Association of air pollution and mortality in individuals with high cardiovascular risk

Inauguraldissertation
zur Erlangung des Grades eines Doktors der Medizin
des Fachbereichs Medizin
der Justus-Liebig-Universität Gießen

vorgelegt von Maitra, Rebecca Anjuli
aus Ludwigsburg

Gießen 2024

Aus dem Fachbereich Medizin der Justus-Liebig-Universität Gießen

JLU, FB11, Medizinische Klinik I, Kardiologie, Biomarkerforschung

Gutachter:

Prof. Dr. med. Till Keller

Gutachter:
Prof. Dr. med. Henning Gall

Tag der Disputation:
Dienstag, 30. September 2025

Table of contents

1. Introduction

According to the World-Health Organisation, non-communicable diseases (NCDs), including cardiovascular diseases, are among the leading causes of mortality worldwide (World Health Organization, 2022b). The establishment of the Framingham Heart study in 1948 opened a new era by identifying major risk factors for cardiovascular diseases (Dawber, Moore and Mann, 1957). The Framingham Heart study is described as the epidemiological study conducted for the largest time period (Andersson *et al.*, 2021). Over the years it has groundbreakingly established that risk factors including high blood pressure, high cholesterol, male sex, high body weight, smoking, and low physical exercise are major risk factors of cardiovascular diseases (ibid.). It thereby introduced a change of paradigm from treatment of the existent cardiovascular diseases towards the prevention of the same (Mahmood *et al.*, 2014). According to Mahmood *et al.* (2014) this lay the premise for the establishment of risk scores.

To predict the total risk of fatal cardiovascular disease events caused by multiple risk factors for a European population Conroy *et al.* (2003) developed the ESC SCORE (Systematic COronary Risk Evaluation). This score includes age, sex, smoking status, systolic blood pressure and total cholesterol to estimate the risk of fatal cardiovascular events (ibid.). This score was updated and revised to current mortality rates: For Europe, the currently proposed risk score to predict cardiovascular diseases and mortality within the guideline of the European Society of cardiology (ESC) (Visseren *et al.*, 2021) is the ESC SCORE 2. In its current form the ESC SCORE 2 was extended to predict the 10-year risk not only of fatal but also of non-fatal cardiovascular events (SCORE2 working group and ESC Cardiovascular risk collaboration, 2021). The updated version of the score is based on five variables: sex, age, current smoker status, systolic blood pressure, and non-HDL-cholesterol (ibid.). It is calculated separately for different European regions based on CVD risk (ibid.). Germany is classified within the group of countries of moderate CVD risk (ibid.). The risk variables of the ESC SCORE 2 are all of patient immanent nature.

For some decades, the association between patient-extrinsic risk factors like air pollution and mortality has been established. The WHO is estimating that 6.7 million premature deaths can be attributed to air pollution annually (World Health Organization, 2022a). Air pollution is among the most important causes of non-communicable diseases (Münzel *et al.*, 2018). Nevertheless, the consideration of patient extrinsic risk factors like air pollution might add valuable information for the prediction of cardiovascular mortality. This forms the basis for the idea behind this dissertation.

After having outlined the background concept of this dissertation, basic pathophysiological pathways of the association between air pollution and mortality are introduced. Thereafter, the aim of this study and the individual chapters of this dissertation are outlined.

1.1. Adverse health effects of air pollution

Air pollutants of particular interest in regard to health are carbon monoxide (CO), nitrogen dioxide (NO_2), ozone (O_3), particulate matter (PM_{10} with an aerodynamic diameter of <10μm and $PM_{2.5}$ with an aerodynamic diameter of <2.5μm), and sulphur dioxide (SO_2) (Brook *et al.*, 2004; Brook, Brook and Rajagopalan, 2003; Mills *et al.*, 2009; Newby *et al.*, 2015). Air pollutants can be classified into primary and secondary air pollutants. Primary air pollutants are those that are directly emitted to the atmosphere (e.g. arise from traffic fossil fuel combustion, heating and industrial processes (Brook *et al.*, 2010; Brook, Brook and Rajagopalan, 2003). CO, NO (nitrogen oxide), NO_2, SO_2 are such primary pollutants (Brook *et al.*, 2010). O_3, instead, is a so called "secondary pollutant" (Brook *et al.*, 2010, p. 2335) because it is produced by chemical reactions with directly exhausted air pollutants within the atmosphere. PM is a special case as it can be emitted directly into the atmosphere (primary) but also may be formed within the atmosphere (secondary pollutant) (ibid.). Particulate pollution, i.e. PM, usually is made up of a complex mixture of liquid and gaseous compounds, with greatly varying sizes, compositions and constitution (Brook et al., 2010; Brook, Brook and Rajagopalan, 2003). $PM_{2.5}$ refers to all particles with a diameter of less than 2.5 μm, while PM_{10} refers to aerodynamic diameters of <10 μm. Particles with a diameter of <10μm can be deeply inhaled into the lungs and the terminal alveoli (Brook, Brook and Rajagopalan, 2003; Mills *et al.*, 2009), while particles > 10μm are likely to remain in nose and throat (Brook, Brook and Rajagopalan, 2003). Of these particles < 10 μm, particles <2.5 μm may reach terminal bronchioles and alveoli and particles <0.1 μm may even penetrate into the bloodstream (Yang, Li and Tang, 2020). The toxicity of PM particles depends on their chemical composition and surface (Mills *et al.*, 2009). A more detailed description of the origin and constitution of air pollutants goes beyond the scope of this dissertation but may be found in the paper by Brook *et al.* (2010).

Over the last 40 to 50 years there has been a growing body of evidence on potential adverse health effects of air pollution- especially in regard to respiratory and cardiovascular diseases, and mortality. In these clinical and epidemiological studies, the focus particularly lies on ambient air pollutants like carbon monoxide (CO), nitrogen monoxide (NO), nitrogen dioxide (NO_2), ozone (O_3), particulate matter (PM_{10} with an aerodynamic diameter of <10μm and $PM_{2.5}$ with an aerodynamic diameter of <2.5μm) and sulphur dioxide (SO_2). A groundbreaking study by Dockery *et al.* (1993) found a relationship between exposure to air- pollution and hospital admissions as well as cardiovascular mortality. It was the so-called six cities study which analysed data over a time period of 14-16 years (ibid.). Ever since, there has been a wave of numerous clinical and epidemiological studies showing strong associations between the adverse impacts of air pollution on health. In general, air pollution is associated with increased morbidity and mortality (see e.g., Atkinson *et al.*, 2014; Brook *et al.*, 2010; Cesaroni *et al.*, 2014; Chen, Goldberg and Villeneuve, 2008; Chen and Hoek, 2020; Lelieveld *et al.*, 2019; Mann *et al.*, 2002; Pope *et al.*, 2002; Pope *et al.*, 2004; Samet *et al.*, 2000; Stieb, Judek and Burnett, 2002). Moreover, air pollutants are increasingly associated with cardiovascular hospitalisation (Collart *et al.*, 2018; Franco *et al.*, 2020; Klompmaker *et al.*, 2021; Mann *et al.*, 2002; Rodríguez, Cobo-Cuenca and Quiles, 2022; Yorifuji, Suzuki and Kashima, 2014), cardiovascular mortality (Lu, Kang and Wang, 2022; Mann *et al.*, 2002; Pope and Turner *et al.*, 2015; Raaschou-Nielsen *et al.*, 2012; Samet *et al.*, 2000; Wong *et al.*, 2002) and all-cause mortality (Carey *et al.*, 2013; Katsouyanni *et al.*, 1997; Pope *et al.*, 2002; Samet *et al.*, 2000).

1.2. The pathophysiological association between air pollution and adverse health outcomes

Definite mechanisms underlying the association between air pollution, cardiovascular diseases and mortality have yet to be established (Mills *et al.*, 2009). Nevertheless, a vast amount of epidemiological, clinical and biomedical studies has dealt with this issue. Within the existing literature, a special focus was put on PM and its impact on human health.

A central effect of air pollution on health is due to its causation of oxidative stress (Münzel *et al.*, 2017; Münzel *et al.*, 2018). Various reviews are comparable in the identified consequences of this oxidative stress: Münzel *et al.* (2018) found that oxidative stress may lead to endothelial disfunction, hypertension and subsequently promote atherosclerosis (Münzel *et al.*, 2018). Adverse effects of this are myocardial infarction, stroke and congestive heart failure (Münzel *et al.*, 2017). In their review Miller and Newby (2020) identified manifold comparable effects of air pollution on the cardiovascular system such as vascular dysfunction, higher vulnerability to ischemic damage and thrombosis. Brook *et al.* (2010) identified comparable consequences including systemic inflammation, systemic oxidative stress, thrombosis, coagulation, systemic and pulmonary arterial blood pressure increases, imbalances in vascular function, atherosclerosis, reduction of heart rate variability and promotion of cardiac ischemia in vulnerable individuals. Various hypotheses of mechanisms leading to these consequences exist. A large amount of literature focuses on the pathophysiological pathway between particulate matter and health which is why this pathway is further explained within this dissertation.

