station with the highest contribution, SAC, is marked with a red cross on the map. Many of the stations that had high sensitivity to this land cover category, including Ope (OPE) and TRN, are also located in the same general area. Second most significant in terms of the stations’ average sensitivity to land cover was “Coniferous forest”, also shown on a map in *Fig. 20* as well as ranked in *Fig. 19*. PAL had the highest sensitivity with 37.4% followed by the two stations Norunda (NOR) and SVB located in northern Sweden with about 25%. LMP had the lowest value with 0.8%, which is no surprise for a station located on an island in the Mediterranean (see *Fig. 2*). The average for all stations was 11.1% sensitivity to coniferous forest. LMP was also the station that saw the least of the category “Broad leaved forest” with 1.4%, whereas CMN saw the most with 32.7%. The average for all stations was 7.6% sensitivity to broad leaved forest. “Pastures” was the fourth most significant land cover category with a station average of 7.1%. Station HPB had the highest value with 17.1%. It is possible to further aggregate the land cover classes and combine their sensitivity values. This was done to categories associated with high population such as “Urban”, “Industrial”, “Dump sites” etc. (see *Table 3*). The

average sensitivity to these classes was 4.5% with higher values for JUE, FRE and HEI with 7.5, 7.6 and 8.7% respectively.



*Figure 19. Distribution of the four major land cover classes the stations were sensitive to (for year 2017). The STILT domain with  $1/12 \times 1/8$  degrees cells are used, and the cells are colored in accordance with percent of land cover within them. The red crosses mark the stations that were the most sensitive to each class year 2017.*

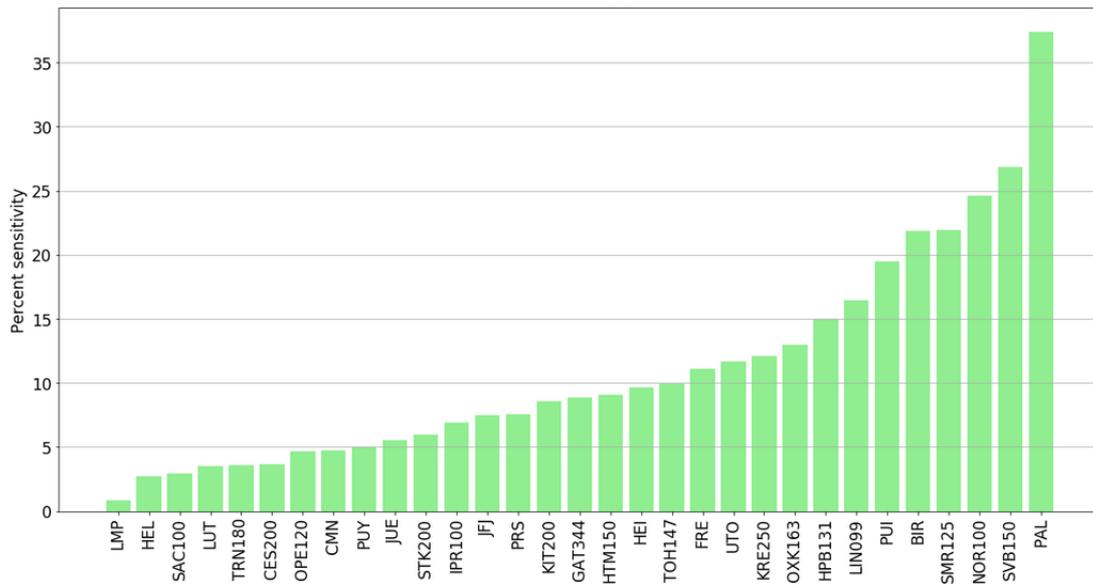


Figure 20. Percent sensitivity to coniferous forests at the different stations. Average year 2017.

The land cover breakdown for the individual stations can be displayed in a bar graph with absolute and percent sensitivity to the 19 different land cover classes with corresponding maps showing where the contribution to the sensitivity to the different land cover classes came from. This is exemplified for HTM in Fig. 21 and Fig. 22. HTM is fairly average in terms of what the stations were sensitive to year 2017, although the average sensitivity of 47.5% to ocean is high above the average for all stations. Values for broad leaved forests and pastures are a little lower than the average value. 2.57% sensitivity to “Urban” indicates that it is more remote than the average for the stations (3.6% year 2017). In terms of the distribution map showing where the sensitivity to “Cropland: staple except rice” came from, it indicates southern Sweden and Denmark with especially high values around the city of Lund.

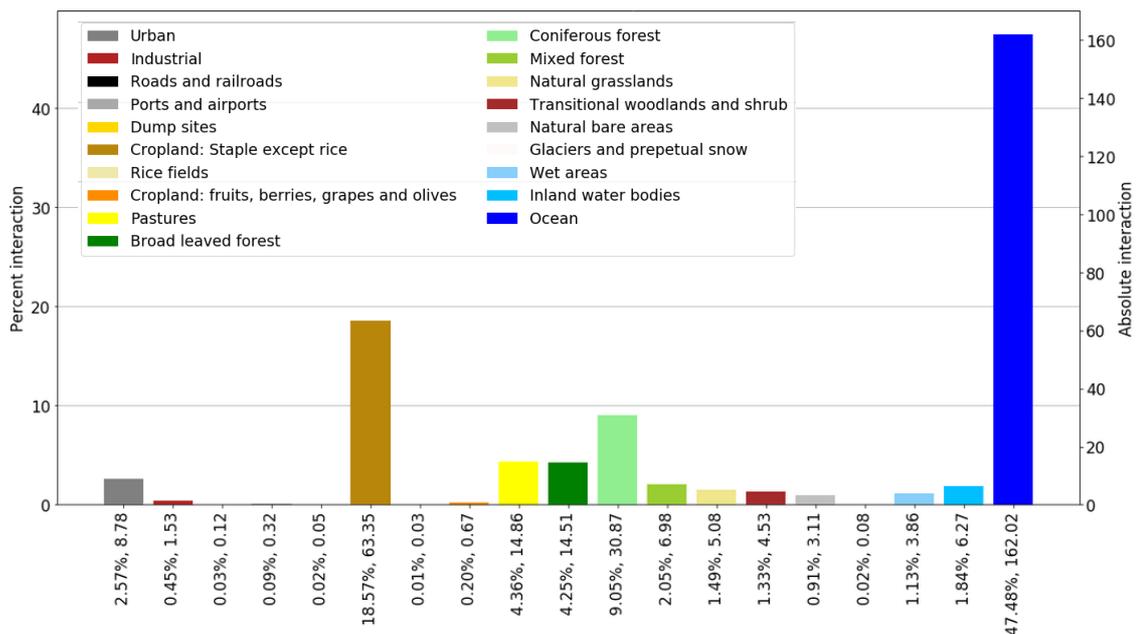


Figure 21. Average land cover breakdown for Hyltemossa year 2017.

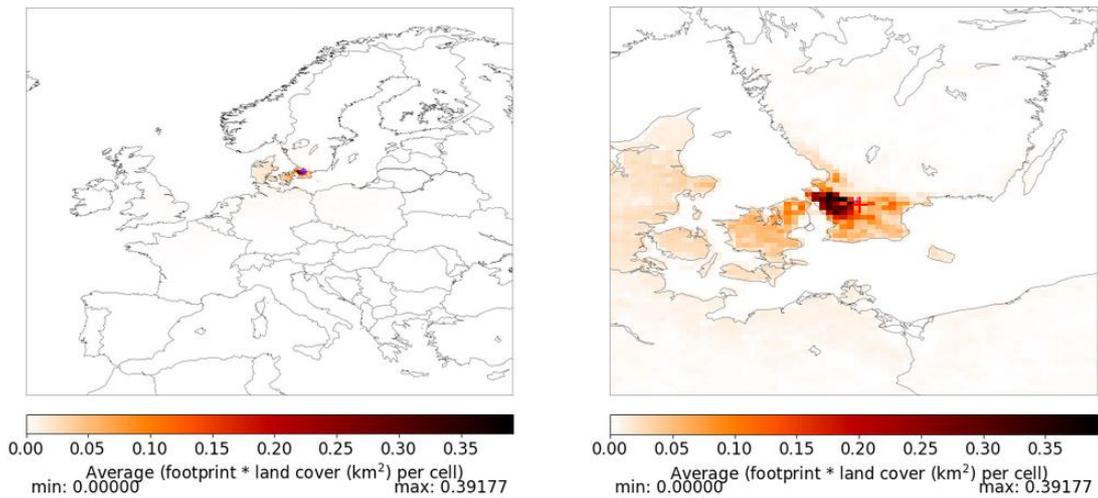


Figure 22. Sensitivity to the land cover class "Cropland: staple except rice" at Hyltemossa.

An important question is if the yearly averages for year 2017 are representative for the other years at the stations. In Fig. 23, the five main land cover classes at HTM are shown with percentage contributions for years 2015-2018. The values were similar all years and the "ranking" among the five land cover classes remained constant. The biggest difference was between year 2017 and 2018 in terms of sensitivity to ocean: there was 6% higher sensitivity to oceans in year 2017.

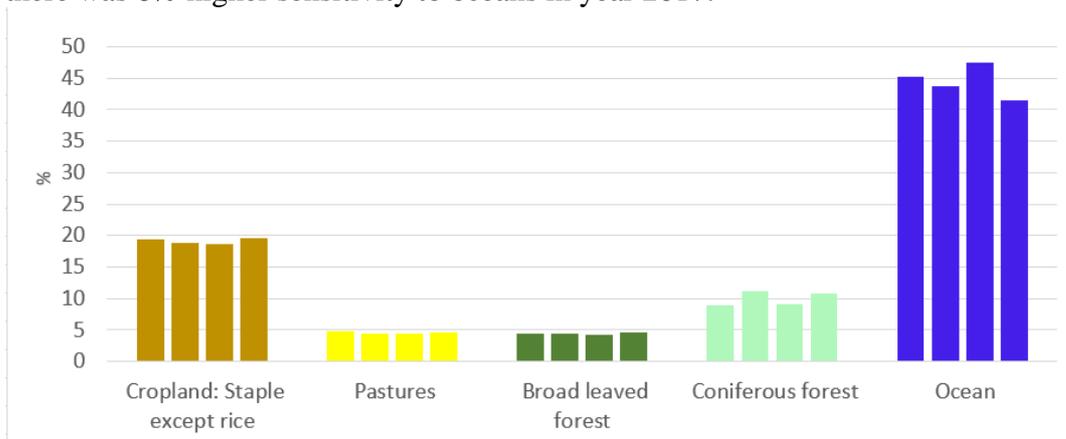


Figure 23. Land cover breakdown for the major land cover classes that station Hyltemossa was sensitive to years 2015-2018

#### 4.3.2 Population

The population density within the STILT domain is shown in a logarithmic scale in Fig 24. The reason for using a logarithmic scale is because the populated areas are small compared to the model domain, and the value in a  $1/12^\circ \times 1/8^\circ$  cell can be very high. The highest cell value is over two million per  $\text{km}^2$ . Areas with significant population, yet insignificant compared to the highest values, would not show up on the map if a linear scale was used. The relatively small areas with high population values also mean that slight changes in atmospheric transport can mean big differences in terms of the sensitivity to population and causes high standard deviation values at the individual stations.

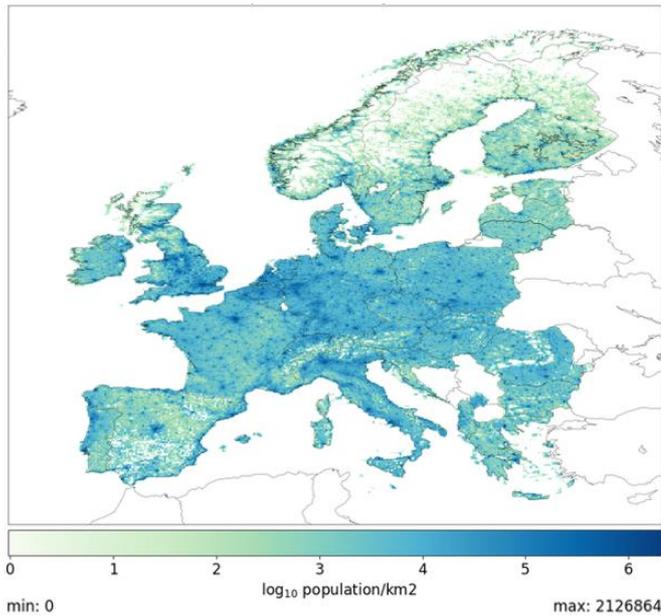


Figure 24. Population density year 2011. Note the logarithmic scale.

The output from multiplying footprints with the population data is only interesting for inter-comparison between the different stations. In Fig. 25, the average footprints for year 2017 have been used and the resulting values are ranked from lowest to highest. The stations that had the highest sensitivity to areas with large populations were the German stations HEI, ahead of FRE. At the other end of the spectrum were the remote stations PAL and SVB. The same rankings in terms of lowest and highest sensitivity to population were true for year 2016. However, there are large variabilities between the individual footprints: the standard deviation ranged between 64% of the year 2017 average at CMN to 218% for SAC.

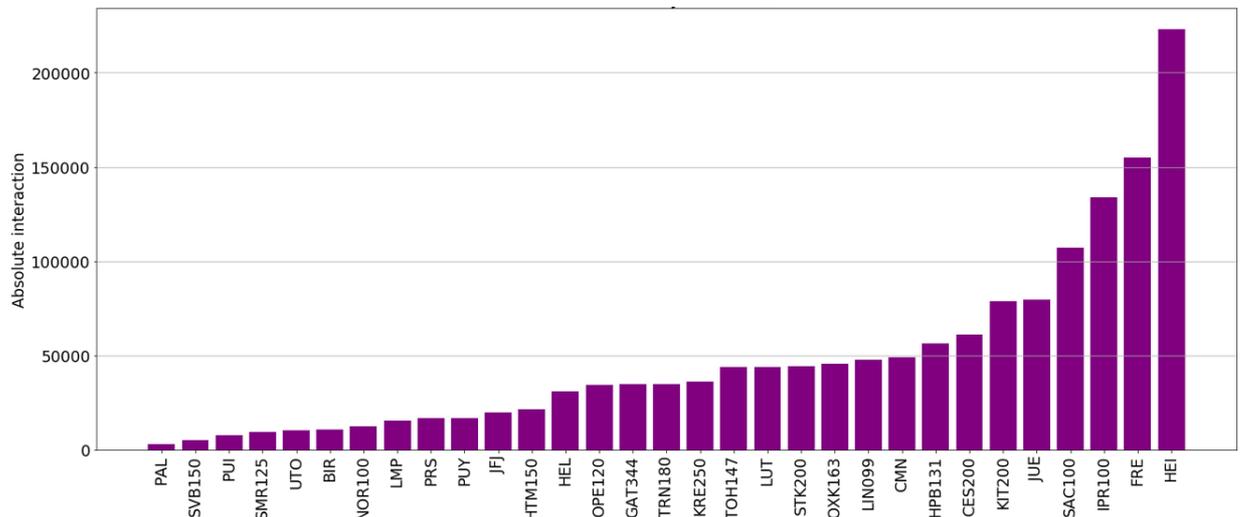


Figure 25. Average sensitivity to population year 2017.