The respiratory uptake of particulate matter is described to lead to a pulmonary inflammation with a subsequent secretion of inflammation mediators into the blood circulation which thereupon enhances the risk of atherosclerosis and cardiovascular diseases (Brook *et al.*, 2010; Brook, Brook and Rajagopalan, 2003). In their review of the literature Brook, Brook and Rajagopalan (2003) identified various pathophysiological pathways of cardiovascular diseases due to $PM_{2.5}$ inhalation. Systematic inflammation which subsequently is associated with arrhythmias, vascular dysfunction, procoagulant changes in the composition of blood properties, advancement of atherosclerosis appears to be the central protagonist (ibid.). PM induced imbalances in the autonomic vascular tone and arrhythmias provoke adverse cardiac events (ibid.). The pathophysiological pathways are reinforced and further elaborated by a review by Brook, Newby and Rajagopalan (2017) suggesting three mediating pathways by which $PM_{2.5}$ promotes cardiovascular events (see Figure 1-1). Within the first pathway, air pollution particles directly penetrate the lung blood barrier and reach systemic circulation (ibid.). The second pathway deals with translocation of the local inflammation and oxidative stress to the systemic circulation and which thus impacts on the cardiac tissue (ibid.). Lastly, inhaled $PM_{2.5}$ interacts with pulmonary tissue and local cells (e.g. macrophages), this induces a local inflammation and oxidative stress (ibid.). According to the authors, this causes an imbalance of the autonomic nervous system favouring the sympathetic over the parasympathetic nervous system (e.g. augmented blood pressure) (ibid.). Brook, Newby and Rajagopalan (2017) stated that these pathways individually or in cooperation can induce vascular dysfunction, modified rheology, arrythmia, and promotion of atherogenesis. The systematic inflammation leads to facilitation of plaque generation, disruption of vulnerable plaques, direct effect of pollutants (Brook and Rajagopalan, 2010). Further, in a secondary data analysis, Chuang *et al.* (2011) identified an association between long-term exposure to air pollution and a rise in IL-6, total cholesterol, fasting glucose and HbA1c in the elderly population. The authors concluded a relationship of these increases with the genesis of atherosclerosis.

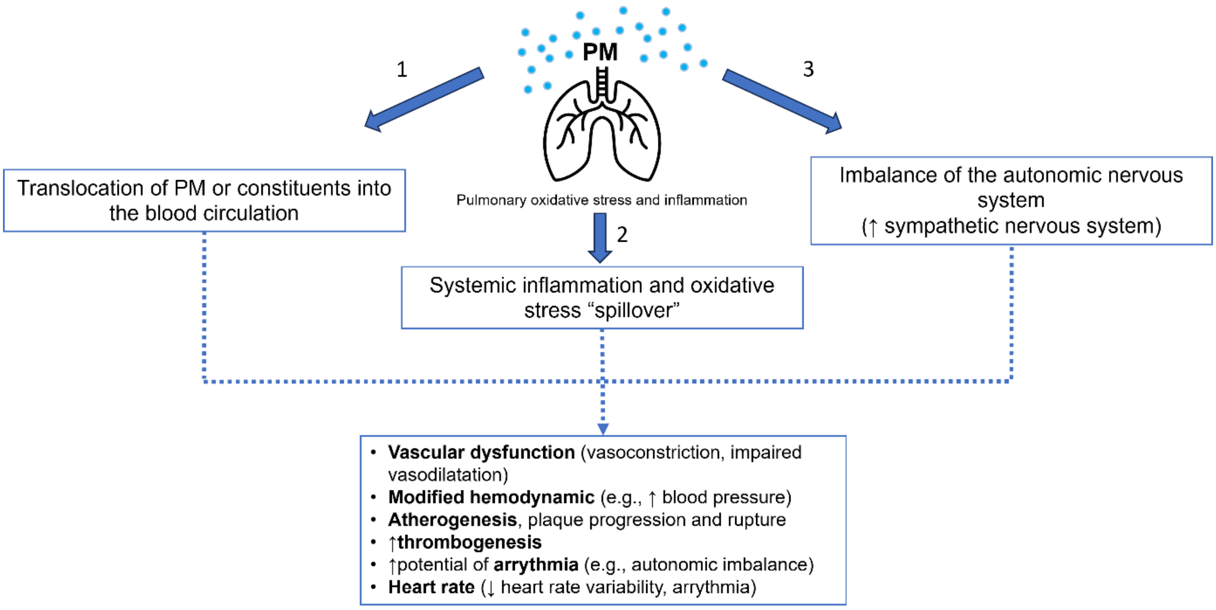


Figure 1-1 Biological pathways linking PM exposure with cardiovascular diseases

Biological pathways linking PM exposure with cardiovascular diseases (modified according to Brook et al.; Mills et al., 2009; Newby et al., 2015 (2010)) and Brook, Newby and Rajagopalan (2017). Three mediating pathways:
1. Translocation of PM or constituents into the systemic circulation
Each of these pathways, individually or in conjunction can cause the listed consequences
2. Mediators of inflammation invade the systemic circulation
3. Imbalance of the autonomic nervous system, favouring the sympathetic nervous system over the parasympathetic nervous system

1.3. The adverse impact of the socioeconomic status on health

The socioeconomic status (SES) is specified as the combination of an individual's economic and social factors including education, income and occupation (American Psychological Association; Baker, 2014). It is assumed to largely interfere with individuals exposure to air pollution (Evans and Kantrowitz, 2002; Hajat, Hsia and O'Neill, 2015; Miller and Newby, 2020). Evans and Kantrowitz (2002) observed an inverse association between the SES and health outcome. Hence, it was assumed that a possible relationship between air pollution and mortality could be confounded by the SES. Therefore, the impact of the socioeconomic status is analysed within this dissertation. For this aim, purchasing power on a postal code level was the closest indicator available to the SES and hence used as its proxy.

2. Research question, objectives and overview

Given the large body of evidence for the association between the patient extrinsic risk factor air pollution and cardiovascular mortality, the aim of this dissertation is to analyse the impact of air pollution on mortality in comparison to and conjunction with the patient immanent risk factors.

In particular, the aim is to explore the potential benefit in risk prediction by use of the extrinsic risk factor air pollution in comparison to the well-established patient immanent ESC SCORE 2 risk factors for a population with pre-existing cardiovascular risk factors in Hesse. The results of this dissertation may inform clinical practice to improve the scoring tools for cardiovascular mortality.

To fulfil this aim, this dissertation aims to answer the following questions:

1. Is there a time dependent association between the extrinsic risk factor air pollution and mortality in a cohort with cardiovascular risk?
2. Does the additional use of the extrinsic risk factor air pollution enhance the result in comparison to only using the patient immanent risk factors of the ESC Score 2?
3. Is there further improvement when the extrinsic risk factor socioeconomic status is included in the analysis?

2.1. Overview

After having introduced to the topic of this dissertation, this subchapter briefly guides through the chapters.

The methods chapter describes how patient-level and air pollution data were collected, conjoined, organised and cleaned. Adjacently, dealing with missing data within this dissertation is circumstantiated. Thereafter, the theoretical foundation for the statistical analysis methodology applied is outlined.

Afterwards, in the results chapter, the findings of the statistical analysis are outlined. In the discussion, the study population, investigation period, geographic location and results in the context of the existing literature are assessed. This dissertation is closed with a concluding chapter.

3. Methods

In this chapter, the dataset that consists of patient-level hospital data and air pollution data from air quality monitoring stations across Hesse is introduced. Afterwards the data cleaning process and the statistical methods applied within this dissertation are described.

3.1. Computing software used within this dissertation

For statistical analysis, the statistical computing software R 4.2.0 (R Core Team, 2022) was used. For data organisation and cleaning, the "tidyverse" package by Wickham *et al.* (2019) was utilised. For the handling of dates and times, the package "lubridate" by Grolemund and Wickham (2011) was applied. To convert environmental data to temporal time series data, the "tsibble" package (Wang, Cook and Hyndman, 2020) was employed.

3.2. Data collection

3.2.1. Study population – Collection of patient-level data

The present study is a post-hoc analysis based on data from the ongoing prospective registry study BioProspective (Ref PMID 31423815). Between 08/2010 and 12/2019, 4836 patients with clinically an indicated invasive cardiac diagnostic procedure (coronary angiography) due to a suspected chronic coronary syndrome were enrolled into this study cohort. Patients with or without previously known coronary artery disease were eligible to participate. Participation was voluntary, each patient gave written informed consent prior to study enrolment. Included patients were informed about content, procedures, and aims of the database. Patients were also informed about their possible withdrawal of the data base at any time without negative consequences. The BioProspective protocol was approved by the local ethics committee as part of the cardiovascular biobank (University Giessen, AZ 199/15). The study was conducted in accordance with the declaration of Helsinki.