It is possible to visualize where the contribution to the population sensitivity came from in a map. This is exemplified for SVB and HEI in Fig. 26. SVB is about 70 km away from the city of Umeå which has a population of just under 85 000 (2018). This city generated the highest sensitivity value for a single cell with 51. Compared to the total average value for year 2017 of 4990, this means that on average 10% of the

population that the station SVB was sensitive to lives within this cell. For HEI the highest cell value is almost 1000 times that of SVB and is located in the same cell as the station itself. It is located in the city of Heidelberg with a population of over 160 000 (2018). Mannheim, located less than 20 km from Heidelberg, is even larger with a population of over 300 000 (2018).

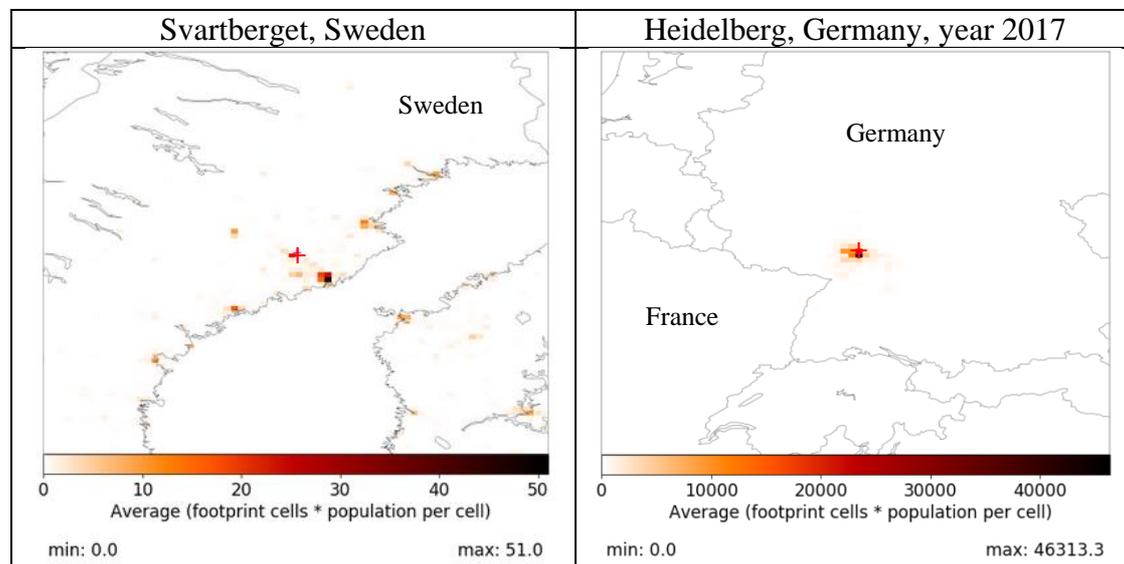


Figure 26. Spatial distribution of average sensitivity to population at Svartberget and Heidelberg year 2017.

#### 4.3.3 Point source emissions

The CO<sub>2</sub> point source emissions quantified in this section are already included in the EDGAR dataset, but there is value in being able to visualize it separately which is exemplified in a later section and Fig. 43. Fig. 27 is a map with the distribution of the point source emission facilities within the STILT domain and their emission quantities. The average influence they had on the CO<sub>2</sub> concentrations at individual stations year 2017 are shown in Fig. 28. PAL had the lowest average with 0.17 ppm, which was only 2.2% of the contribution from point source emissions at JUE which had an average of 7.55 ppm. The average for all stations year 2017 was 1.73 ppm, but over half of the stations had an average lower than 1 ppm. The German stations FRE, HEI and JUE brought up the average. The variabilities between the individual footprints at the stations were high with an average standard deviation value of 3.84 ppm for all stations. HEI experienced the most variability with a standard deviation value of 23.36 ppm. Point source emissions can, like population, be concentrated in few cells and slight differences in transport to the station can mean big differences in the estimated point source contribution to the CO<sub>2</sub> concentrations at the stations.

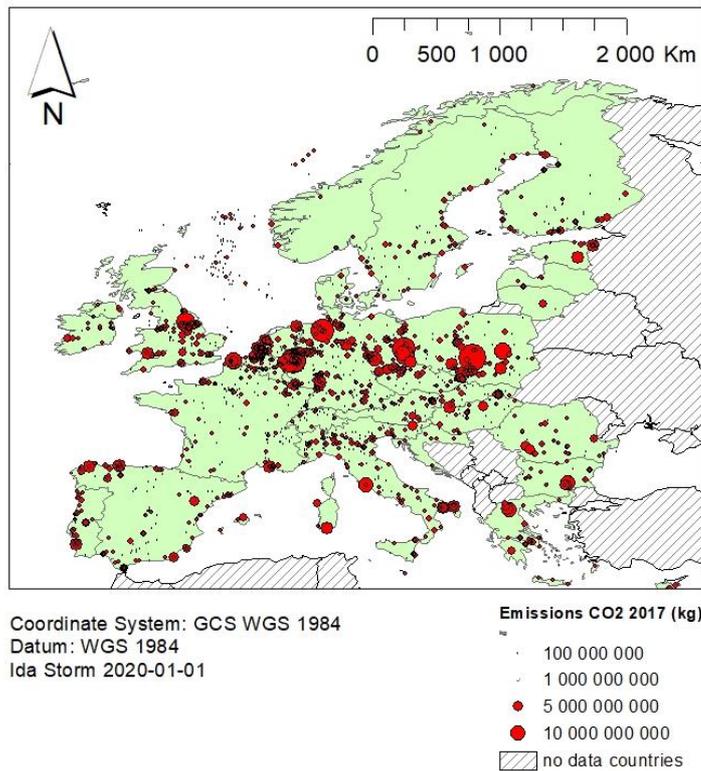


Figure 27. Point source emissions within the STILT domain year 2017.

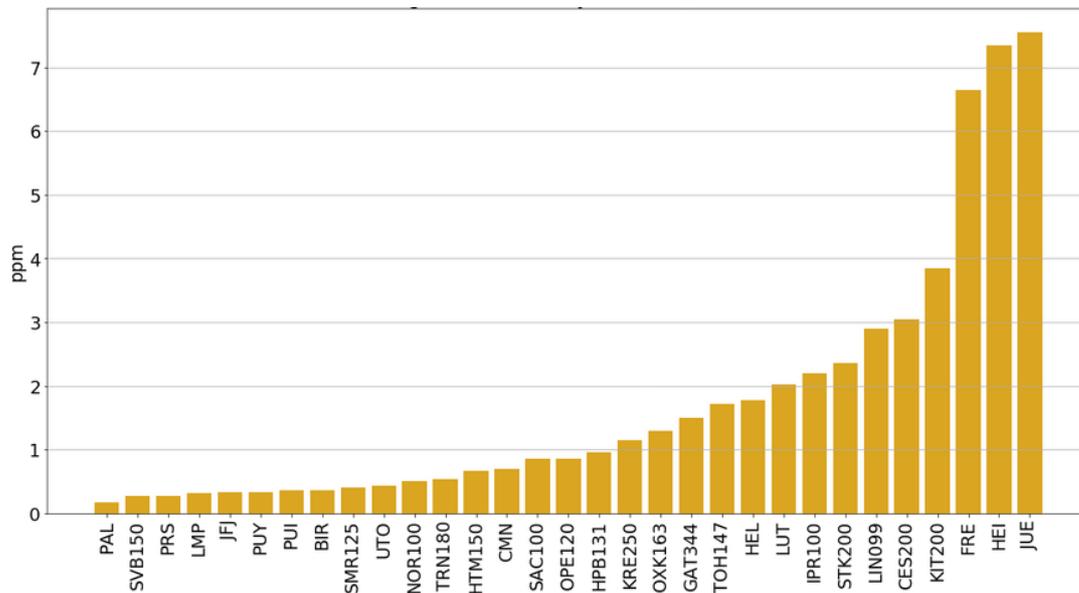


Figure 28. Expected contribution from point source emissions on the CO<sub>2</sub> concentrations year 2017.

It is interesting to consider point source emissions within the averaged footprints compared to how many people live in the area. Fig. 29 demonstrates this relationship, and it is not entirely linear: generally, stations that had a relatively high contribution from point source emissions were also sensitive to areas with large populations. However, stations SAC and IPR saw a lot of population and relatively low point source emission contributions whereas the opposite was true for station JUE.

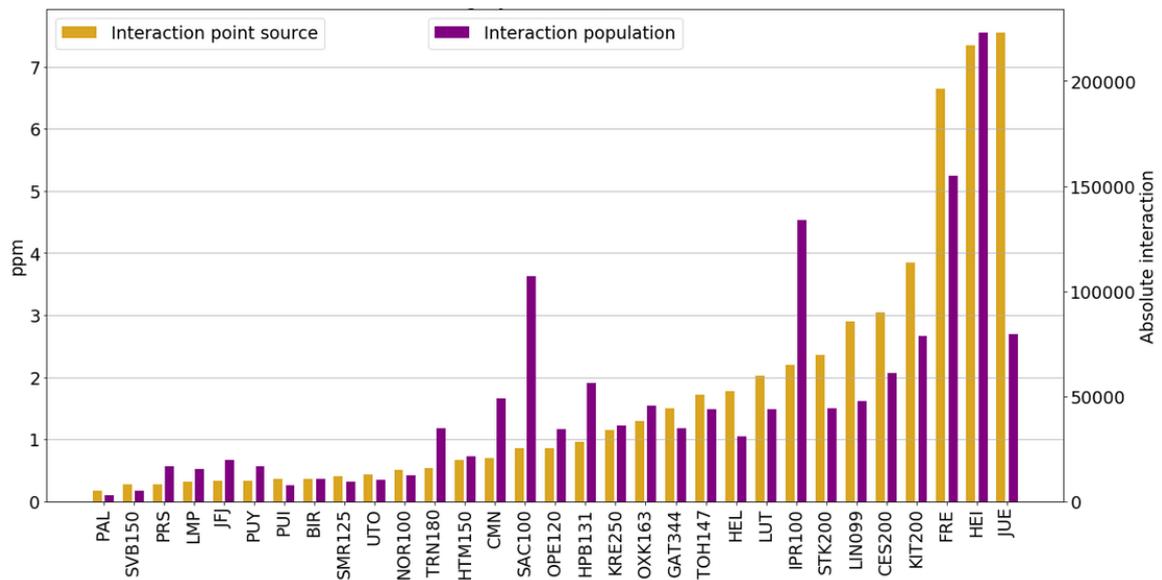


Fig 29. Sensitivity to point source emissions compared to sensitivity to population year 2017. Sorted lowest to highest contribution from point source emissions.

Regarding how representative these annual averages are for other years, there are some differences between years 2016 and 2017: in general, the values were higher in year 2016 compared to year 2017 with an average difference of just over 11%. The difference was largest for PAL with 47% higher contribution from point source emissions year 2016 (0.24 ppm compared to 0.16 ppm). The same point source emissions data was used both years which means that the differences are due to differences in the transports to the stations. As mentioned in section 4.1, the total average sensitivity values for year 2016 were generally higher than year 2017 and is likely the explanation.

#### 4.3.4 Radiocarbon (<sup>14</sup>CO<sub>2</sub>)

Nuclear power plants and fuel reprocessing stations that emit radiocarbon as <sup>14</sup>CO<sub>2</sub> are spread out within the STILT domain in accordance with Fig. 5 (section 3.2.4). The average expected shift from the normal ratio of <sup>14</sup>C to <sup>12</sup>C due to emissions from these facilities are shown in Table 5 for year 2016. HEI was on average influenced the most with an average estimated shift of 2.3 ‰. Other stations with relatively high values are SAC, Cabauw (CES), and TRN. Although small, these emissions have effect on the ratio all stations and the standard deviation values are relatively high for many stations. A high standard deviation value indicate that the influence can be quite a bit higher for certain time-steps.

Table 5. Estimated average shift in the ratio between  $^{14}\text{C}$  and  $^{12}\text{C}$  due to radiocarbon emissions at nuclear power plants and fuel reprocessing stations year 2016.

Station	$\Delta_{14}\text{C}$ (‰)	Standard deviation	Station	$\Delta_{14}\text{C}$ (‰)	Standard deviation
HEI	2.3	6.63	OXK	0.72	1.07
SAC	1.95	6.39	SMR	0.64	1.33
CES	1.62	3.78	KRE	0.59	0.9
TRN	1.51	3.59	LIN	0.59	1.07
KIT	1.34	2.06	HTM	0.56	0.95
FRE	1.31	3.87	PUY	0.51	1.15
JUE	1.25	2.7	IPR	0.48	0.66
NOR	1.22	5.02	BIR	0.46	0.92
LUT	1.09	2.35	JFJ	0.43	0.92
OPE	1.09	2.15	PUI	0.43	0.98
STK	0.94	1.66	SVB	0.37	0.78
HEL	0.9	1.94	PRS	0.35	1.17
HPB	0.84	1.52	CMN	0.32	0.48
UTO	0.8	2.16	PAL	0.21	0.4
TOH	0.77	1.37	LMP	0.18	0.32
GAT	0.72	1.23			

It is possible to see exactly which facilities contributed to the estimated shifts. For HEI, this is shown on a map in *Fig. 30* with the facilities that shifted the ratio on average more than 0.05‰ year 2016. The most significant influence came from Phillipsburg 2, located about 30 km northwest of HEI, with an average ratio shift of 1.2‰. Hence, just over half the influence at HEI came from there. The second largest contributor is located much further away – La Hauge in France – and was responsible for an average shift of 0.4‰. Radiocarbon emissions with 19100 TBq radioactivity year 2016 made La Hauge’s influence appear as significant in footprints for many stations around Europe. It is by far the largest anthropogenic source of radiocarbon within the domain. The second largest emitting facility, Leningrad, only emitted 20% of that, and Phillipsburg just about 0.5% (89 TBq). The influence from the individual facilities can also be visualized in a line graph, exemplified by influence at HEI during January 2016 in *Fig. 31*.