Clinical data on enrolled patients were collected and stored in a password protected database (REDCap) in a pseudonymised manner according to the standard operating procedures of the cardiovascular biobank. REDCap (Research Electronic Data Capture) is a web-based database software to facilitate data collection for research studies (Harris *et al.*, 2009; Harris *et al.*, 2019).

Intrinsic cardiovascular risk factors were prevalent among the cohort. They were recorded according to the established classification system SCORE 2 of the European Society of Cardiology (SCORE2 working group and ESC Cardiovascular risk collaboration, 2021). SCORE 2 consists of 5 variables – namely sex (male/female), age (continuous in years), smoking (smoking or non-smoking), systolic blood pressure (continuous in mmHg), non-HDL cholesterol (continuous in mg/dL) to predict the 10-year risk of fatal and non-fatal cardiovascular events in individuals across Europe (ibid). These data were gathered upon study enrolment by health care professionals. Non-HDL-cholesterol was calculated as “total cholesterol – HDL-cholesterol”. Ex-smokers were categorised as non-smokers.

The primary outcome for purpose of this dissertation was overall mortality. Mortality information was received by a request to the German national mail office on the 26th September 2022. Data in regard to the patient registry partly was already published within other studies (Diouf *et al.*, 2019; Elsner *et al.*, 2020).

3.2.2. Collection of environmental Data – Markers of air pollution

For this work, Mrs. Bärbel Oehme and Dr. Florian Ditas have provided publicly available air pollution data of the HLNUG Hessisches Landesamt für Naturschutz, Umwelt und Geologie (Hessian State Agency for Nature Conservation, Environment and Geology) (HLNUG, 2024). In total, the air pollution data were collected by 40 continuous HLNUG air quality stations across the state Hesse (Table 3-1). Those are measuring stations with automated analysers. They employ physical measurement techniques to measure air pollutants utilised as air pollution markers within this dissertation (Table 3-2). Among these air monitoring stations, measured values are inquired approximately every 5s in an automated process (HLNUG, 2022b). Half-hour averages are then transmitted to measuring network centres. Additionally, meteorological variables are measured (global radiation, atmospheric pressure, temperature, precipitation, relative humidity, wind direction, wind velocity) (ibid.). Daily average concentration values were inquired. For reasons of comparability, only data from continuously measuring air pollution stations were used in this dissertation.

Table 3-1 – Overview of the air pollution monitoring stations of the HLNUG in Hesse

**Longitude and latitude were rounded to 4 decimal places*
colours according to the location of the measurement station: yellow = urban background, green = rural background, white = urban traffic area

	Station ID	Station name	Station environ-ment	Longitude*	Latitude*	Time restrictions
1	**DEHE001**	Darmstadt	urban background	8.6646	49.8723	
2	**DEHE005**	Frankfurt Höchst	urban background	8.5425	50.1017	
3	**DEHE008**	Frankfurt Ost	urban background	8.7463	50.1253	
4	**DEHE011**	Hanau	urban background	8.9215	50.1357	
5	**DEHE013**	Kassel Mitte	urban background	9.4834	51.3143	
6	**DEHE018**	Raunheim	urban background	8.4515	50.0103	
7	**DEHE020**	Wetzlar	urban background	8.5006	50.5672	
8	**DEHE022**	Wiesbaden Süd	urban background	8.2449	50.0503	
9	**DEHE030**	Marburg	urban background	8.7693	50.8043	
10	**DEHE032**	Bebra	urban background	9.8003	50.9700	
11	**DEHE044**	Limburg	urban background	8.0610	50.3832	
12	**DEHE045**	Michelstadt	urban background	9.0020	49.6725	
13	**DEHE134**	Fulda Zentral	urban background	9.6801	50.5464	
14	**DEHE135**	Frankfurt Schwanheim	urban background	8.5763	50.0755	
15	**DEHE150**	Frankfurt Niedwald	urban background	8.5945	50.1134	24.04.2019-10.01.2021
16	**DEHE160**	Mörfelden	urban background	8.5659	49.9650	
17	**DEHE161**	Flörsheim	urban background	8.4287	50.0190	

18	**DEHE162**	Frankfurt Lerchesberg	urban background	8.6830	50.0815	
19	**DEHE024**	Witzenhausen Wald	rural background	9.7746	51.2918	
20	**DEHE026**	Spessart	rural background	9.3994	50.1644	
21	**DEHE028**	Fürth Odenwald	rural background	8.8172	49.6535	
22	**DEHE039**	Burg Herzberg	rural background	9.4594	50.7704	
23	**DEHE042**	Linden	rural background	8.6844	50.5330	
24	**DEHE043**	Riedstadt	rural background,	8.5168	49.8251	
25	**DEHE046**	Bad Arolsen	rural background	8.9282	51.4309	
26	**DEHE050**	Zierenberg	rural background	9.2712	51.3608	
27	**DEHE051**	Wasserkuppe	rural background	9.9359	50.4977	
28	**DEHE052**	Kleiner Feldberg	rural background	8.4461	50.2219	
29	**DEHE060**	Kellerwald	rural background	9.0318	51.1548	
30	**DEHE037**	Wiesbaden Ringkirche	urban traffic area	8.2303	50.0772	
31	**DEHE040**	Darmstadt Hügelstraße	urban traffic area	8.6538	49.8695	
32	**DEHE041**	Frankfurt Friedberger Landstraße	urban traffic area	8.6919	50.1246	
33	**DEHE049**	Kassel Fünffensterstraße	urban traffic area	9.4911	51.3121	
34	**DEHE059**	Fulda Petersberger Straße	urban traffic area	9.6848	50.5500	
35	**DEHE061**	Gießen Westanlage	urban traffic area	8.6686	50.5841	
36	**DEHE062**	Marburg Universitätsstraße	urban traffic area	8.7705	50.8072	
37	**DEHE063**	Heppenheim Lehrstraße	urban traffic area	8.6420	49.6432	
38	**DEHE112**	Wiesbaden Schiersteiner Straße	urban traffic area	8.2289	50.0721	
39	**DEHE116**	Offenbach Untere Grenzstraße	urban traffic area	8.7848	50.1015	
40	**DEHE131**	Limburg Schiede	urban traffic area	8.0599	50.3864	

Table 3-2 - Overview of the measured air quality parameters used within the dissertation and their measurement techniques

Air pollutant	**Measurement techniques** (HLNUG, 2022a)
Inorganic pollutants	
Carbon dioxide (**CO_2**)	Gas filter correlation (Gfc)
Carbon monoxide (**CO**)	IR absorption
Ozone (**O_3**)	UV-absorption
Sulphur dioxide (**SO_2**)	UV- fluorescence
Nitrogen monoxide (**NO**)	Chemiluminescence
Nitrogen dioxide (**NO_2**)	Chemiluminescence passive collection: Adsorption of NO_2 to triethanolamine, followed by quantitative-chemical laboratory analysis
Particles	
Particulate Matter 2,5 (**$PM_{2,5}$**) (≤ 2,5 µm)	Hybrid process (Nephelometer and ß-absorption)
Particulate Matter 10 (**PM_{10}**) (≤ 10 µm)	Hybrid process (Nephelometer and ß-absorption)

3.2.3. Geocoding- assigning air pollution data to patients

Mapping patients to the relevant air monitoring stations on postal code level was achieved through retrieving spatial data from the function “GetData” of the “raster” package by Hijmans (2023), Bundesamt für Kartographie und Geodäsie (2019) and OpenStreetMap contributors (2019). The package “sf” by Pebesma (2018) was utilised to prepare the data for mapping. Eventually, data were mapped with the “ggplot2” package by Wickham (2016). In order to work with the spatial data, it was ensured that all different sources of data use the same coordinate reference system (CRS). A CRS can be specified by an EPSG projection. The HLNUG uses the coordinates of the World Geodetic System 1984 (WGS 84). Thus, the EPSG (European Petroleum Survey Group Geodesy) projection 4326 was used which can be seen as its EPSG identifier within the analysis of this dissertation.

Via geocoding air pollution data of the HLNUG measurement stations were assigned to the patients. For this geocoding process, the “sf” Package in R developed by Pebesma (2018) was used. Stations were assigned to individuals’ residing areas (postal code level) according to distance and availability of air quality data. For each air pollutant, the closest station within a radius of 35 km that covered at least 90% of air pollution data for the last 1094 days (3 years) prior to study enrolment and the day of study enrolment was assigned, respectively.