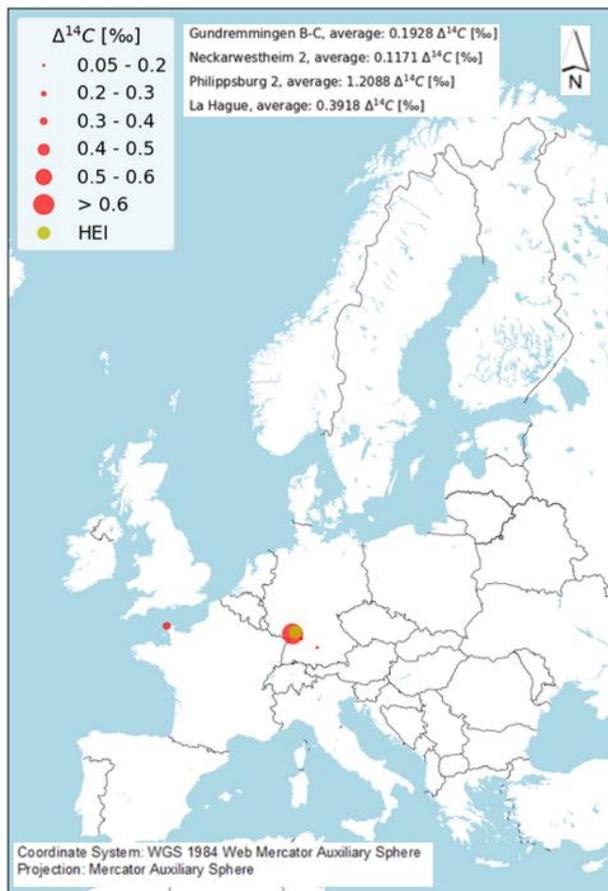


Figure 30. Heidelberg's estimated shift in ratio  $^{14}\text{C}$  to  $^{12}\text{C}$  attributed to radiocarbon emissions at different facilities year 2016. Threshold for inclusion for a facility:  $>0.05$  ‰.

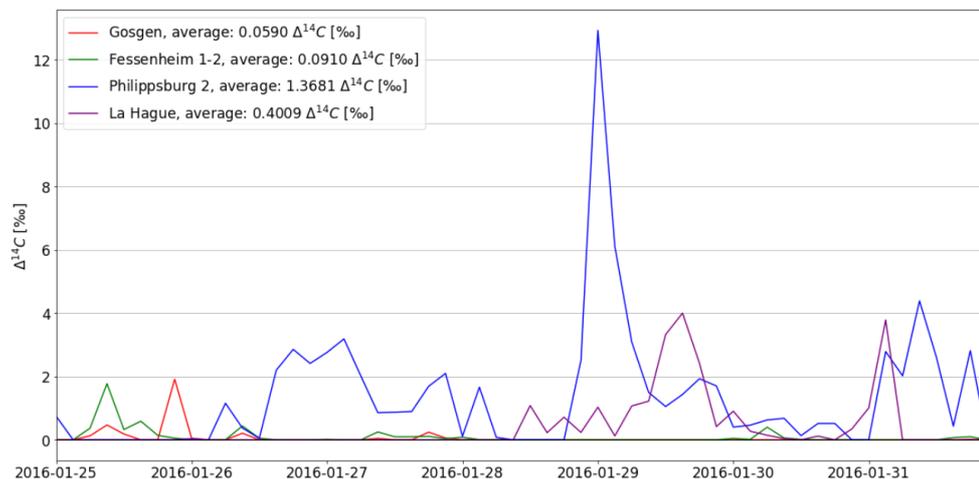


Figure 31. Time series showing the influence from facilities with radiocarbon emissions on the ratio  $^{14}\text{C}$  to  $^{12}\text{C}$  at Heidelberg, last week of January year 2016. Threshold for inclusion for a facility: average of  $>0.05$  ‰ for January.

The standard deviation value of 6.63‰ for HEI year 2016 indicates high variability in contribution from nuclear power plants and fuel reprocessing plants over time. This is clear when considering the time series for HEI in Fig. 32: days with no influence are followed by high peaks in the next three-hour period. The footprints with the peaking

values, defined as shift in radiocarbon at the station by more than 50‰, have been subset and considered in terms of their sensitivity values (see section 4.1) as well as which facilities contributed the most to the shift. The result point to very high local influence with major contribution to the shift from the closest nuclear power plan Phillipsburg 2.

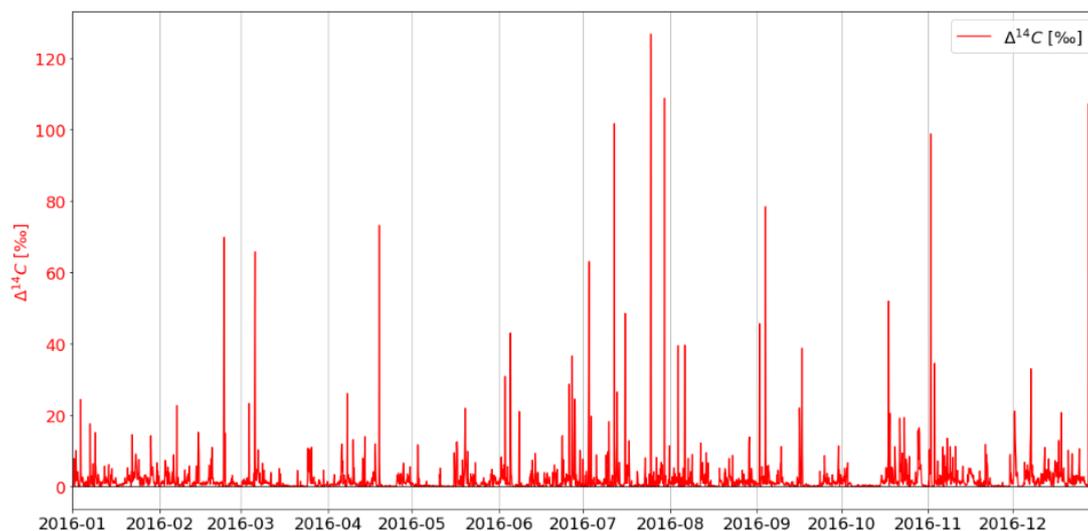


Figure 32. Shift in ratio  $^{14}\text{C}$  to  $^{12}\text{C}$  at Heidelberg year 2016 due to emission of radiocarbon from nuclear power plants and fuel reprocessing stations.

#### 4.4 Modelled concentration compared to measured concentration

For stations that have measure  $\text{CO}_2$  concentrations these can be compared to the corresponding modelled  $\text{CO}_2$  concentrations. There were 15 stations with measured concentrations available for year 2017 (see Table 2). The average model data differences for these are displayed in Table 6. The averages ranged from an underestimate of 4.73 ppm at Lutjewad (LUT) to an overestimate of 0.29 ppm at PRS. On the whole, 13/15 stations had an average underestimate for year 2017. The NaN-values in Table 6 refer to individual footprint during the year that either had no modelled  $\text{CO}_2$  concentration, or lacked measured concentration. The latter was the most common reason for NaN-values.

Table 6. Average model data differences for year 2017.

Station	Average year 2017 (ppm)	$\sigma$ of model data differences	NaN-values (out of 2920)
LUT	-4.73	8.00	339
LIN	-2.89	9.02	54
NOR	-2.56	4.41	775
TRN	-2.22	5.70	513
KRE	-2.06	4.60	1138
OPE	-1.9	5.18	316
SMR	-1.74	3.17	112
HTM	-1.67	3.66	921

HPB	-1.46	5.69	100
SVB	-1.44	3.06	1311
GAT	-1.28	4.09	905
PUY	-0.7	3.51	74
JFJ	-0.53	3.08	166
HEI	0.18	35.52	451
PRS	0.29	2.30	520

Fig. 33 displays a time series with TRN’s modelled and the measured concentrations, as well as the resulting model data differences in a separate time series at the bottom. In the model data difference time series, the zero-line means perfect agreement between the modelled and the measured concentration values. Many footprints have underestimates larger than 20 ppm and there are only a couple of overestimates of the same magnitude. There can be big differences from one footprint to the next separated only by three hours.

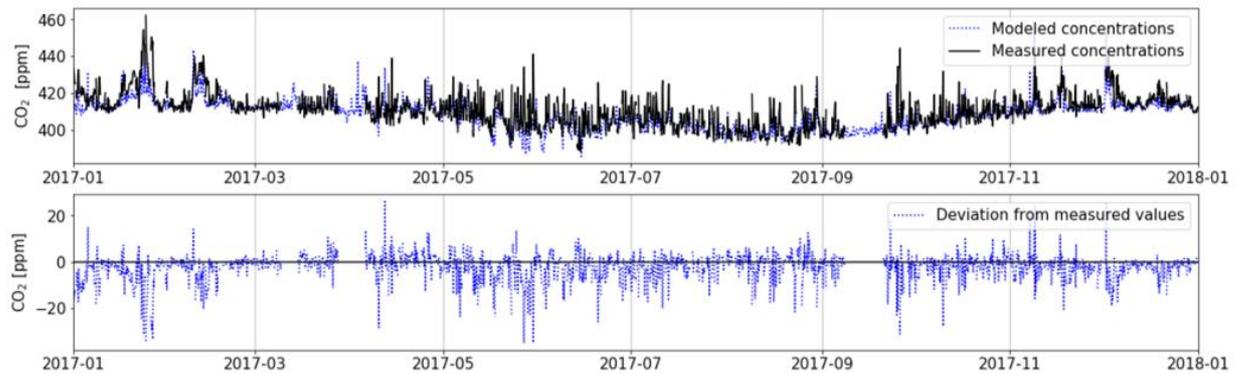


Figure 33. Time series with modelled and measured CO<sub>2</sub> concentrations at Trainou year 2017. Model data differences – the modelled concentration minus the measured concentration for each footprint – are displayed in the bottom graph.

Already in Fig. 33, there are indications of the model having more difficulties estimating the CO<sub>2</sub> concentrations during certain parts of year 2017. This was closer examined for all stations by aggregating the model data differences by month as well as counting the number of footprints that had a model data differences above or below one standard deviation of the model data differences. It is important to consider both ways of quantifying the model data differences: with the count, possible trends of stations with months that have both over- and underestimates become clear. Another benefit from using the standard deviation values as thresholds is that stations with low variability, such as JFJ and PRS, can also indicate poorer model performance for certain months. Results are presented in Figures 34-36. It should be noted that the high number of NaN-values especially for stations KRE, SVB, Gartow (GAT) and HTM (see Table 6) mean months with averages based on few, and sometimes zero, model data differences. Generally, the figures indicate monthly variations in model performance, but there is no overall trend that applies for all stations. However, most months (over 80%) had an average underestimate. HEI stands out in Fig. 34 as a station with large differences between the different months: January, May and December were months with average underestimates and especially high

underestimates – greater than 10 ppm – in January and December. For April the opposite was true with an average overestimate of 10 ppm. PRS, JFJ and HPB are other stations where the average model date differences vary between overestimates and large underestimates. For all three stations the winter months were associated with overestimates and the summer months with underestimates.

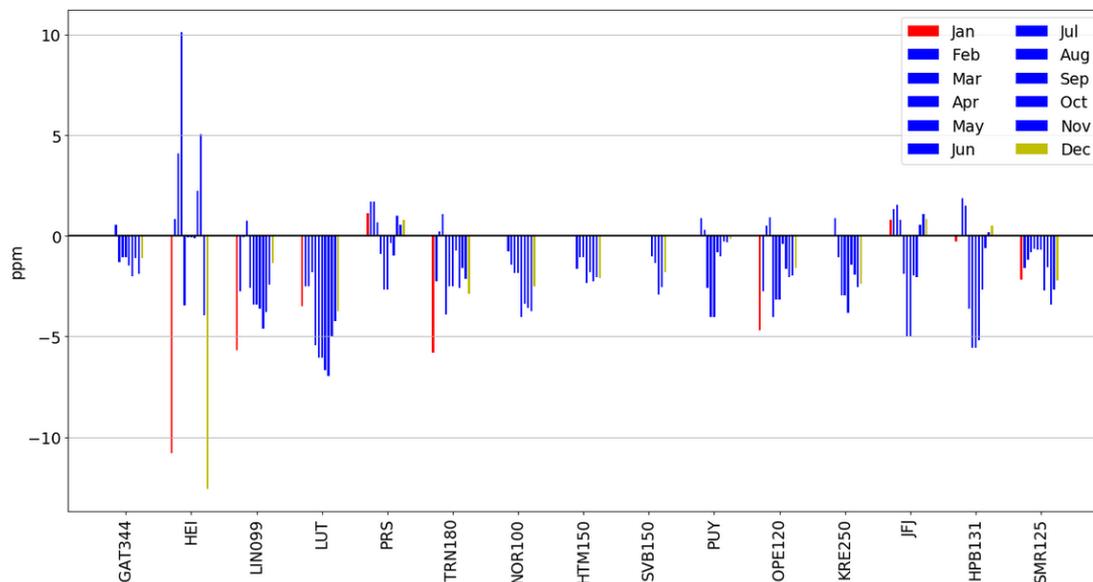


Figure 34. Average model data difference by month year 2017

In Figures 35 and 36, the data are aggregated in terms of count of model data differences above or below one standard deviation of the year 2017 average model data differences. When comparing the two figures, it is important to notice that the y-axes have different scales: the counts of underestimates are much higher. For most stations it seems that the model is more prone to underestimating the CO<sub>2</sub> concentration in the summer months, and the times the model overestimates are more often in the winter months. For LIN (Lindenberg), TRN and SMEAR/Hyytiälä (SMR), this did not apply. Stations where high model underestimates at least partly were taken out by high overestimates become clear: HEI and PRS are the only stations with similar counts of over- as underestimates, witnessing of more and higher over- and underestimates than the averages in Fig. 34. It is also interesting to consider the stations that had low variability: the summer months at JFJ had a high count of underestimates with June having highest with over 140 (out of 248). For PRS, January, February and March were months with many overestimates. The highest count of overestimates was for February with just under 70.

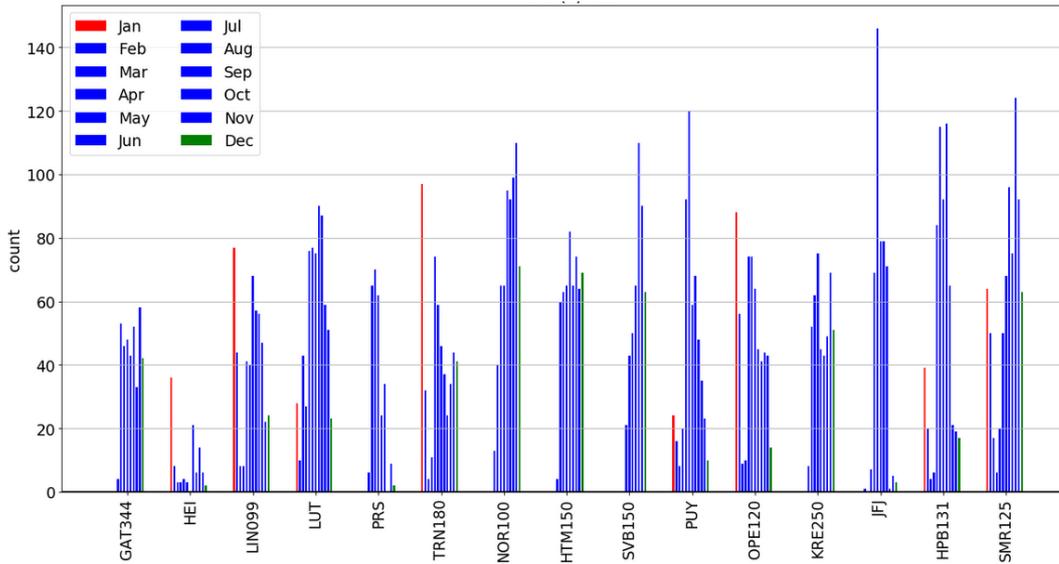


Figure 35. Count model data differences one standard deviation below measured concentrations by month year 2017.

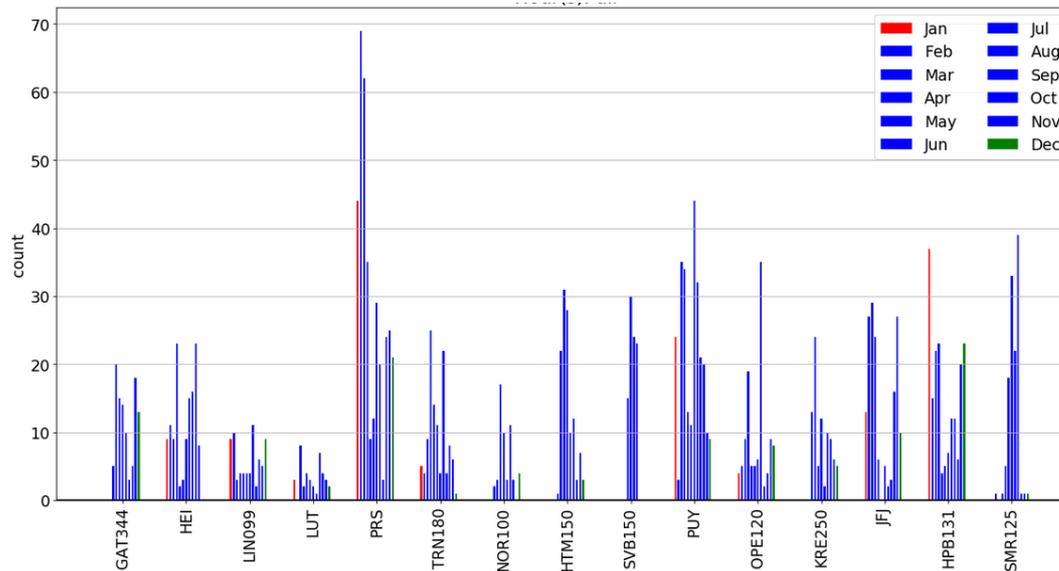


Figure 36. Count model data differences one standard deviation above measured concentrations by month year 2017.

Model data differences have also been aggregated and averaged by the different hours that the STILT model system is set to run for: 00:00, 03:00, 6:00, 9:00, 12:00, 15:00, 18:00, and 21:00. Fig. 37 shows the average model data difference and the general problem of the STILT model system underestimating the CO<sub>2</sub> concentrations is clear also here: other than average overestimates during hours 21:00, 0:00 and 3:00 at HEI, all hours at PRS, and 12:00 and 15:00 at PUY, all hours at all stations are on average underestimates. Better averages are associated with day-time for most stations with eight stations having the lowest average at 15:00.

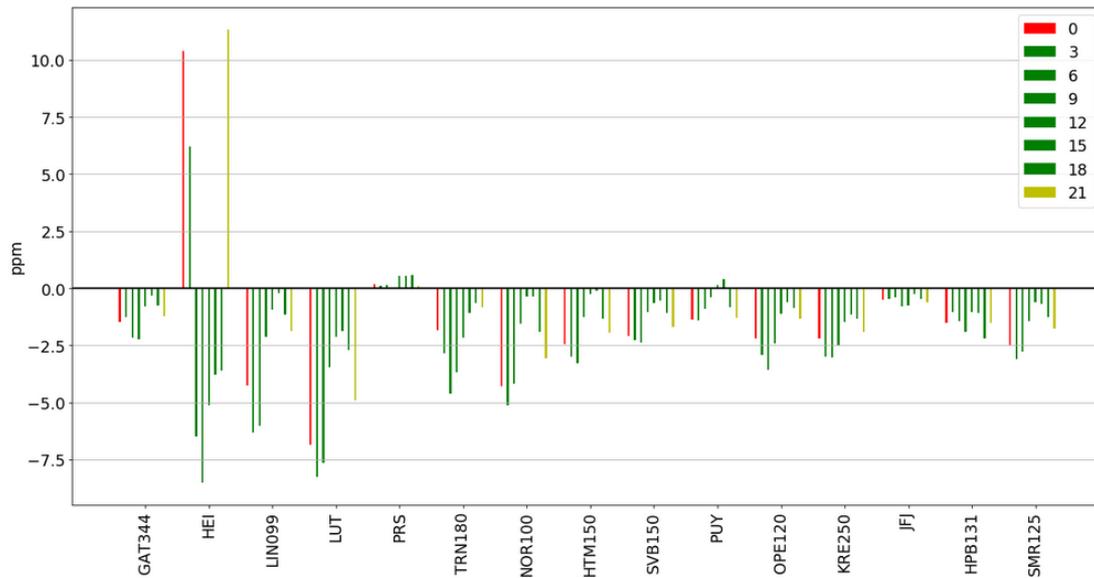


Figure 37. Average model data difference by hour year 2017.

When aggregating the model data difference by hour in terms of count of over- and under one standard deviation of the model data differences, KRE, HPB and JFJ indicate similar patterns over the different hours. Again, day-time model runs show better performance with hours 12:00 and 15:00 being best for most stations. This trend is not as apparent in terms of count of overestimates with stations HTM, PUY, GAT and SMR that rather had more overestimates during these hours. There were really high counts of night-time underestimates for stations LIN, LUT, NOT OPE, and SMR.

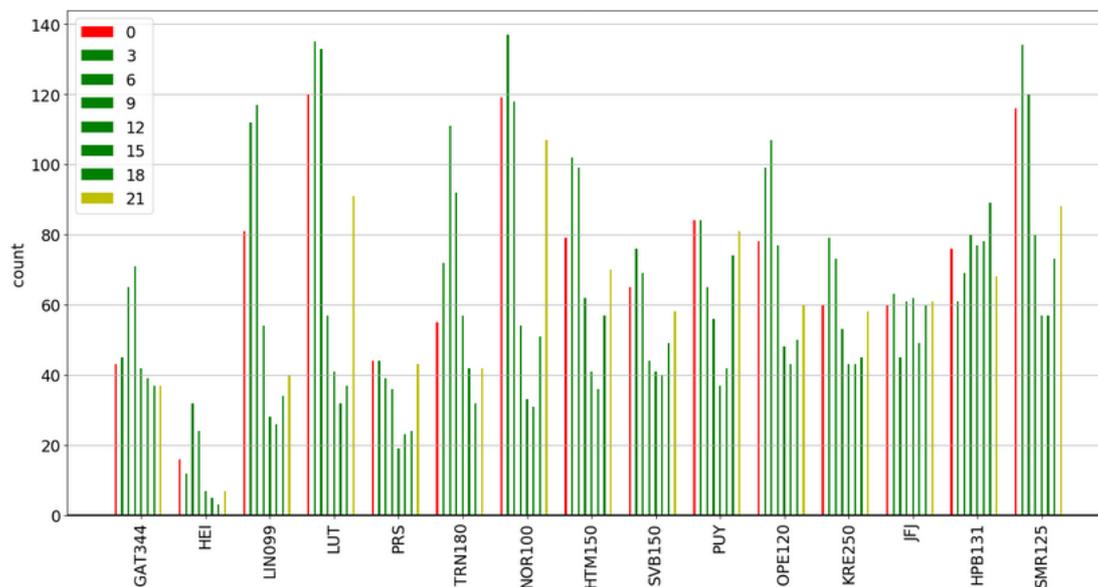


Figure 38. Count model data differences one standard deviation below measured concentrations by hour year 2017.

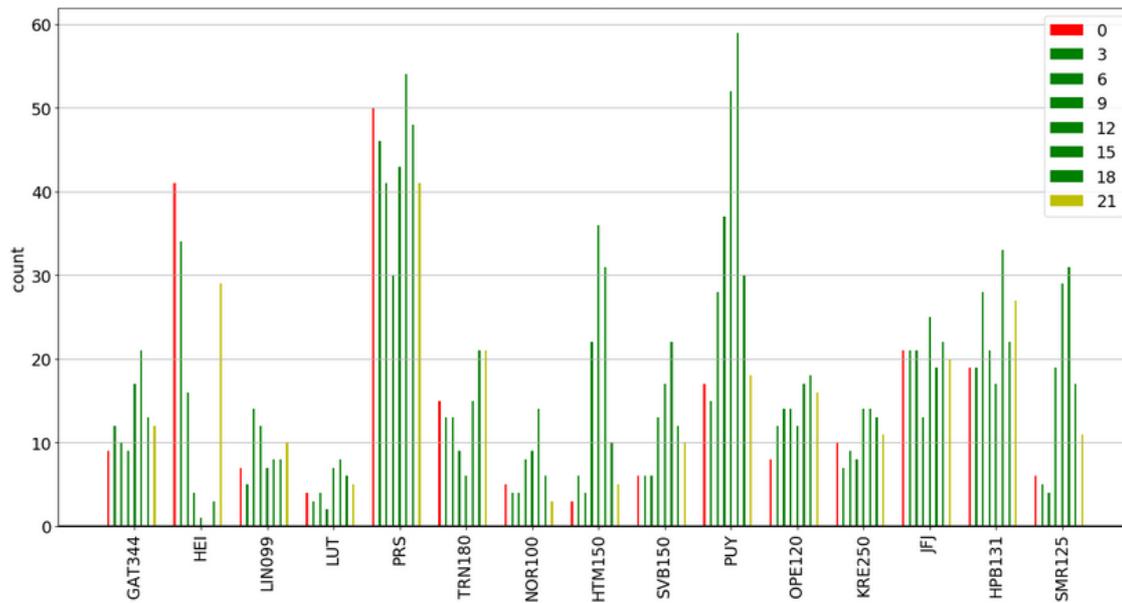
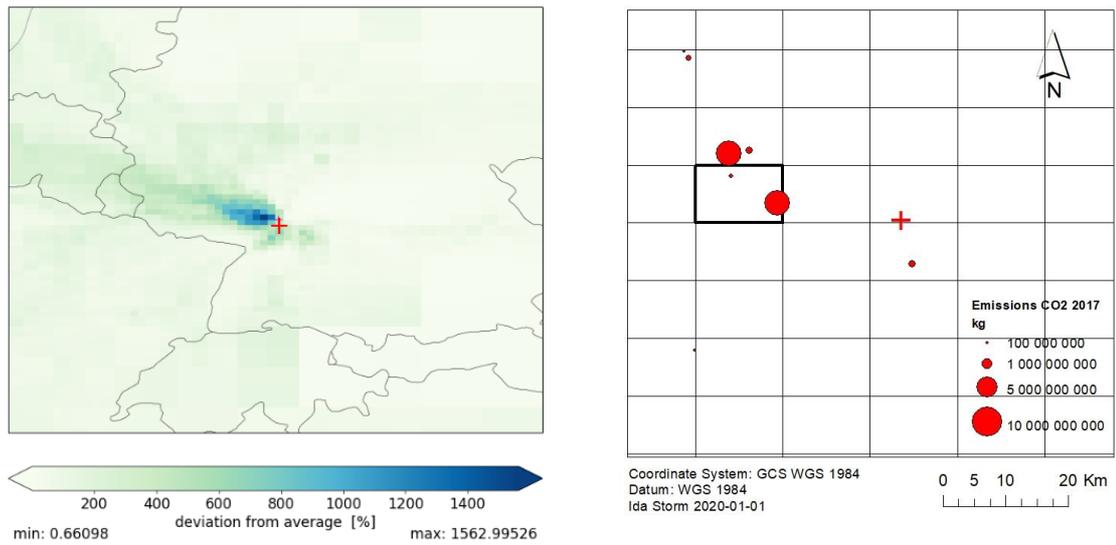


Figure 39. Count model data differences one standard deviation above measured concentrations by hour year 2017.

The process of sub-setting footprints based on certain attributes, such as hour and month, can also be applied to each footprint’s model data difference. In *Fig. 43*, spatial patterns of overestimates larger than one standard deviation (126) of the model data differences for HEI year 2017 are considered. The area immediately west of HEI was associated with overestimates with one cell having almost 1600% higher sensitivity in the subset of the 126 overestimating footprints compared to the year 2017 average. To gain a better understanding of why this is, running analyzes on the subset in the different Notebooks is a good start: the average contribution from point source emission during the high overestimates is 60.8 ppm, which is very high compared to the year 2017 average of just over seven for HEI (see *Fig. 28*). The cell with a value close to 1600% – highlighted by an extra thick outline in the map *in Fig. 40* – alone contributed with an average of 37.7 ppm. The cell contains Power Plant Mannheim (Grosskraftwerk Mannheim) which emitted 6860000 tons of CO<sub>2</sub> year 2017. Power plants emit from high stacks, and the hot air rises fast. This is not accounted for in the STILT model system where the emissions are estimated to mix with the air passing in the lower half of the PBL, making overestimates of their influence likely. Also, the emissions are distributed into the whole cell equally, as opposed to at a specific location within the cell, and hence transport of air from west of HEI will always “hit” this cell and be subjected to the error. The sensitivity to the area within 100 km of the station is 3.5 times higher for the subset than the average which is how point source emission close to the station can have such high influence on the modelled concentration.