3.2.4. Calculation of the ESC SCORE 2

The ESC SCORE 2 was calculated according to the R package "estimateCVRisk" by Grün (2023) that calculates the SCORE 2 according to the SCORE2 working group and ESC Cardiovascular risk collaboration (2021). The variables used for the calculation process are non-HDL-cholesterol, sex, age, systolic blood pressure, and smoking.

3.2.5. Collection of socioeconomic status data

Sociodemographic data such as the purchasing power were acquired from the commercial provider Digital Data Services (DDS) GmbH. To ensure data privacy this data were acquired only on the level of postal code regions. This concept was approved as post-hoc analysis of the data from the BioProspective cohort of the local ethics committee (Giessen University, AZ 147/11). Purchasing power was also mapped on a postal code level.

3.3. Dealing with missing values

Within the dataset there were missing data among patient and air pollution data. Thus, this section deals with the handling of missing data within this dissertation. Not dealing with missing data can cause sacrifice of statistical power (Graham, 2009). Different types of missing data exist. According to Rubin (1976) and Little and Rubin (2020) processes causing missing data may be classified as missing completely at random (MCAR), missing at random (MAR), or missing not at random (MNAR).

3.3.1. Dealing with missing patient- level data

3.3.1.1. ESC SCORE 2 variables

The variables within the patient data set needed to calculate the ESC SCORE 2 (non-HDL cholesterol (total cholesterol – HDL-cholesterol), sex, age, systolic blood pressure, smoking) contained missing values. It was assumed that the missing patterns were at least MAR or even MCAR. For a better understanding of the missingness patterns, analyses of missingness patterns with the R packages “naniar” (Tierney and Cook, 2023), “mice“ (van Buuren and Groothuis-Oudshoorn, 2011), and “finalfit”

(Harrison, Drake and Ots, 2023) were conducted. Pair plots were created to observe connections between missing and observed values with the function "missing_pairs".

To impute the missing values of the ESC SCORE 2 variables, the R package mice (van Buuren and Groothuis-Oudshoorn, 2011) was used. 1 was set for the number of imputations. Predictive mean matching was set as the method for the imputation of the continuous variables (cholesterol, age, systolic blood pressure) and logistic regression imputation for binary variables (sex, smoking).

3.3.1.2. Mortality

The 224 Patients without mortality data were excluded from the analysis (complete case analysis). Complete case analysis may lead to possible bias (Jamshidian and Mata, 2007). Nevertheless, this decision was made for economic reasons and the high probability for the presence of MCAR missing structure given the marginal share of missing values (4.6%) for this variable. The same procedure was conducted for the same reasons for the 2 patients without residing area information.

3.3.2. Dealing with missing air- pollution data

The air quality data set also had missing values. The missingness of air pollution data is generally not caused by the missing value itself (Junninen *et al.*, 2004) but may be explainable by maintenance reasons or measurement issues (Gómez-Carracedo *et al.*, 2014; Plaia and BONDI, 2006) that also are likely to have occurred with the air pollution data used within this study. Like Hildebrandt *et al.* (2009), MCAR conditions were assumed for the air pollution data as their missingness is unlikely to depend on health/ patient variables. Before assigning the air pollution data to the patients, air pollution variables were imputed with the R package "mtsdi" for multivariate time series imputation of air pollution (Junger and Leon, 2018). This package uses an expectation maximisation (EM) algorithm based imputation method for multivariate normal time series (ibid.). An EM algorithm appears a good fit to estimate missing data close to the actual value (Dong and Peng, 2013; Musil *et al.*, 2002). In EM, an iterative algorithm generates maximum likelihood estimates (Graham, 2009; Musil *et al.*, 2002). In their package, Junger and Ponce de Leon (2015) describe a four-step iterative algorithm including the estimation and re-estimation of missing values that is conducted up to the achievement of a convergence criterion. The package has frequently been applied to studies on air quality and health (e.g. Motesaddi Zarandi *et al.*, 2022; Nguyen, Malig and Basu, 2021; Thongphunchung *et al.*, 2022).

Within this dissertation, imputation was separately conducted for each air quality monitoring station. Given that the authors of the "mtsdi" package declare that an estimates' validity decreases when missingness exceeds 10% (Junger and Ponce de Leon, 2015), imputation was conducted when at least 90% of an air pollution variable at an air pollution monitoring station were available (i.e. e.g. SO_2, NO,…). Otherwise, values were not imputed. Some of the imputations yielded negative and therefore implausible values. Given the few occurrences of this happening, it was decided to keep imputed values as they were. Moreover, a restriction of the range of imputed values to prevent the imputation of implausible values leads to potential bias (Hippel, 2013; Horton, Lipsitz and Parzen, 2003; Rodwell *et al.*, 2014).

3.4. Statistical analysis

The previously described stages of data manipulation created a dataset that contained mortality information for each patient, as well as the ESC SCORE 2 and air pollution information for 1094 days (3 years) prior to study enrolment and the day of study enrolment.

The following section describes the process of statistical analyses. This includes Receiver-operator characteristic analysis and logistic regression analysis.

3.4.1. Analysis of time-dependent association between air pollution and mortality

This subsection describes the method applied to analyse the time-dependent association between mortality and air pollution markers – Receiver-operating characteristic (ROC).

It was decided to analyse the time-dependent association of mortality with air pollution markers for the respective last three years (i.e. 1094 days) before enrolment to the study and the day of study enrolment for each individual. To this end the means of each air pollution marker from any of the 1094 days prior to and the day of study enrolment were calculated to represent the average air pollution over a specific time period. In this mean calculation missing values were excluded. The following example will illustrate this:

The average air pollution for one patient for the time period starting 2 days prior to study enrolment was calculated in the following manner:

mean ([air pollutant 2 days prior to enrolment],[air pollutant 1 days prior to enrolment], [air pollutant day of enrolment]**)**

If a day had a missing value, no mean was calculated but the missing value was preserved. This was done to prevent that longer time periods with missing values of air pollutants would have the same mean value. Each of these mean values for a certain time period within the 1095 days prior to study enrolment were used to calculate a ROC curve. From here on, the statistical basics of ROC analysis are described.

Receiver-operating characteristic (ROC) analysis is a well-known tool to analyse a diagnostic test's precision (see e.g., Hajian-Tilaki, 2013; Metz, 1978; Zou, O'Malley and Mauri, 2007; Zweig and Campbell, 1993). Diagnostic tests often produce continuous results (Hoo, Candlish and Teare, 2017). The aim of ROC analysis in this dissertation is to identify a threshold of the mean air pollution exposure over a certain time period to distinguish "dead" or "not dead" – or in other words "the selection of a threshold to distinguish a 'positive' test from a 'negative' test" (Hoo, Candlish and Teare, 2017, p. 357). To produce a ROC curve, contingency tables produced for different thresholds along the continuous line of test results are of great help. A contingency table (e.g., see Figure 3-1 – Contingency table of a diagnostic test) visually summarises the four possible outcomes of a diagnostic test (true positive, false positive, false negative, and true negative) at a certain threshold/ cut-off.

From this contingency table, frequently used performance metrics like the true positive rate (TPR or sensitivity) and the false positive rate (FPR or 1- Specificity) can be calculated. For each threshold (different classifiers) of air pollution the contingency table the True Positive Rate (TPR) and False Positive Rate (FPR) can be calculated.

		actual results	
		dead	not dead
predicted	dead	**True positive (TP)**	**False positive (FP)**
	not dead	**False negative (FN)**	**True negative (TN)**
		True positive rate (TPR) = Sensitivity = (TP/(TP+FN)	**False positive rate (FPR) = 1-Specificity = FP/(FP/TN)**

Figure 3-1 – Contingency table of a diagnostic test

The results can be graphically displayed by plotting the FPR (x- axis) against the TPR (y-axis). This graph (i.e. the ROC Curve) is a visual means to summarise contingency tables for different classifiers. Basic parts of the graph are described to locate the reader within the ROC space. Several points on the ROC space are of importance to better understand the performance of diagnostic classifiers (see Figure 3-2). The blue diagonal line (x = y) displays the ROC curve of a random classifier while the red curve represents the gold standard. Generally, classifiers to the left of the blue diagonal line are good classifiers and classifiers to the right of the blue diagonal line are bad classifier (because they are worse than a random classifier). The accuracy of a classifier increases the closer it gets to (0,1) (high TPR, low FPR) (see e.g., Altman and Bland, 1994). Thus, the diagnostic performance of the green ROC curve is better than that of the orange curve. The area under the curve (AUC) can be defined as a one-dimensional summary of a ROC curve and thus a one-dimensional summary of the tests' discrimination ability or its diagnostic performance (see e.g. Fawcett, 2006; Hajian-Tilaki, 2013; Hoo, Candlish and Teare, 2017; Zou, O'Malley and Mauri, 2007). The AUC of the blue diagonal line (random classifier) is 0.5, while the AUC of the red curve (gold standard/ perfect classifier) is 1.

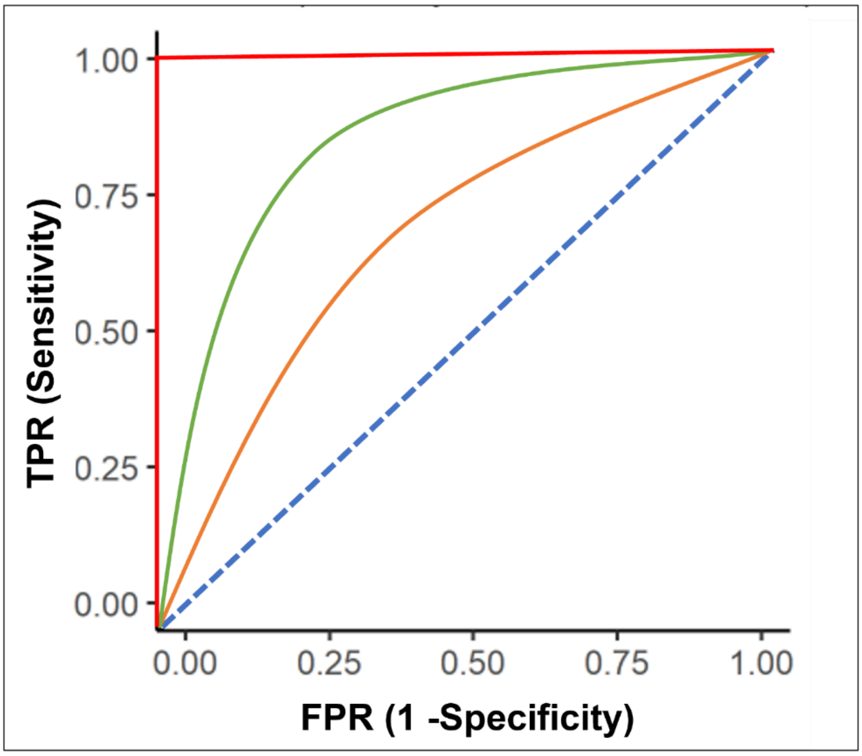


Figure 3-2- performance of different ROC curves

TPR (True positive rate), FPR (False positive rate)

After having introduced the basic concepts of ROC curve analysis, the following section explains how ROC curve analysis was applied within this dissertation.

The mean exposure values for 1095 time periods (3 years) was used to calculate ROC curves of mortality prediction. To compare the precision of these 1095 ROC curves, AUC values of each curve as a one-dimensional summary for a certain time period were used. Afterwards, AUC values (y- axis) against the time prior to study enrolment were plotted. This approach is subsequently explained in further detail.

Figure 3-3 visualises this proceeding: the centre of the x- axis (marked by the icon of a blue hospital) represents the day of study enrolment. The aim was to predict mortality which was queried in September 2022 and therefore lay in the future from the perspective of study enrolment. To this end, the mean exposure to an air pollutant from a certain day until study enrolment was analysed. Hence, the red backward arrow represents the time investigated for the analysis of air pollution exposure. From the perspective of study enrolment this time lay in the past. The ROC graphs resemble the 1095 ROC graphs created, while the red curve above illustrates the 1095 AUC values as a means of summaries of those 1095 ROC graphs. By doing this, the maximum AUC value and therefore the maximum time-dependent association between mean exposure to an air pollutant and mortality could be identified. A ROC graph for the association between ESC SCORE 2 on the day of study enrolment and mortality was also created. To perform the analysis in R, the “pROC” package by Robin *et al.* (2011) was used.

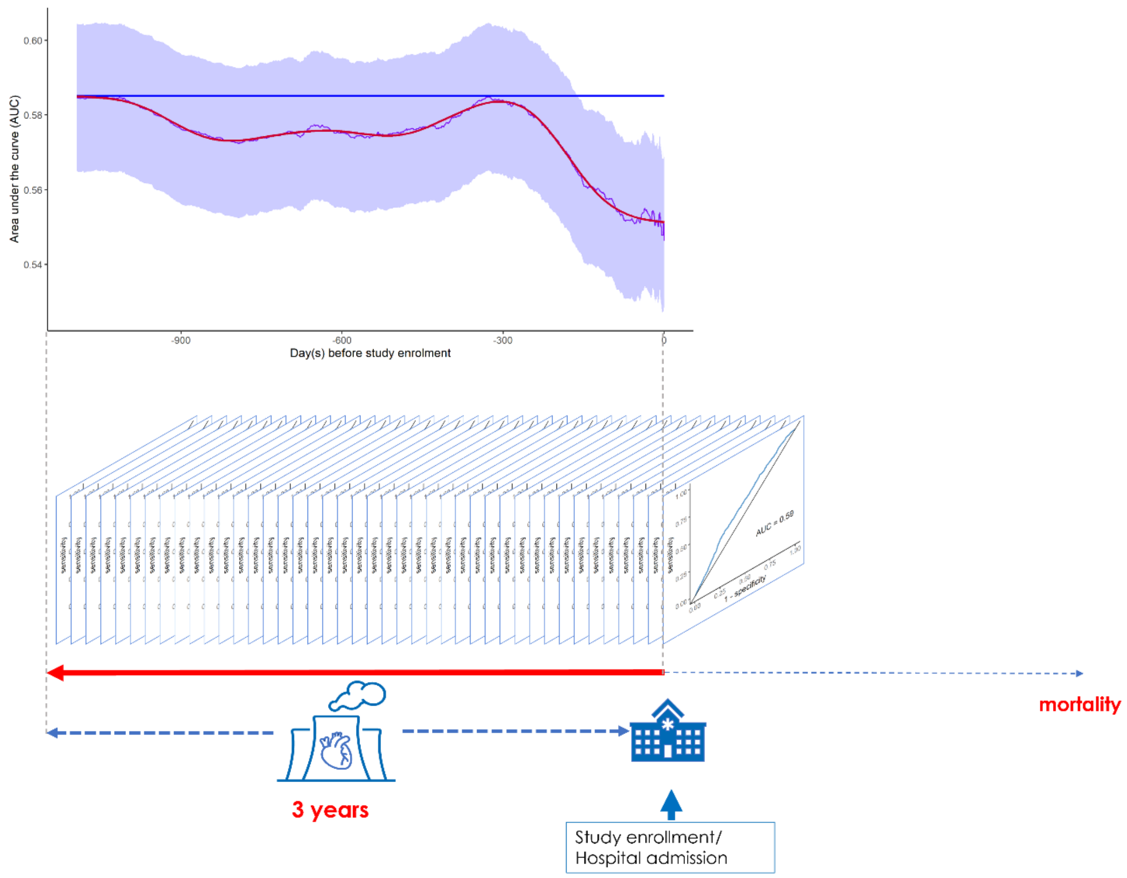


Figure 3-3- ROC curves and curve of AUC values over time

The graph is organised into three parts. From the bottom to the top the first part displays the chronological orientation of the study. The centre of the x- axis (blue hospital) marks the day of study enrolment/hospital admission. The forward arrow points towards the future. In September 2022 mortality was queried. The backward arrow (red) points three years towards the past. This red arrow represents the time investigated for the analysis of air pollution exposure. From the perspective of study enrolment this time lay in the past. The ROC graphs above the x-axis resemble 1095 ROC graphs of the association between the respective exposure to mean air pollution concentrations for different time periods until study enrolment and mortality. The upper graph illustrates the 1095 AUC values as a means of summaries of those 1095 ROC graphs.

3.4.2. The role of air pollution exposure compared to patient immanent risk factors in predicting mortality

To analyse the benefit of considering the mean air pollution exposure in mortality prediction in comparison with only using patient-immanent risk factors like the variables of the ESC SCORE 2, logistic regression was used. For this aim, logistic regressions models including the mean air pollution exposure for their maximum time-dependent association were created. Comparable to the previous subsection, logistic regression is first described theoretically and according to its use within this dissertation.

In logistic regression modelling, there is a sample of n independent observations in the form (x_i, y_i), i = 1, 2, 3, …, n. y_i is a binary response e.g. 0 or 1 or death or alive (see e.g., Hosmer, Lemeshow and Sturdivant, 2013; Nwanganga and Chapple, 2020) and x is the value of the variable for the i^{th} patient. The logistic regression function is used to predict the probability p of the binary response Y (e.g. 0 or 1 or death or alive).

Multivariate versions of logistic regressions exist where a binary outcome variable is modelled as a function of n explanatory variables. To fit this regression, a sigmoid logistic curve with the following function is used (see e.g., Hosmer, Lemeshow and Sturdivant, 2013; Nwanganga and Chapple, 2020; Rossi, 2022).