*Figure 40. Areas associated with overestimates at station Heidelberg year 2017. The percentage values of the cells represent how much sensitivity the footprints associated with high overestimates (126 footprints) had compared to the average of all footprints year 2017 (2469 footprints). The map to the right shows the point source emission locations surrounding Heidelberg.*



## 5. Discussion

Connecting back to the overarching research question, there has indeed been a spatial characterization, but also a general characterization, of the different stations. Key has been to characterize the stations by using the other stations as reference. Year 2017's footprints have been used to generate annual averages that in turn have been used to rank the stations among themselves. This has many times been complemented with averages for year 2016 and by and large the rankings among the stations have been similar between the two years. To characterize more years is needed to confirm if the rankings are stable and hence if the characterizations of the stations for year 2017 are representative. There are clear connections between a few of the different areas of characterization making for similar rankings: anthropogenic emissions are by definition emissions that humans emit and can be expected to be strongly correlated to the size of the population in an area. Furthermore, point source emissions come from factories and other facilities that often require large work forces and the infrastructure of built up areas and are hence usually found in relatively close proximity to populated areas. Also, the point source input dataset from E-PRTR is part of the EDGAR dataset that represents anthropogenic emissions. Eleven stations had an average anthropogenic contribution lower than two ppm year 2017: all seven stations further north than Birkenes (BIR) in Norway – including BIR – and only mountain station stations JFJ, PRS and PUY south of BIR except LMP which is located on a remote island in the Mediterranean. Ten out of these eleven station also had the lowest sensitivity to population and point source emissions within their average footprints. Although mostly situated at remote locations in central Europe, the remaining stations are simply too close to areas of high populations to have a very low contribution of anthropogenic emissions to the CO<sub>2</sub> concentrations at the stations. However, it will entirely depend on the transport to the station and filtering out footprints with high anthropogenic contribution can be done if appropriate for the intended analysis. Contribution from anthropogenic emissions within the footprint areas of central European stations is also indicated by the categorization in Henne et al. (2010) (see section 2.4). Their ranking, LMP, BIR/JFJ, PUY/CMN, IPR/CES - from least to most subjected to fluxes - agrees with this thesis's population ranking where IPR and CES were ranked the highest among the seven ICOS stations considered in the paper (see *Fig. 25*). Reproducing the classification with the same parameters as in the paper could be an idea for the future. The necessary values are already derived for the 31 stations considered in this thesis.

With regards to land cover, the stations have been considered in terms of percentage of the 19 different land cover classes within their average footprint areas for year 2017. Large spatial extents of footprints are an argument to do this kind of breakdown to fully understand what land cover classes are within the sensitivity area of the stations. Other than oceans, the land cover classes with the highest values were “Cropland: staple except rice”, “Coniferous forest”, “Broad leaves forest” and “Pastures”. Similarities between what stations are sensitive to include 25-35% sensitivity to “Cropland: staple except rice” for 13 stations located on similar longitudes reaching from TRN, SAC and OPE in France to KRE in Czech Republic and high sensitivity to coniferous forests for northern stations such as PAL, SVB and NOR. Categories associated with urban activities constitute a small percentage for all stations, reflecting high concentrations of Europeans living in relatively small areas compared to the size of the STILT domain. The biospheric component amounted to

an average sink for almost all station based on their year 2017 average footprints and stations, largest at JFJ, PUY and HPB. Exceptions were CMN, FRE and HEI and for CMN the biospheric component was also an average source in the summer. This is not realistic and may reflect an underlying problem with the VPRM model. The biospheric component will be affected by what land cover types are found within the footprints. However, no connections were made between annual average land cover breakdown and the annual biospheric components at the stations: the land cover map used to calibrate the parameters for the VPRM model (SYNMAP, Jung et al., 2006), which in turn was used to derive the biospheric component, is different than that used for the land cover breakdown (CORINE). SYNMAP might be interesting to quantify and use in the future, but has fewer land cover classes than CORINE and has been specifically developed for parameterization in a biospheric model and hence did not fit the purpose of this thesis. Another interesting venue would be to correlate land cover breakdown to model performance on the basis of individual footprints to see if any connections between land cover type and model performance can be made. However, the fluxes between the biosphere and atmosphere are the hardest to model with many additional factors other than land cover at play (see section 2.2).

The final annual averages to consider are the influence of radiocarbon emissions within the footprint areas on the ratio between the radioactive carbon isotope  $^{14}\text{C}$  and  $^{12}\text{C}$ . Other than the emissions from these facilities, as well as dilution of the ratio from burning fossil which has no  $^{14}\text{C}$ , the background ratio between the two isotopes are stable. Hence, the measured ratio can be used to estimate the fossil fuel emissions within the footprints, and – although generally only a small influence – accounting for the positive influx of  $^{14}\text{C}$  from radiocarbon emissions make for better estimates. Considering the average for year 2016, all stations saw at least a very small shift in the ratio with more significant values for stations HEI, SAC, CES, TRN and NOR. The stations had generally high standard deviation values with highest for NOR with 400% compared to the average shift. The explanation lies in NOR's close proximity to Forsmark – one of the nuclear power plants with the highest emissions year 2016 – while the station is relatively far from other sources of radiocarbon: air transport over Forsmark will cause high values, and otherwise there will be very low values. 400% is the highest standard deviation value compared to the average for all annual values considered, but rare land cover classes have values up to over 200%. Also, population, with relatively few cells of high or very high concentrations, have high variability with an average standard deviation considering all stations of 220% compared to the average. The large variabilities suggest that whereas the annual averages are good for a general overview, there is value in considering footprints individually and possibly subset footprints based on desirable characteristics before analysis. The input ancillary datasets are static in time, which means that the differences are only due to differences in footprints stemming from changes in transports to the stations.

Footprints are important to understand and to connect to unexpected characterization values for the stations. Footprints vary both in terms of total sensitivity and their distribution. Total sensitivity is the resulting value from adding all the cell values and footprints of atmospheric stations, with tall towers, generally have low total sensitivity values. The elevation of the station clearly plays an important role for the total average sensitivity values for the year 2017, with three of the stations located on mountains having the lowest values (see *Fig. 8*). This is the reason low contributions

from anthropogenic emissions were seen at these stations year 2017 despite being located in central Europe. However, the mountain station CMN is one of the stations with highest average sensitivity which indicate the importance of other factors as well, including meteorological conditions effecting the transports to the stations. The difference between the station with the lowest total average sensitivity, PRS, and the station with the highest, LMP, is almost 400%. However, the bulk (21/31) of the stations range between just under 4 and 5.6 ppm / ( $\mu\text{mol} / \text{m}^2\text{s}$ ). The difference in total sensitivity between footprints at the individual stations are even bigger with standard deviation values up to 12 ppm / ( $\mu\text{mol} / \text{m}^2\text{s}$ ) for LMP. Footprints can have really high total sensitivity values exemplified by the 2017 years' time series for HTM (*Fig. 9*) with an average of 3.7 ppm / ( $\mu\text{mol} / \text{m}^2\text{s}$ ) having a maximum value close to 30 ppm / ( $\mu\text{mol} / \text{m}^2\text{s}$ ). High footprint sensitivity values can in turn lead to outliers when it comes to the characterization values since they are derived from multiplying footprints with underlying data. In terms of the distribution of the sensitivity an average of 48% were to the area within 300 km of the stations considering all stations' averages year 2017. However, the distributions are also subjected to high variability and sometimes most of the footprint is concentrated within 100 km of the station. Differences in general meteorological conditions between summer and winter caused for most stations to have higher sensitivity values in the winter (Dec. 2016- Feb. 2017) compared to summer (Jun. 2017- Aug. 2017). This was true for 25/31 stations with 13 stations having an average of over 150% higher in the winter. For three stations, LMP, JFJ and PRS, the sensitivity values were lower in the winter with 25, 48 and 51% of what they were in the summer. This can in turn help explain why the average anthropogenic contributions were lower in the winter than in the summer for these stations which is opposite of what is true for all other stations. There are also differences between nighttime hours 21:00, 0:00, 3:00 and 6:00 compared to 9:00, 12:00, 15:00 and 18:00 with 28/31 stations having higher average sensitivity values in the daytime.

The setup of the radiocarbon analysis, where radiocarbon emissions from nuclear power plants and fuel reprocessing stations influence are estimated, will be implemented with measured ratios between  $^{12}\text{C}$  and  $^{14}\text{C}$  to estimate fossil fuel emissions within the footprints. A measured ratio represents a short window of time so ratio shifts for individual footprints will be used after annual averages have confirmed influence high enough to motivate accounting for these emissions. It is possible to breakdown the influence between the individual nuclear power plants and fuel reprocessing plant which has proven useful to understand spikes in expected ratio shifts at HEI: they are associated with very high influence from Phillipsburg 2 - located only 30 km from HEI - due to unusually high sensitivity to the area within 100 km of the station. Decisions on whether the spikes are realistic, or if they should be excluded from analyzes, can be based partly on this information. Also, the French nuclear power plant La Hauge has proven to be a facility that has significant influence on the ratio at HEI with second highest influence after Phillipsburg 2 in terms of annual averages year 2016. The sensitivity to the cell containing this facility, located 900 km away, is never high. However, the huge quantity of radiocarbon emissions relative to the other facilities (19100 TBq compared to Phillipsburg 2's 89 TBq year 2016) meant that even very limited sensitivity to the cell will significantly shift the ratio at HEI and many other stations located all over Europe. This also meant a significant influence more often than facilities much closer to HEI (see *Fig. 31*). La Hauge was not considered in the paper by Kuderer et al. (2018) where only the five

nuclear power plants within 40 km of HEI were considered in terms of influence on the ratio. This is likely partly the explanation for the different result in the paper: between 2012 to 2014 the mean values were 0.44 ‰ ± 0.32 ‰, compared to 2.3‰ in the results of this thesis (see *Table 5*). Another reason might be the high spikes of contribution from Phillipsburg 2 associated with strong local contribution predicted by the STILT transport model. In the Kuderer paper, HYSPLIT dispersion model was used. The ability to break down the influence into contribution from different facilities can also be applied to the 2000+ facilities in the E-PRTR database that have been used to calculate the influence of point source emissions on the CO<sub>2</sub> concentrations at the stations. Time series showing point source emission influences also contain spikes that in turn could be better explained. Whereas yet to be implemented permanently in the Notebook, it has proven useful to explain high model overestimates at HEI which is discussed in a later paragraph.

All output values from multiplying data with footprints have uncertainties that are stemming from several different sources. First, the footprints are the result of a best guess regarding how air particles move depending on current meteorological conditions, which are estimated based on interpolation of measurements both in space and time. The natural randomness of particle movements is also simulated as an acting force in the estimation of each particle's trajectory. There is no telling exactly how well the simulated transport at any given time have succeeded in representing the reality, but big differences in resulting values depending on which transport model are used indicate that it is hard to generate footprints with high precision (Peylin et al., 2011). When the footprints are multiplied by any of the ancillary data layers representing population, land cover, point source emissions of CO<sub>2</sub> or radiocarbon, the uncertainties associated with reducing the complex reality to a map representation is an additional factor. Whereas population is relatively easily represented because it is simply a count of people living within the cell, land cover is associated with greater uncertainties: the minimum mapping unit is 0.25 km<sup>2</sup>, and any islands of land cover with smaller areas were simply not accounted for in the CORINE 2018 land cover. For linear elements, the corresponding value is a width of 100 m. The processing of land cover into the cell size and extent of the footprints resulted in one total km<sup>2</sup> value for each land cover class per cell, and did hence not add additional uncertainties. To calculate the area values with best possible accuracy without using more than one projection, Lambert azimuthal equal-area Europe was used. The footprint cell resolution is defined in degrees which cause for big differences in cell area between northern and southern cells in the domain, ranging between 38 and 108 km<sup>2</sup>. This means that point source emissions of radiocarbon and CO<sub>2</sub> are distributed over increasingly bigger area moving south in the domain. If bigger cell sizes lead to worse estimates of contribution for the emissions at the stations is unsure, but it will definitely slightly influence the results. In the processing of the point source emission data, the emissions from a facility was simply assigned to the cell in which it was contained. An option that could possibly reduce uncertainties slightly would be to interpolate the emissions so they can be distributed into more than one cell. This would be more realistic especially for facilities located close to the edge of a cell, or in a corner where the emissions could then be distributed into four cells rather than one. To estimate the anthropogenic emissions' influence on the CO<sub>2</sub> concentrations at the stations, the footprints have been multiplied by an underlying emission database. Many different datasets have been used to create the EDGAR dataset used in the STILT model setup, with additional datasets used as proxies to spatially disaggregate

data that has been reported as a country total into  $0.1^{\circ} \times 0.1^{\circ}$  cells. Some data, such as point source emissions, have higher precision than other. Additional uncertainties are added from distributing the resulting annual dataset over the year using seasonal and daily cycles for different emission categories – especially since the same cycles are used for the whole domain despite differences in emission patterns for the individual countries. In terms of the biospheric component, the values are generated by a separate model – VPRM – which estimates the respiration and photosynthesis. Model parameters include current conditions such as temperature, shortwave radiation and satellite derived parameters that are used to estimate light-use efficiency. Other necessary parameters are associated with land cover types and have been calibrated based on measured NEE associated with each land cover type.

Modelled CO<sub>2</sub> concentrations at the stations are subjected to uncertainties associated with the footprints, the emission inventory, and the biospheric component. The result of these uncertainties are differences between modelled and measured concentrations, referred to as model data differences. These have been quantified for 15 stations that currently have measured CO<sub>2</sub> concentrations available for download. Yearly averages of the model data differences indicate that model underestimates are more common than overestimates with negative annual values for 13/15 stations year 2017. The model data differences have in turn been averaged by hour and month which point to differences in the STILT mode system's ability to correctly estimate CO<sub>2</sub> concentration. A count of model data differences above or below one standard deviation of the model data differences have also been used to characterize the different month's and hour's model performance. One advantage of using standard deviation values is that possible trends become clear for stations with low variability. Furthermore, in terms of the average values, high and low model data differences take each other out and indicate better model performance than is true. Daytime estimates are generally more accurate than night-time estimates, and the same is true for the summer months compared to the winter months. Many stations have average underestimates for all months, whereas other stations, including HEI, PRS, HPB and JFJ, have relatively high average overestimates for some month, and high underestimates for other months. For the three latter stations, the average overestimates are in the winter months. The model data differences have also been used to exemplify how footprints can be subset based on model data difference to help explain poor model performance: e.g. for the station HEI the footprints associated with model data differences greater than one standard deviation (35.52 ppm) higher than the measured concentrations were aggregated and compared to the average footprint for HEI year 2017. The resulting relative footprint map's cell values represent the percentage differences between the average footprint and the average of the subset of footprints to identify possible spatial trends. The area immediately west of HEI had high cell values with the highest being 1600%, indicating its very strong association with overestimates. In turn, large emission quantities from a power plant within that cell was identified to contribute with an average of 37 ppm to the modelled CO<sub>2</sub> concentrations associated with the footprints with overestimates. This is likely a big source of error because power plants emissions, despite high emission towers and warm emissions that rise fast, are assumed to mix with air traveling within the lower half of the PBL.

All the results produced for the purposes of this thesis can be reproduced for any user defined date range. Analysis can also be performed on one specific footprint, or

subset group of footprints. High or low resulting values can possibly be explained, leaving the users more informed of the processes behind how they were generated. This was the case in the example of spikes in radiocarbon, as well as the high model overestimates, at HEI. The users are also empowered by the option to exclude subset of footprints such as those associated with high over- or underestimates resulting from uncertainties in the STILT model system. Exclusion or inclusion could also be based on any of the other characterizations such as land cover breakdown within each footprint. Furthermore, more stations can be added to the analysis and be presented along with the values of the current stations or separately. These stations can be within the ICOS station network, from a different network, or simply be a hypothetical station. Hypothetical station testing can be valuable when planning for new station locations in the growing ICOS network. Different locations can be characterized and compared to the current stations, and in turn the best location given the planned station's intended purpose can be selected. Furthermore, the prepared ancillary data – population, point source emissions, radiocarbon and land cover – can be used within future Notebooks.

## 6. Conclusions

Many different aspects of the 31 stations' footprint areas have been considered to offer a general characterization of what types of areas potentially influences the measurements at the stations. The foundation for all other analysis is the footprints themselves, and a couple of specific research questions regard how their spatial characteristics change between stations, as well as between footprints at the individual stations. All three-hourly footprints for year 2017 were used to generate averages in total sensitivity. Total sensitivity is the resulting values from all footprint cells added together and represents how much of the air travelling to the station did so in the area closest to the surface. When the air travels here, its composition is subject to influence from surface fluxes and emissions of CO<sub>2</sub>. This also means potentially higher values from multiplying the footprints with ancillary data such as population and radiocarbon emissions. The difference between the station with lowest total average sensitivity (PRS) and the highest (LMP) was almost 400%. However, 21 of the stations have relatively similar values. Another aspect is how much of the sensitivity is to the area close to the station, as opposed to far away. Considering all stations, on average 48% of the sensitivity was to the area within 300 km of the stations. However, for all average values there are large differences between individual footprints, both with regards to total sensitivity and sensitivity within certain distances of the stations. The total average sensitivity values for the different stations have standard deviation values ranging from 49% of the average at KRE to 124% at LMP. These differences are what in turn leads to large variability in all annual values derived from multiplying the footprints with underlying data.

The results from multiplying the footprints with data on land cover is connected to more specific research questions: what are the main land cover classes the stations are sensitive to, and how well does this represent the breakdown of land cover in the whole domain? For year 2017, ocean was the most significant class with an average for all stations of 31.3%, ranging from 11.4% for PAL to 66.0% for HEL. Second was staple crops, not including rice, with a station average of 19.5% with higher values associated with stations in central Europe. Third was coniferous forests with 11.1% followed by broad leaved forests, 7.6%, and pastures, 7.1%. These are also the major land cover types within the domain, but 49% oceans within the domain make for lower values for all other land cover classes. The large extent of many footprints is what make the annual footprint areas of the combined stations representative of the domain.

The final research question regards the agreement between modelled and measured concentrations, and how the model performance might be related to what hour of the day, or month, it is. Some general patterns, such as better estimates during daytime and summer, have been identified for many of the stations. Also, underestimating the measured concentrations has been identified as the most frequent error when estimating modelled concentrations with the STILT model system. The model data differences can also be used to subset footprints, either to analyze what might have caused the differences, or to exclude them from further analysis. For HEI, footprints associated with high overestimates could be connected to high local influences and modelled concentration values hiked up by facilities emitting CO<sub>2</sub> close to the station. Information about structural errors associated with different hours and/or months can be used in the process of improving the STILT model system. For instance,

anthropogenic emissions in the EDGAR database stem from yearly quantities that have been distributed in time with the same time-profiles for the different emission sectors for all European countries. The process of tailoring these to the different countries could include running this Notebook.

With a continuously increasing understanding of the natural fluxes in the carbon cycle, more precise emission statistics, and better transport models simulating the air travel, modelled CO<sub>2</sub> concentrations can be expected to get closer to the reality. Also, planned measurements of radiocarbon at the ICOS stations will mean an additional way to estimate the fossil fuel component within the footprints independent of emission statistics. As a result, better verification of reported CO<sub>2</sub> emissions, such as in national emission inventories, will be possible. In turn, different actors can be held accountable for their emissions, whether it be paying a carbon tax or tracking mitigation efforts in accordance with international agreements. Furthermore, to understand the changing climate's effect on the carbon cycle – such as the carbon reservoirs' abilities to continue to absorb about 50% of the anthropogenic emissions to the atmosphere – it is important to be able to model the carbon cycle with confidence. It will give us a better idea about what the future global CO<sub>2</sub> concentrations, and associated consequences, might be.

## References

- Aalto T., Hatakka J., Kouznetsov R. and Stanislawska K. (2015): Background and anthropogenic influences on atmospheric CO<sub>2</sub> concentrations measured at Pallas: Comparison of two models for tracing air mass history. *Boreal Env. Res.*, 20: 213–226.
- Andres, R. J., Boden, T. A., and Higdon, D. M. (2016): Gridded uncertainty in fossil fuel carbon dioxide emission maps, a CDIAC example, *Atmos. Chem. Phys.*, 16: 14979-14995, <https://doi.org/10.5194/acp-16-14979-2016>
- BP: Statistical Review of World Energy 2018 (2019) – data workbook, available at: <https://www.bp.com/content/dam/bp/business-sites/en/global/corporate/xlsx/energy-economics/statistical-review/bp-stats-review-2019-all-data.xlsx>
- CLC2018 Technical Guidelines (2017). Retrieved 13 March 2019, from <https://land.copernicus.eu/user-corner/technical-library>
- Cooperative Global Atmospheric Data Integration Project. (2019). Multi-laboratory compilation of atmospheric carbon dioxide data for the period 1957-2018; obspack\_co2\_1\_GLOBALVIEWplus\_v5.0\_2019\_08\_12 [Data set]. NOAA Earth System Research Laboratory, Global Monitoring Division. <https://doi.org/10.25925/20190812>
- CORINE land cover (2018). Land cover data. Downloaded 13 March 2019, from <https://land.copernicus.eu/pan-european/corine-land-cover/clc2018>
- Denier van der Gon, H. D., Hendriks, C., Kuenen, J., Segers, A., and Visschedijk, A. (2011): Description of current temporal emission patterns and sensitivity of predicted AQ for temporal emission patterns, *TNP Report, EU FP7 MACC deliverable report D\_D-EMIS\_1.3*, Retrieved from: [https://gmes-atmosphere.eu/documents/deliverables/d-emis/MACC\\_TNO\\_del\\_1\\_3\\_v2.pdf](https://gmes-atmosphere.eu/documents/deliverables/d-emis/MACC_TNO_del_1_3_v2.pdf)
- European Centre for Medium-Range Weather Forecasts (2006): ECMWF operational analysis: Assimilated Data. NCAS British Atmospheric Data Centre <http://catalogue.ceda.ac.uk/uuid/c46248046f6ce34fc7660a36d9b10a71>
- E-PRTR (2019). Pollutant release data. Downloaded 9 May 2019, from <https://www.eea.europa.eu/data-and-maps/data/member-states-reporting-art-7-under-the-european-pollutant-release-and-transfer-register-e-prtr-regulation-22>
- Geels, C., Gloor, M., Ciais, P., Bousquet, P., Peylin, P., Vermeulen, A. T., Dargaville, R., Aalto, T., Brandt, J., Christensen, J. H., Frohn, L. M., Haszpra, L., Karstens, U., Rödenbeck, C., Ramonet, M., Carboni, G., and Santaguida, R. (2007): Comparing atmospheric transport models for future regional inversions over Europe – Part 1: mapping the atmospheric CO<sub>2</sub> signals. *Atmos. Chem. Phys.*, 7: 3461–3479, <https://doi.org/10.5194/acp-7-3461-2007>

- GEOSTAT (2011). Population data. Downloaded 17 October 2019, from <https://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/population-distribution-demography/geostat>
- GEOSTAT 1A – Representing Census data in a European population grid (2011) Retrieved June 8 2019, from: <https://www.efgs.info/wp-content/uploads/geostat/1a/GEOSTAT1A-final-report.pdf>
- Henne, S., Brunner, D., Folini, D., Solberg, S., Klausen, J., and Buchmann, B. (2010): *Assessment of parameters describing representativeness of air quality in-situ measurement sites*, *Atmos. Chem. Phys.*, 10, 3561-3581, <https://doi.org/10.5194/acp-10-3561-2010>
- ICOS (2019): ICOS Handbook, Knowledge through observations. Eds: S. Ashraful Alam J-M Rintala, K Ahlgren and E. Juurola. Publisher ICOS ERIC ISBN 978-952-94-1950-0 Version 2. Retrieved 1 December 2019, from: <https://www.icos-ri.eu/sites/default/files/cmis/ICOS%20Handbook%202019.pdf>
- ICOS Carbon Portal. STITL result viewer. Retrieved 17 October 2019, from <https://stilt.icos-cp.eu/viewer/>
- ICOS Carbon Portal. ICOS Stations. Retrieved September 12 2019, from <https://www.icos-cp.eu/stations>
- ICOS RI (2019): ICOS Atmospheric Greenhouse Gas Mole Fractions of CO<sub>2</sub>, CH<sub>4</sub>, CO, <sup>14</sup>CO<sub>2</sub> and Meteorological Observations September 2015 - April 2019 for 19 stations (49 vertical levels), final quality-controlled Level 2 data. <https://doi.org/10.18160/ce2r-cc91>
- ICOS Strategy (2019). <https://doi.org/10.18160/TV1J-4XQY>
- IPCC (2018): Global Warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty. Eds: Masson-Delmotte, V., P. Zhai, H.-O. Pörtner, D. Roberts, J. Skea, P.R. Shukla, A. Pirani, W. Moufouma-Okia, C. Péan, R. Pidcock, S. Connors, J.B.R. Matthews, Y. Chen, X. Zhou, M.I. Gomis, E. Lonnoy, T. Maycock, M. Tignor, and T. Waterfield (eds.)]. In Press.
- Janssens-Maenhout, G., Crippa, M., Guizzardi, D., Muntean, M., Schaaf, E., Dentener, F., Bergamaschi, P., Pagliari, V., Olivier, J. G. J., Peters, J. A. H. W., van Aardenne, J. A., Monni, S., Doering, U., and Petrescu, A. M. R. (2017): EDGAR v4.3.2 Global Atlas of the three major Greenhouse Gas Emissions for the period 1970–2012, *Earth Syst. Sci. Data Discuss.*, <https://doi.org/10.5194/essd-2017-79>
- Janssens-Maenhout, G., Pagliari, V., Guizzardi, D., and Muntean, M.: Global emission inventories in the Emission Database for Global Atmospheric

- Research (EDGAR) – Manual (I): Gridding: EDGAR emissions distribution on global grid-maps, JRC Report, EUR 25785 EN, ISBN 978-92-79-28283-6, doi.10.2788/81454, 2013.
- Jung, M., Henkel, K., Herold, M. and Churkina, G. (2006): Exploiting synergies of global land cover products for carbon cycle modeling, *Remote Sensing of Environment*, 101(4), 534–553, <https://doi.org/10.1016/j.rse.2006.01.020>
- Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B., Bussonnier, M., Frederic, J., Kelley, K., Hamrick, J., Grout, J., Corlay, S., Ivanov, P., Avila, D., Abdalla, S., Willing, C. and Jupyter development team. (2016): Jupyter Notebooks – a publishing format for reproducible computational workflows. Loizides, Fernando and Schmidt, Birgit (eds.) *In Positioning and Power in Academic Publishing: Players, Agents and Agendas*. IOS Press., 87-90, <https://doi.org/10.3233/978-1-61499-649-1-87>
- Kuderer, M., Hammer, S., Levin, I. (2018): The influence of  $^{14}\text{CO}_2$  releases from regional nuclear facilities at the Heidelberg  $^{14}\text{CO}_2$  sampling site (1986–2014), *Atmos. Chem. Phys.*, 18, 7951–7959, <https://doi.org/10.5194/acp-18-7951-2018>
- Lansø, A. S., Smallman, T. L., Christensen, J. H., Williams, M., Pilegaard, K., Sørensen, L., Geels, C. (2019): Simulating the atmospheric  $\text{CO}_2$  concentration across the heterogeneous landscape of Denmark using a coupled atmosphere–biosphere mesoscale model system, *Biogeosciences*, 16, 1505–1524, <https://doi.org/10.5194/bg-16-1505-2019>
- Lin, J. C., C. Gerbig, S. C. Wofsy, A. E. Andrews, B. C. Daube, K. J. Davis, and C. A. Grainger (2003): A near-field tool for simulating the upstream influence of atmospheric observations: The Stochastic Time-Inverted Lagrangian Transport (STILT) model, *J. Geophys. Res.*, 108(D16), 4493, <https://doi.org/10.1029/2002JD003161>
- Lin, J. C., M. R. Pejam, E. Chan, S. C. Wofsy, E. W. Gottlieb, H. A. Margolis, and J. H. McCaughey (2011): Attributing uncertainties in simulated biospheric carbon fluxes to different error sources, *Global Biogeochem. Cycles*, 25, GB2018, <https://doi.org/10.1029/2010GB003884>,
- Mahadevan, P., Wofsy, S. C., Matross, D. M., Xiao, X., Dunn, A. L., Lin, J. C., Gerbig, C., Munger, J. W., Chow, V. Y., and Gottlieb, E. W. (2008): A satellite-based biosphere parameterization for net ecosystem  $\text{CO}_2$  exchange: Vegetation Photosynthesis and Respiration Model (VPRM), *Global Biogeochem. Cy.*, 22, GB2005, <https://doi.org/10.1029/2006GB002735>
- Natural Earth. Ocean data. Version 4.1.0. Retrieved 7 June 2019, from <https://www.naturalearthdata.com/downloads/10m-physical-vectors/10m-ocean/>
- Peylin, P., Houweling, S., Krol, M. C., Karstens, U., Rödenbeck, C., Geels, C., Vermulen, A., Badawy, B., Aulagnier, C., Pregger, T., Delage, F., Pieterse, G.,