Equation 3-1 Logistic regression model

$$p = P(Y|\, X_1, \mathrm{X}_2, \dots, X_n) = \frac{e^{b_0 + b_1 X_1 + b_2 X_2 + b_3 X_3 + \cdots + b_n X_n}}{1 + e^{b_0 + b_1 X_1 + b_2 X_2 + b_3 X_3 + \cdots + b_n X_n}}$$

Y	*Binary response variable*
	Y = 1 Occurrence of an event (e.g. dead)
	Y = 0 Non-occurrence of an event
p	*Probability/ Proportion that p = P(Y=1)*
$X_1, X_2, \ldots, X_n$	*Set of explanatory variables*
$b_1, b_2, \ldots, b_n$	*Coefficients/slopes of the logistic regression,* b_0 *is the intercept*

To fit a logistic regression model (Equation 3-1) to a dataset, the values of $b = b_0, b_1, b_2, \cdots, b_n$ need to be estimated which is done by applying the maximum likelihood method (see e.g., Hilbe, 2015; Hosmer, Lemeshow and Sturdivant, 2013; James *et al.*, 2021; Rossi, 2022). For reasons of mathematical applicability the logarithm of the likelihood function is used to fulfil this task (see e.g., Hosmer, Lemeshow and Sturdivant, 2013)

The residual deviance compares the log-likelihood of the perfect (saturated) model where the fitted values are equal to the observed values with the log-likelihood of the fitted model (see e.g., García-Portugués, 2023; Hilbe, 2015). The null deviance compares the log-likelihood of the saturated model with the log- likelihood of the null model which only uses the intercept (see e.g., Nwanganga and Chapple, 2020). The greater the difference between the null and residual deviance, the better a models' performance (ibid.). The log-likelihood of a saturated model in logistic regression equals 0.

3.4.2.1. *Interpretation of the results of logistic regression analysis with the Odds Ratio*

For the interpretation of the coefficients of logistic regression, the odds ratio (OR) is used. The OR is utilised to identify the influence of an explanatory variable on the outcome. The odds ratio is defined as the ratio of two odds. An odds is defined as the ratio of the probability of an occurrence of an event (p) divided by the probability of its non-occurrence (1-p) (see e.g., Nick and Campbell, 2007; Nwanganga and Chapple, 2020).

Equation 3-2- Odds

$$
\begin{aligned}
odds(Y = 1 | X_1, \mathrm{X}_2, \ldots, \mathrm{Xn}) &= \frac{P(occurring)}{P(not\ occurring)} = \frac{p}{1-p} \\
&= \frac{\dfrac{e^{b_0+b_1X_1+b_2X_2+b_3X_3+\cdots+b_nX_n}}{1+e^{b_0+b_1X_1+b_2X_2+b_3X_3+\cdots+b_nX_n}}}{1-\dfrac{e^{b_0+b_1X_1+b_2X_2+b_3X_3+\cdots+b_nX_n}}{1+e^{b_0+b_1X_1+b_2X_2+b_3X_3+\cdots+b_nX_n}}} = \frac{\dfrac{e^{b_0+b_1X_1+b_2X_2+b_3X_3+\cdots+b_nX_n}}{1+e^{b_0+b_1X_1+b_2X_2+b_3X_3+\cdots+b_nX_n}}}{\dfrac{1}{1+e^{b_0+b_1X_1+b_2X_2+b_3X_3+\cdots+b_nX_n}}} \\
&= e^{b_0+b_1X_1+b_2X_2+b_3X_3+\cdots+b_nX_n}
\end{aligned}
$$

3.4.2.2. *Logistic regression within this dissertation*

In the following section, describes the use of logistic regression analysis to estimate the possible benefit of adding air pollution to a model predicting mortality. Earlier, it was described how the time period of mean air pollution exposure that yielded a maximum AUC was identified. For the fitting the multiple logistic regression model, the mean air pollution of this time period which yielded the maximum AUC was added to the core data frame of ESC SCORE 2 variables.

Even after the earlier described EM imputation of air pollution data, there was considerable missingness among the variables of mean air pollution exposure for the time period with the maximum AUC which was used for the logistic regression analysis. As introduced earlier, like Hildebrandt *et al.* (2009), it was assumed that the missingness of air pollution data was completely at random because it is unlikely that the missingness depends on the patient variables. Thus, for the process of logistic regression, a complete case analysis (listwise deletion) was performed. While complete case analysis may lead to substantial loss of data it is very convenient in its application and does not result in bias for data that are MCAR (see e.g., Bartlett, Harel and Carpenter, 2015; Pepinsky, 2018; Ross, Breskin and Westreich, 2020; van Buuren, 2018).

After evaluating the data frame structure, the function “glm” from R (R Core Team, 2022) was used to fit a logistic model. The “glm()” function is a function in the R base stats for generalised models. Numeric variables were z transformed to establish comparability between different scales and value ranges.

To assess the goodness-of-fit of the model, the Hosmer-Lemeshow Test with the function “performance_hosmer” of the “performance” package by Lüdecke *et al.* (2021) which builds on Hosmer and Lemeshow (2000) was used. A p-value greater than 0.05 indicates that there is no significant evidence that the model is poorly fit (see e.g., Hilbe, 2015; Rossi, 2022).

With the aim of identifying the surplus information possibly gained from adding air pollution markers and the socioeconomic status to the variables of the ESC SCORE 2, the likelihood ratio test (LRT, Equation 3-3) (Hilbe, 2015) was used. It compares the deviances of two models (see e.g., Chatterjee and Simonoff, 2013). For this aim, the “lrtest” function of the r package “lmtest” (Zeileis and Hothorn, 2002) was used. By using the likelihood ratio test, a model containing the ESC SCORE 2 variables (model 1) was compared with a model containing the ESC SCORE 2 variables and an air pollution marker (model 2). Afterwards model 2 was compared with a model containing the ESC variables, an air pollution marker and the purchasing power (model 3). This was done in single air pollution marker models for each air pollution marker respectively.

Equation 3-3- Likelihood ratio test

$$Likelihood\ ratio\ test = -2\{\mathcal{L}_{reduced} - \mathcal{L}_{full\mathcal{L}}\}$$

$\boldsymbol{\mathcal{L}_{reduced}}$	Log-likelihood reduced model
$\boldsymbol{\mathcal{L}_{full\mathcal{L}}}$	Log-likelihood full model

To meaningfully interpret the coefficients of the regression model, coefficients of the regression model were exponentiated to extract their ORs.

The hypotheses underlying these steps were:

H$_0$: Model 1 and model 2 fit the data equally well. There is no difference between the residual deviance in the full and the reduced model.

H$_1$: Model 2 surpasses model 1 in terms of fitting the data significantly.

<u>Model 1</u>

ESC SCORE 2 variables

(SCORE2 working group and ESC Cardiovascular risk collaboration, 2021, p. 2444)

Mortality ~ variables of the ESC Score 2

<u>Model 2</u>

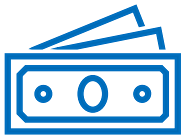
ESC SCORE 2 variables

\+ 

Mortality ~ variables of the ESC Score 2 + mean air pollution exposure (time period with maximum AUC)

<u>Model 3 (full model)</u>

ESC SCORE 2 variables

\+  + 

Mortality ~ variables of the ESC Score 2 + mean air pollution exposure (time period with maximum AUC) + purchasing power

Figure 3-4: Logistic regression models

Within this dissertation, further packages were used in the analysis process: gridExtra (Auguie, 2017), units (Pebesma, Mailund and Hiebert, 2016), stringr (Wickham, 2022), hmisc (Harrell, 2022), and corrplot (Wei and Simko, 2021).

3.5. Flowchart of patients and processes

Figure 3-5 – Flowchart of patients and processes gives an overview of patients and processes. The whole cohort consisted of 4836 patients. 224 patients were without mortality data, and for 2 patients there was no information about the residing area. The removal of these patients resulted in a cohort of 4610 patients. According to the processes described earlier, air pollution and SES data were added. ROC and AUC analyses were conducted and the time period for resulting in a maximum AUC was identified. Multivariate logistic regression was conducted only for those patients with an available mean air pollution concentration for this time period (complete case analysis). This resulted in different sizes of analysis groups for logistic regression among different air pollution markers.

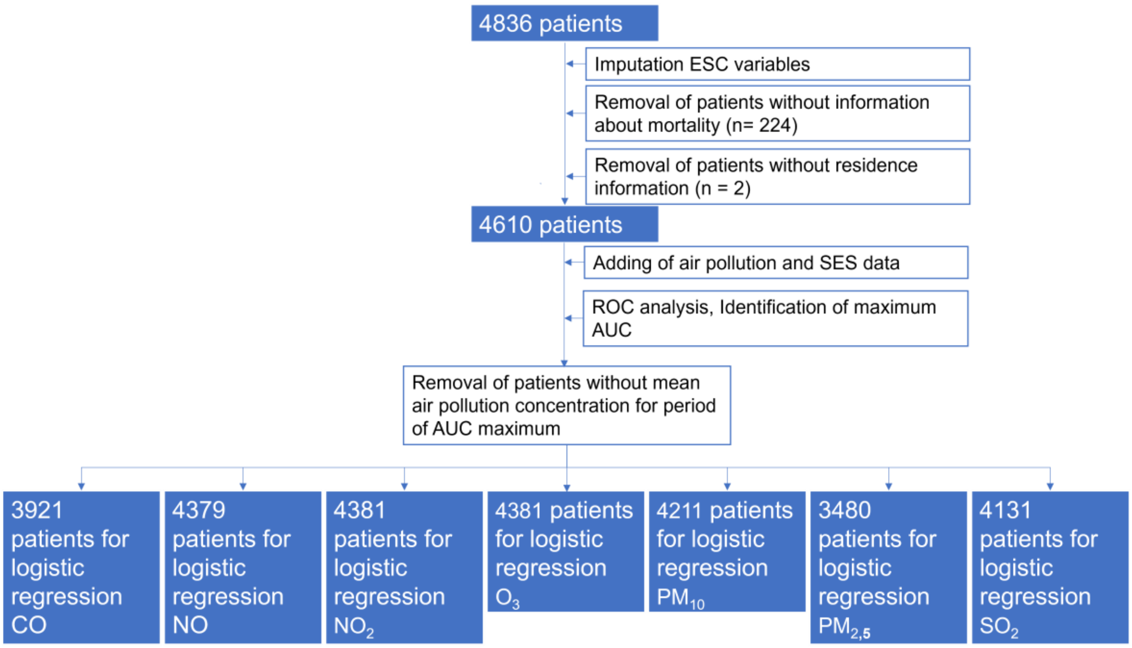


Figure 3-5 – Flowchart of patients and processes

After having introduced the methods, the results are presented in the next chapter.

4. Results

4.1. Descriptive Analysis of the data set

4.1.1. Descriptive analysis of patient-level data

For an overview see Table 4-1- Baseline characteristics

The cohort consisted of 4610 individuals that were scheduled for coronary angiography within 2010 and 2019. 32% patients were female and 68% male. The median age of the cohort was 69.0 years [Confidence Interval (CI): 59.8 - 75.6 years]. A total of 1122 (24.3%) of the patients were dead when mortality information was inquired in

September 2022. The median follow-up time amounted to approximately 10 years (3603 days [IQR (Interquartile Range): 3267 days – 4051 days]. On the day of study inclusion, arterial hypertension was detected in 85%. 81% suffered from dyslipidaemia, 28% were diagnosed with diabetes mellitus, 19% were currently smoking, 35% were obese, and 35% had a family history of cardiovascular diseases. More than half of the cohort was previously diagnosed with coronary artery disease, 35% had undergone a previous percutaneous coronary intervention, 13% had undergone coronary artery bypass surgery, and 20% had a history of myocardial infarction. The ESC SCORE 2 which was calculated with the imputed variables yielded a median of 15 [IQR: 12-20].

Table 4-1- Baseline characteristics of the study cohort

This table provides an overview of the baseline characteristics of the study cohort.

Definition of abbreviations: BMI = body mass index, CAD = coronary artery disease, PCI = percutaneous coronary intervention, CABG = coronary artery bypass grafting
Data are presented as count (percentage) or median [interquartile range]

	Type and Unit	**Data availability**	**All**	**female**	**male**	**p-value**
Gender	n (%)		n = 4610 (100)	n = 1473(32.0)	n = 3137(68.1)	
Age	median (IQR), [years]	4610 / 4610	69.03 [59.8-75.6]	70.15 [60.8-76.1]	68.41 [59.4-75.3]	0.001
Cardiovascular risk factors						
Arterial Hypertension	n (%)	4592 / 4610	3882 (84.5)	1217 (83.0)	2665 (85.3)	0.05
Dyslipidaemia	n (%)	4586 / 4610	3730 (81.3)	1220 (83.3)	2510 (80.4)	0.02
Diabetes Mellitus	n (%)	4596 / 4610	1273 (27.7)	378 (25.7)	895 (28.6)	0.04
Smoking	n (%)	4309 / 4610	839 (19.5)	229 (16.7)	610 (20.8)	0.001
Family History	n (%)	4070 / 4610	1412 (34.7)	514 (39.4)	898 (32.5)	<0.001
Obesity (BMI>30)	n (%)	4588 / 4610	1639 (35.7)	549 (37.5)	1090 (34.9)	0.1
History						
Known CAD	n (%)	4604 / 4610	2369 (51.5)	545 (37.0)	1824 (58.2)	<0.001
Known PCI	n (%)	4600 / 4610	1591 (34.6)	352 (24.0)	1239 (39.6)	<0.001
Known CABG	n (%)	4600 / 4610	560 (12.2)	98 (6.7)	462 (14.8)	<0.001

History of myocardial infarction	n (%)	4595 / 4610	926 (20.2)	196 (13.3)	730 (23.4)	<0.001
ESC Score 2 (2021)*[1]	median (IQR), [n]	4610 / 4610	15 (12-20)	13 (9-17)	17 (13-21)	<0.001

[1] ESC SCORE 2 calculated from imputed variables

4.2. Descriptive analysis of the geocoding process and the assignation of air pollution data with individual patients

The level of greenness indicates the number of patients living in a post code area. Most patients are located within central Hesse. Areas where most patients are located are only sparsely covered with air monitoring sites (Figure 4-1). The level of blueness indicates the amount of purchasing power for a post code. For the purchasing power there appears to be little variation for the areas where most patients are located (Figure 4-1). While the metropolitan area around Frankfurt accounts a higher density of air monitoring sites and higher purchasing power.

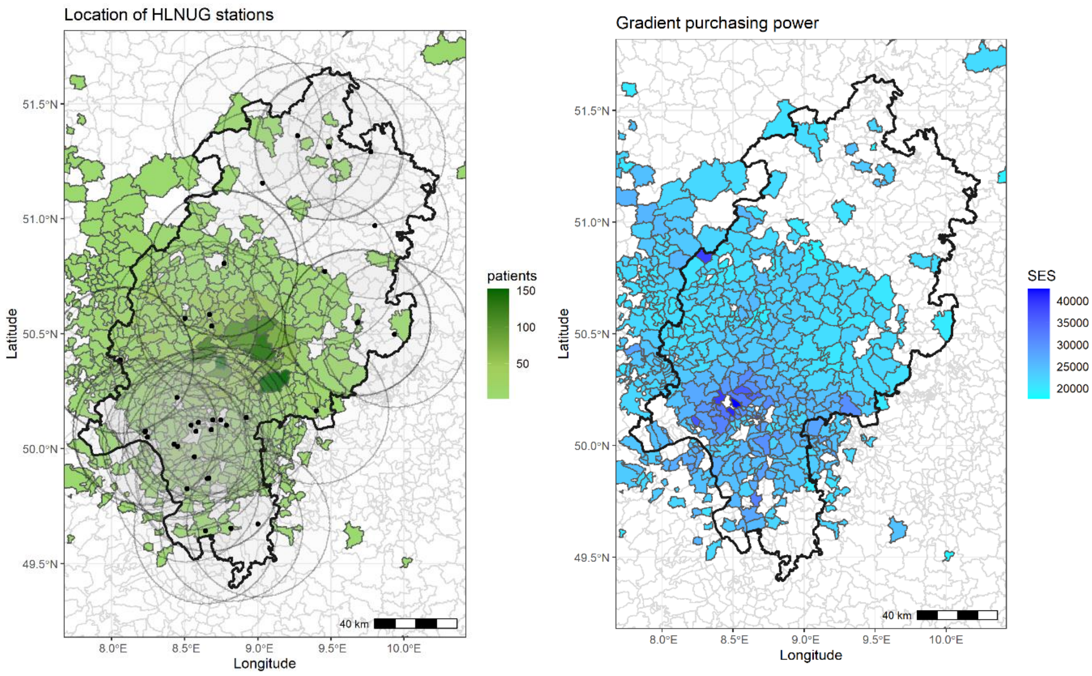


Figure 4-1- Location of HLNUG stations, patients and gradient of purchasing power

The map on the left side represents the location of patients according to their postal code. The structures surrounded in light grey are the postal code areas which are coloured in green gradients according to the number of patients that live there. Thus, the dark green areas in the centre are those areas where most of the patients from the cohort are located. The black dots are the locations of the air monitoring sites while the black circles represent the 35 km circumferences surrounding an air pollution monitoring station. The map on the right side displays the socioeconomic status (SES, €) of the postal code areas in form of the purchasing power.