- Ciais, P., and Heimann, M. (2011): Importance of fossil fuel emission uncertainties over Europe for CO<sub>2</sub> modeling: model intercomparison, *Atmos. Chem. Phys.*, *11*, 6607–6622, <https://doi.org/10.5194/acp-11-6607-2011>
- Steinbach, J., Gerbig, C., Rödenbeck, C., Karstens, U., Minejima, C., and Mukai, H. (2007): The CO<sub>2</sub> release and Oxygen uptake from Fossil Fuel Emission Estimate (COFFEE) dataset: effects from varying oxidative ratios, *Atmos. Chem. Phys.*, *11*, 6855–6870, <https://doi.org/10.5194/acp-11-6855-2011>
- Uglietti, C., Leuenberger, M., and Brunner, D. (2011): European source and sink areas of CO<sub>2</sub> retrieved from Lagrangian transport model interpretation of combined O<sub>2</sub> and CO<sub>2</sub> measurements at the high alpine research station Jungfraujoch, *Atmos. Chem. Phys.*, *11*, 8017-8036, <https://doi.org/10.5194/acp-11-8017-2011>
- Vardag, S. N., Gerbig, C., Janssens-Maenhout, G., and Levin, I. (2015): Estimation of continuous anthropogenic CO<sub>2</sub>: model-based evaluation of CO<sub>2</sub>, CO, δ<sup>13</sup>C(CO<sub>2</sub>) and Δ<sup>14</sup>C(CO<sub>2</sub>) tracer methods, *Atmospheric Chemistry and Physics*, *15*, 12705–12729, <https://doi.org/10.5194/acp-15-12705-2015>
- Zazzeri, G., Acuña Yeomans, E., and Graven, H. (2018): Global and Regional Emissions of Radiocarbon from Nuclear Power Plants from 1972 to 2016. *Radiocarbon*, *60*(4), 1067-1081. <https://doi.org/10.1017/RDC.2018.42>
- Karstens, U. (2019): Global anthropogenic CO<sub>2</sub> emissions for 2006-2019 based on EDGARv4.3 and BP statistics 2016. doi:10.18160/y9qv-s113
- Zazzeri et al. Radiocarbon emissions. Retrieved June 6 2019 from <https://static.cambridge.org/content/id/urn:cambridge.org:id:article:S0033822218000425/resource/name/S0033822218000425sup001.csv>

#### Personal Communications

Karstens, U., Researcher, ICOS Carbon Portal, Lund University, Sweden.

## Appendices

### Appendix 1.

“The Jupyter Notebook is an open-source web application that allows you to create and share documents that contain live code, equations, visualizations and narrative text.” (<https://jupyter.org/>)

The output Jupyter Notebooks from this thesis are uploaded as separate HTML files on LUP Student Papers:

Sensitivity area evaluation Notebook:

<http://lup.lub.lu.se/luur/download?func=downloadFile&recordId=9007298&fileId=9007311>

Ancillary data Notebook:

<http://lup.lub.lu.se/luur/download?func=downloadFile&recordId=9007298&fileId=9007303>

Breakdown modelled concentrations Notebook:

<http://lup.lub.lu.se/luur/download?func=downloadFile&recordId=9007298&fileId=9007305>

Modelled concentrations compared to measured concentrations year 2017 Notebook:

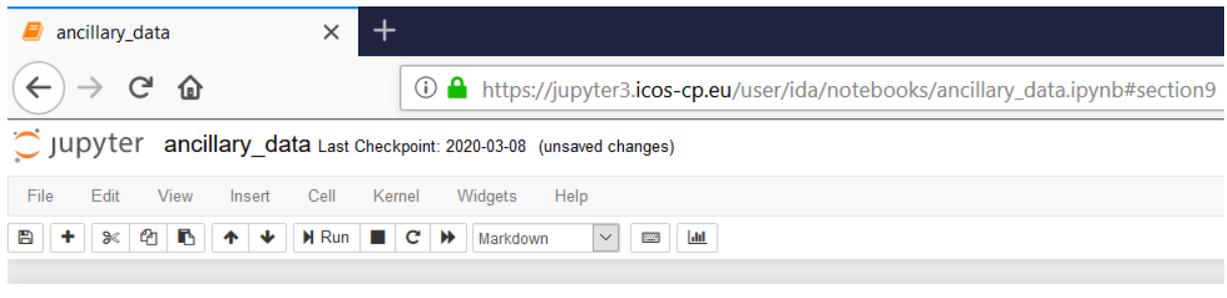
<http://lup.lub.lu.se/luur/download?func=downloadFile&recordId=9007298&fileId=9007307>

Radiocarbon emissions influence Notebook:

<http://lup.lub.lu.se/luur/download?func=downloadFile&recordId=9007298&fileId=9007309>

In the future it will be possible to use them as a service at the Carbon Portal.

Notebooks are shareable documents built from cells which are either filled with Markdown language or program code (Python 3) with output from running code output directly below the cell. This is exemplified by screenshots from the “Ancillary data” Notebook:



## Ancillary data

[Import modules and functions](#)

[Select which stations to include in the analysis](#)

[Set the date range and the hour\(s\) for the analysis](#)

1. Land cover data: CORINE
  - A. [Import land cover data from NetCDF](#)
  - B. [Average breakdown \(abosolute and percent\) of what land cover is seen by the stations](#)
  - C. [Spatial distribution of land cover classes seen by the selected stations](#)
4. Population data
  - A. [Import population data from NetCDF](#)
  - B. [Bar grap with average population interaction. option to show influence map for each station](#)
3. Point source emission
  - A. [Import point source emission data from NetCDF](#)
  - B. [Bar grap with average point source emission contribution to concentration. option to show influence map for each station](#)

*Initial cell with markdown language introducing the Notebook. The underlined blue texts are links to different parts of the document.*

**Run this cell to finalize the land cover selection**

```
In [22]: only_selected_land_cover_classes_spatial_dist=land_cover_classes_spatial_dist.value
list_only_selected_land_cover_classes_spatial_dist=[]
for value in only_selected_land_cover_classes_spatial_dist:
    list_only_selected_land_cover_classes_spatial_dist.append(value)

print('Your selection is: ', list_only_selected_land_cover_classes_spatial_dist)

Your selection is: ['Cropland: Staple except rice', 'Broad leaved forest', 'Coniferous forest']
```

```
In [23]: #create a dictionary where the key is the land use class/ute classification.
dictionary = {'Urban': urban, 'Industrial': industrial, 'Roads and railroads': road_and_rail,
             'Ports and airports': ports_and_apirports, 'Dump sites': dump_sites,
             'Cropland: Staple except rice': staple_cropland_not_rice, 'Rice fields': rice_fields,
             'Cropland: fruits, berries, grapes and olives': cropland_fruit_berry_grapes_olives,
             'Pastures': pastures, 'Broad leaved forest': broad_leaved_forest,
             'Coniferous forest': coniferous_forest, 'Mixed forest': mixed_forest,
             'Natural grasslands': natural_grasslands, 'Transitional woodlands and shrub': transitional_woodland_shrub,
             'Natural bare areas': bare_natural_areas,
             'Glaciers and prepetual snow': glaciers_prepetual_snow, 'Wet areas': wet_area,
             'Inland water bodies': inland_water_bodies, 'Ocean': oceans}

matplotlib.rcParams.update({'font.size': 17})

lon=all_corine_classes.variables['lon'][:]
lat=all_corine_classes.variables['lat'][:]

station_iterator=iter(all_stations)

for station in station_iterator:

    for selection in list_only_selected_land_cover_classes_spatial_dist:

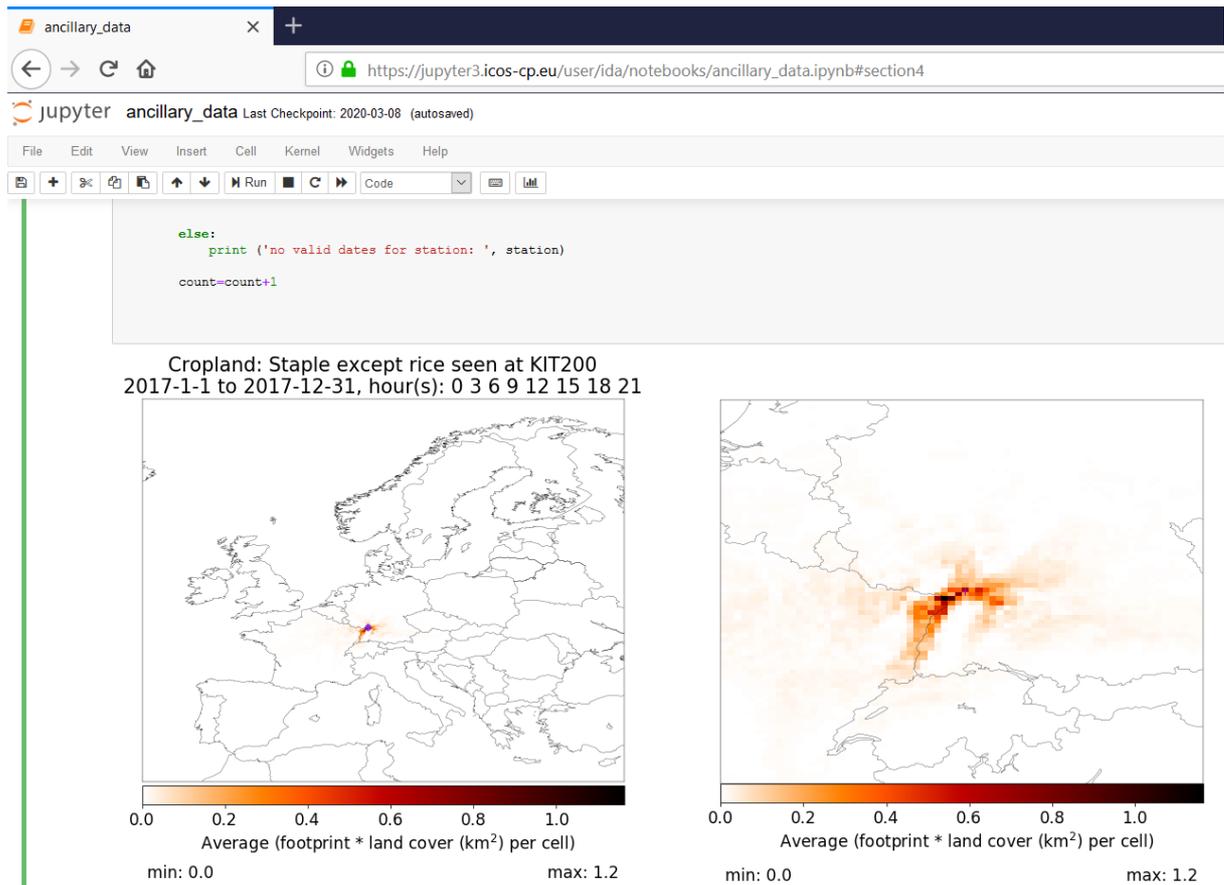
        fp_selected_land_cover=dictionary[selection]

        #want to look at the maps for all the stations
        list_selected_land_cover=[]
        list_stations=[]

        station_iterator=iter(all_stations)
```

*Two of the cells in the land cover data section of the Notebook. The code is written in Python 3.*

When a cell is “run”, meaning when the code of a particular cell is executed, the results will appear at the bottom of the cell. Upon running code, the user is sometimes asked for input. For instance, what station(s) to analyze and for which date range need to be specified in all Notebooks.



*Part of the output from running a cell in the land cover data section of the Notebook. Station Karlsruhe (KIT200) had been selected, and the map represents the spatial distribution of the average sensitivity to land cover type “Cropland: staple except rice”.*

**Master Thesis in Geographical Information Science**

1. *Anthony Lawther*: The application of GIS-based binary logistic regression for slope failure susceptibility mapping in the Western Grampian Mountains, Scotland (2008).
2. *Rickard Hansen*: Daily mobility in Grenoble Metropolitan Region, France. Applied GIS methods in time geographical research (2008).
3. *Emil Bayramov*: Environmental monitoring of bio-restoration activities using GIS and Remote Sensing (2009).
4. *Rafael Villarreal Pacheco*: Applications of Geographic Information Systems as an analytical and visualization tool for mass real estate valuation: a case study of Fontibon District, Bogota, Columbia (2009).
5. *Siri Oestreich Waage*: a case study of route solving for oversized transport: The use of GIS functionalities in transport of transformers, as part of maintaining a reliable power infrastructure (2010).
6. *Edgar Pimiento*: Shallow landslide susceptibility – Modelling and validation (2010).
7. *Martina Schäfer*: Near real-time mapping of floodwater mosquito breeding sites using aerial photographs (2010).
8. *August Pieter van Waarden-Nagel*: Land use evaluation to assess the outcome of the programme of rehabilitation measures for the river Rhine in the Netherlands (2010).
9. *Samira Muhammad*: Development and implementation of air quality data mart for Ontario, Canada: A case study of air quality in Ontario using OLAP tool. (2010).
10. *Fredros Oketch Okumu*: Using remotely sensed data to explore spatial and temporal relationships between photosynthetic productivity of vegetation and malaria transmission intensities in selected parts of Africa (2011).
11. *Svajunas Plunge*: Advanced decision support methods for solving diffuse water pollution problems (2011).
12. *Jonathan Higgins*: Monitoring urban growth in greater Lagos: A case study using GIS to monitor the urban growth of Lagos 1990 - 2008 and produce future growth prospects for the city (2011).
13. *Mårten Karlberg*: Mobile Map Client API: Design and Implementation for Android (2011).
14. *Jeanette McBride*: Mapping Chicago area urban tree canopy using color infrared imagery (2011).
15. *Andrew Farina*: Exploring the relationship between land surface temperature and vegetation abundance for urban heat island mitigation in Seville, Spain (2011).
16. *David Kanyari*: Nairobi City Journey Planner: An online and a Mobile Application (2011).
17. *Laura V. Drews*: Multi-criteria GIS analysis for siting of small wind power plants - A case study from Berlin (2012).
18. *Qaisar Nadeem*: Best living neighborhood in the city - A GIS based multi criteria evaluation of ArRiyadh City (2012).

19. *Ahmed Mohamed El Saeid Mustafa*: Development of a photo voltaic building rooftop integration analysis tool for GIS for Dokki District, Cairo, Egypt (2012).
20. *Daniel Patrick Taylor*: Eastern Oyster Aquaculture: Estuarine Remediation via Site Suitability and Spatially Explicit Carrying Capacity Modeling in Virginia's Chesapeake Bay (2013).
21. *Angeleta Oveta Wilson*: A Participatory GIS approach to *unearthing* Manchester's Cultural Heritage 'gold mine' (2013).
22. *Ola Svensson*: Visibility and Tholos Tombs in the Messenian Landscape: A Comparative Case Study of the Pylion Hinterlands and the Soulima Valley (2013).
23. *Monika Ogden*: Land use impact on water quality in two river systems in South Africa (2013).
24. *Stefan Rova*: A GIS based approach assessing phosphorus load impact on Lake Flaten in Salem, Sweden (2013).
25. *Yann Buhot*: Analysis of the history of landscape changes over a period of 200 years. How can we predict past landscape pattern scenario and the impact on habitat diversity? (2013).
26. *Christina Fotiou*: Evaluating habitat suitability and spectral heterogeneity models to predict weed species presence (2014).
27. *Inese Linuza*: Accuracy Assessment in Glacier Change Analysis (2014).
28. *Agnieszka Griffin*: Domestic energy consumption and social living standards: a GIS analysis within the Greater London Authority area (2014).
29. *Brynja Guðmundsdóttir*: Detection of potential arable land with remote sensing and GIS - A Case Study for Kjósarhreppur (2014).
30. *Oleksandr Nekrasov*: Processing of MODIS Vegetation Indices for analysis of agricultural droughts in the southern Ukraine between the years 2000-2012 (2014).
31. *Sarah Tressel*: Recommendations for a polar Earth science portal in the context of Arctic Spatial Data Infrastructure (2014).
32. *Caroline Gevaert*: Combining Hyperspectral UAV and Multispectral Formosat-2 Imagery for Precision Agriculture Applications (2014).
33. *Salem Jamal-Uddeen*: Using GeoTools to implement the multi-criteria evaluation analysis - weighted linear combination model (2014).
34. *Samanah Seyedi-Shandiz*: Schematic representation of geographical railway network at the Swedish Transport Administration (2014).
35. *Kazi Masel Ullah*: Urban Land-use planning using Geographical Information System and analytical hierarchy process: case study Dhaka City (2014).
36. *Alexia Chang-Wailing Spitteler*: Development of a web application based on MCDA and GIS for the decision support of river and floodplain rehabilitation projects (2014).
37. *Alessandro De Martino*: Geographic accessibility analysis and evaluation of potential changes to the public transportation system in the City of Milan (2014).
38. *Alireza Mollasalehi*: GIS Based Modelling for Fuel Reduction Using Controlled Burn in Australia. Case Study: Logan City, QLD (2015).
39. *Negin A. Sanati*: Chronic Kidney Disease Mortality in Costa Rica; Geographical Distribution, Spatial Analysis and Non-traditional Risk Factors (2015).

40. *Karen McIntyre*: Benthic mapping of the Bluefields Bay fish sanctuary, Jamaica (2015).
41. *Kees van Duijvendijk*: Feasibility of a low-cost weather sensor network for agricultural purposes: A preliminary assessment (2015).
42. *Sebastian Andersson Hylander*: Evaluation of cultural ecosystem services using GIS (2015).
43. *Deborah Bowyer*: Measuring Urban Growth, Urban Form and Accessibility as Indicators of Urban Sprawl in Hamilton, New Zealand (2015).
44. *Stefan Arvidsson*: Relationship between tree species composition and phenology extracted from satellite data in Swedish forests (2015).
45. *Damián Giménez Cruz*: GIS-based optimal localisation of beekeeping in rural Kenya (2016).
46. *Alejandra Narváez Vallejo*: Can the introduction of the topographic indices in LPJ-GUESS improve the spatial representation of environmental variables? (2016).
47. *Anna Lundgren*: Development of a method for mapping the highest coastline in Sweden using breaklines extracted from high resolution digital elevation models (2016).
48. *Oluwatomi Esther Adejoro*: Does location also matter? A spatial analysis of social achievements of young South Australians (2016).
49. *Hristo Dobrev Tomov*: Automated temporal NDVI analysis over the Middle East for the period 1982 - 2010 (2016).
50. *Vincent Muller*: Impact of Security Context on Mobile Clinic Activities A GIS Multi Criteria Evaluation based on an MSF Humanitarian Mission in Cameroon (2016).
51. *Gezahagn Negash Seboka*: Spatial Assessment of NDVI as an Indicator of Desertification in Ethiopia using Remote Sensing and GIS (2016).
52. *Holly Buhler*: Evaluation of Interfacility Medical Transport Journey Times in Southeastern British Columbia. (2016).
53. *Lars Ole Grottenberg*: Assessing the ability to share spatial data between emergency management organisations in the High North (2016).
54. *Sean Grant*: The Right Tree in the Right Place: Using GIS to Maximize the Net Benefits from Urban Forests (2016).
55. *Irshad Jamal*: Multi-Criteria GIS Analysis for School Site Selection in Gorno-Badakhshan Autonomous Oblast, Tajikistan (2016).
56. *Fulgencio Sanmartín*: Wisdom-volkano: A novel tool based on open GIS and time-series visualization to analyse and share volcanic data (2016).
57. *Nezha Acil*: Remote sensing-based monitoring of snow cover dynamics and its influence on vegetation growth in the Middle Atlas Mountains (2016).
58. *Julia Hjalmarsson*: A Weighty Issue: Estimation of Fire Size with Geographically Weighted Logistic Regression (2016).
59. *Mathewos Tamiru Amato*: Using multi-criteria evaluation and GIS for chronic food and nutrition insecurity indicators analysis in Ethiopia (2016).
60. *Karim Alaa El Din Mohamed Soliman El Attar*: Bicycling Suitability in Downtown, Cairo, Egypt (2016).
61. *Gilbert Akol Echelai*: Asset Management: Integrating GIS as a Decision Support Tool in Meter Management in National Water and Sewerage Corporation (2016).
62. *Terje Slinning*: Analytic comparison of multibeam echo soundings (2016).

63. *Gréta Hlín Sveinsdóttir*: GIS-based MCDA for decision support: A framework for wind farm siting in Iceland (2017).
64. *Jonas Sjögren*: Consequences of a flood in Kristianstad, Sweden: A GIS-based analysis of impacts on important societal functions (2017).
65. *Nadine Raska*: 3D geologic subsurface modelling within the Mackenzie Plain, Northwest Territories, Canada (2017).
66. *Panagiotis Symeonidis*: Study of spatial and temporal variation of atmospheric optical parameters and their relation with PM 2.5 concentration over Europe using GIS technologies (2017).
67. *Michaela Bobeck*: A GIS-based Multi-Criteria Decision Analysis of Wind Farm Site Suitability in New South Wales, Australia, from a Sustainable Development Perspective (2017).
68. *Raghdaa Eissa*: Developing a GIS Model for the Assessment of Outdoor Recreational Facilities in New Cities Case Study: Tenth of Ramadan City, Egypt (2017).
69. *Zahra Khais Shahid*: Biofuel plantations and isoprene emissions in Svea and Götaland (2017).
70. *Mirza Amir Liaquat Baig*: Using geographical information systems in epidemiology: Mapping and analyzing occurrence of diarrhea in urban - residential area of Islamabad, Pakistan (2017).
71. *Joakim Jörwall*: Quantitative model of Present and Future well-being in the EU-28: A spatial Multi-Criteria Evaluation of socioeconomic and climatic comfort factors (2017).
72. *Elin Haettner*: Energy Poverty in the Dublin Region: Modelling Geographies of Risk (2017).
73. *Harry Eriksson*: Geochemistry of stream plants and its statistical relations to soil- and bedrock geology, slope directions and till geochemistry. A GIS-analysis of small catchments in northern Sweden (2017).
74. *Daniel Gardevärn*: PPGIS and Public meetings – An evaluation of public participation methods for urban planning (2017).
75. *Kim Friberg*: Sensitivity Analysis and Calibration of Multi Energy Balance Land Surface Model Parameters (2017).
76. *Viktor Svanerud*: Taking the bus to the park? A study of accessibility to green areas in Gothenburg through different modes of transport (2017).
77. *Lisa-Gaye Greene*: Deadly Designs: The Impact of Road Design on Road Crash Patterns along Jamaica's North Coast Highway (2017).
78. *Katarina Jemec Parker*: Spatial and temporal analysis of fecal indicator bacteria concentrations in beach water in San Diego, California (2017).
79. *Angela Kabiru*: An Exploratory Study of Middle Stone Age and Later Stone Age Site Locations in Kenya's Central Rift Valley Using Landscape Analysis: A GIS Approach (2017).
80. *Kristean Björkmann*: Subjective Well-Being and Environment: A GIS-Based Analysis (2018).
81. *Williams Erhunmonmen Ojo*: Measuring spatial accessibility to healthcare for people living with HIV-AIDS in southern Nigeria (2018).
82. *Daniel Assefa*: Developing Data Extraction and Dynamic Data Visualization (Styling) Modules for Web GIS Risk Assessment System (WGRAS). (2018).
83. *Adela Nistora*: Inundation scenarios in a changing climate: assessing potential impacts of sea-level rise on the coast of South-East England (2018).

84. *Marc Seliger*: Thirsty landscapes - Investigating growing irrigation water consumption and potential conservation measures within Utah's largest master-planned community: Daybreak (2018).
85. *Luka Jovičić*: Spatial Data Harmonisation in Regional Context in Accordance with INSPIRE Implementing Rules (2018).
86. *Christina Kourdounouli*: Analysis of Urban Ecosystem Condition Indicators for the Large Urban Zones and City Cores in EU (2018).
87. *Jeremy Azzopardi*: Effect of distance measures and feature representations on distance-based accessibility measures (2018).
88. *Patrick Kabatha*: An open source web GIS tool for analysis and visualization of elephant GPS telemetry data, alongside environmental and anthropogenic variables (2018).
89. *Richard Alphonse Giliba*: Effects of Climate Change on Potential Geographical Distribution of *Prunus africana* (African cherry) in the Eastern Arc Mountain Forests of Tanzania (2018).
90. *Eiður Kristinn Eiðsson*: Transformation and linking of authoritative multi-scale geodata for the Semantic Web: A case study of Swedish national building data sets (2018).
91. *Niamh Harty*: HOP!: a PGIS and citizen science approach to monitoring the condition of upland paths (2018).
92. *José Estuardo Jara Alvear*: Solar photovoltaic potential to complement hydropower in Ecuador: A GIS-based framework of analysis (2018).
93. *Brendan O'Neill*: Multicriteria Site Suitability for Algal Biofuel Production Facilities (2018).
94. *Roman Spataru*: Spatial-temporal GIS analysis in public health – a case study of polio disease (2018).
95. *Alicja Miodońska*: Assessing evolution of ice caps in Suðurland, Iceland, in years 1986 - 2014, using multispectral satellite imagery (2019).
96. *Dennis Lindell Schettini*: A Spatial Analysis of Homicide Crime's Distribution and Association with Deprivation in Stockholm Between 2010-2017 (2019).
97. *Damiano Vesentini*: The Po Delta Biosphere Reserve: Management challenges and priorities deriving from anthropogenic pressure and sea level rise (2019).
98. *Emilie Arnesten*: Impacts of future sea level rise and high water on roads, railways and environmental objects: a GIS analysis of the potential effects of increasing sea levels and highest projected high water in Scania, Sweden (2019).
99. *Syed Muhammad Amir Raza*: Comparison of geospatial support in RDF stores: Evaluation for ICOS Carbon Portal metadata (2019).
100. *Hemin Tofiq*: Investigating the accuracy of Digital Elevation Models from UAV images in areas with low contrast: A sandy beach as a case study (2019).
101. *Evangelos Vafeiadis*: Exploring the distribution of accessibility by public transport using spatial analysis. A case study for retail concentrations and public hospitals in Athens (2019).
102. *Milan Sekulic*: Multi-Criteria GIS modelling for optimal alignment of roadway by-passes in the Tlokweng Planning Area, Botswana (2019).
103. *Ingrid Piirisaar*: A multi-criteria GIS analysis for siting of utility-scale photovoltaic solar plants in county Kilkenny, Ireland (2019).

104. *Nigel Fox*: Plant phenology and climate change: possible effect on the onset of various wild plant species' first flowering day in the UK (2019).
105. *Gunnar Hesch*: Linking conflict events and cropland development in Afghanistan, 2001 to 2011, using MODIS land cover data and Uppsala Conflict Data Programme (2019).
106. *Elijah Njoku*: Analysis of spatial-temporal pattern of Land Surface Temperature (LST) due to NDVI and elevation in Ilorin, Nigeria (2019).
107. *Katalin Bunyevácz*: Development of a GIS methodology to evaluate informal urban green areas for inclusion in a community governance program (2019).
108. *Paul dos Santos*: Automating synthetic trip data generation for an agent-based simulation of urban mobility (2019).
109. *Robert O' Dwyer*: Land cover changes in Southern Sweden from the mid-Holocene to present day: Insights for ecosystem service assessments (2019).
110. *Daniel Klingmyr*: Global scale patterns and trends in tropospheric NO<sub>2</sub> concentrations (2019).
111. *Marwa Farouk Elkabbany*: Sea Level Rise Vulnerability Assessment for Abu Dhabi, United Arab Emirates (2019).
112. *Jip Jan van Zoonen*: Aspects of Error Quantification and Evaluation in Digital Elevation Models for Glacier Surfaces (2020).
113. *Georgios Efthymiou*: The use of bicycles in a mid-sized city – benefits and obstacles identified using a questionnaire and GIS (2020).
114. *Haruna Olayiwola Jimoh*: Assessment of Urban Sprawl in MOWE/IBAFO Axis of Ogun State using GIS Capabilities (2020).
115. *Nikolaos Barmpas Zachariadis*: Development of an iOS, Augmented Reality for disaster management (2020).
116. *Ida Storm*: ICOS Atmospheric Stations: Spatial Characterization of CO<sub>2</sub> Footprint Areas and Evaluating the Uncertainties of Modelled CO<sub>2</sub> Concentrations (2020).