4.3. Missing data analysis and imputation

4.3.1. Missing patient-level data

4.3.1.1. Missing data among the ESC SCORE 2 variables

Figure 4-2 shows the missing values among the variables needed to calculate the ESC SCORE 2. The variables contained the following missing percentages: smoking (6.5%), total cholesterol (5.5%), HDL-cholesterol (24,6%), systolic blood pressure (4.2%), and sex and age (each below 0.1%). In total, 3.3% of the variables needed to calculate the ESC SCORE 2 were missing. Figure 4-3 illustrates the combinations of missingness.

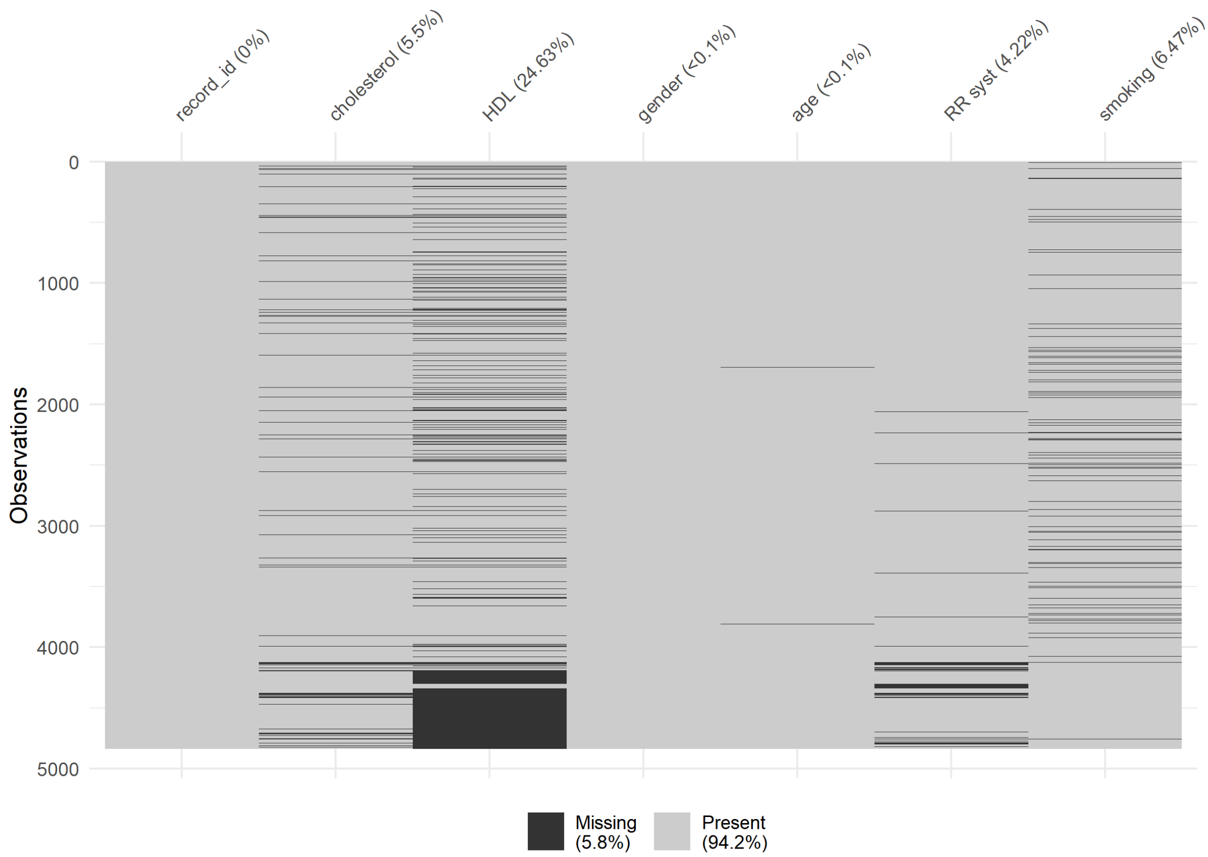


Figure 4-2 – Plot of missing data for the ESC SCORE 2 variables

The columns show the ESC SCORE 2 variables and the rows are the observations of the patients for those variables. There is no row order but one row entails the variables for one patient. Missing values are illustrated in black, available values in grey. The percentage of missing values for each variable is provided in brackets.
cholesterol = total cholesterol, HDL = HDL cholesterol

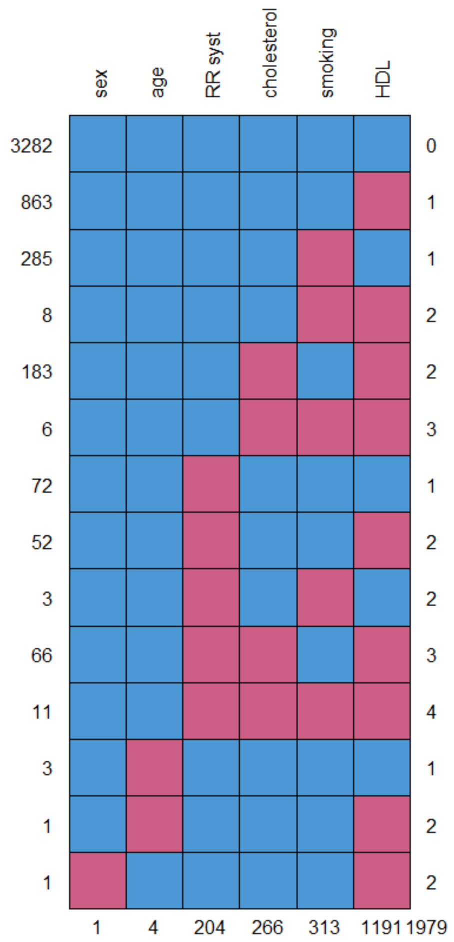


Figure 4-3 plot of patterns of missing data among the ESC SCORE 2 variables

The rows and the columns of this graph are ordered according to the number of missing values. The columns are the variables, to the left are the numbers of missing values in total, to the right the number of missing values in the combination of a certain row (e.g. row 4: there are 8 patients where the values for the variables HDL and smoking are missing (2 variables)) and the bottom line disaggregates the number of missing values within each variable (e.g. the age is missing for 4 patients). (cholesterol = total cholesterol, HDL = HDL cholesterol)

Given the absence for plausible reasons of the missingness pattern to be not coincidental among these variables, it was assumed that the missing patterns were at least MAR or even MCAR.

Missing values were imputed. After the imputation process, sensitivity analysis was conducted to compare imputed and not imputed values. Table 4-2 illustrates the imputed dataset of the ESC SCORE variables in comparison to the original dataset. From this analysis, it could be concluded that the imputed values appear plausible.

Table 4-2 Comparison of original and imputed ESC SCORE 2 variables

This table provides an overview of the original ESC SCORE 2 variables in comparison to the imputed variables. The variables total cholesterol and HDL cholesterol were imputed and subtracted afterwards to build the variable Non-HDL cholesterol.

ESC SCORE 2 variables	original	imputed
Total Cholesterol in mg/dL (median [IQR])	185.0 [157.0 – 219.0]	186.0 [157.0 – 220.0]
HDL cholesterol in mg/dL (median [IQR])	50.0 [41.0 – 61.0]	49.0 [40.0 - 60.0]
Sex (male)	3288 1 NA	3288
Age in years 4(median[IQR])	69.0 [59.7 – 75.5]	69.0 [59.7 – 75.5]
Systolic blood pressure in mmHg (median[IQR])	135 [120 - 150]	135 [120-150]
Smoking (smoker)	896 313 NA	934

4.3.1.2. Missing mortality data

Figure 4-4 demonstrates the percentage of missing values among the variable mortality. As already mentioned in the methods chapter, missingness of the MCAR type was assumed due to the marginal share of missing values (4.6%) for this variable. Hence, it was decided to rely on complete case analysis for the mortality variable.

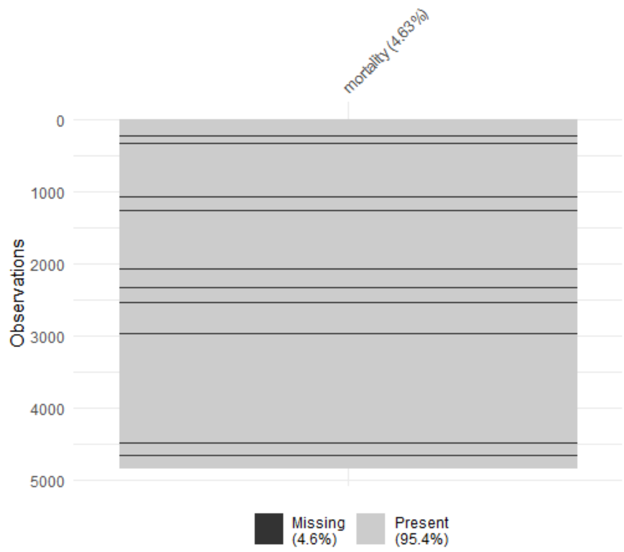


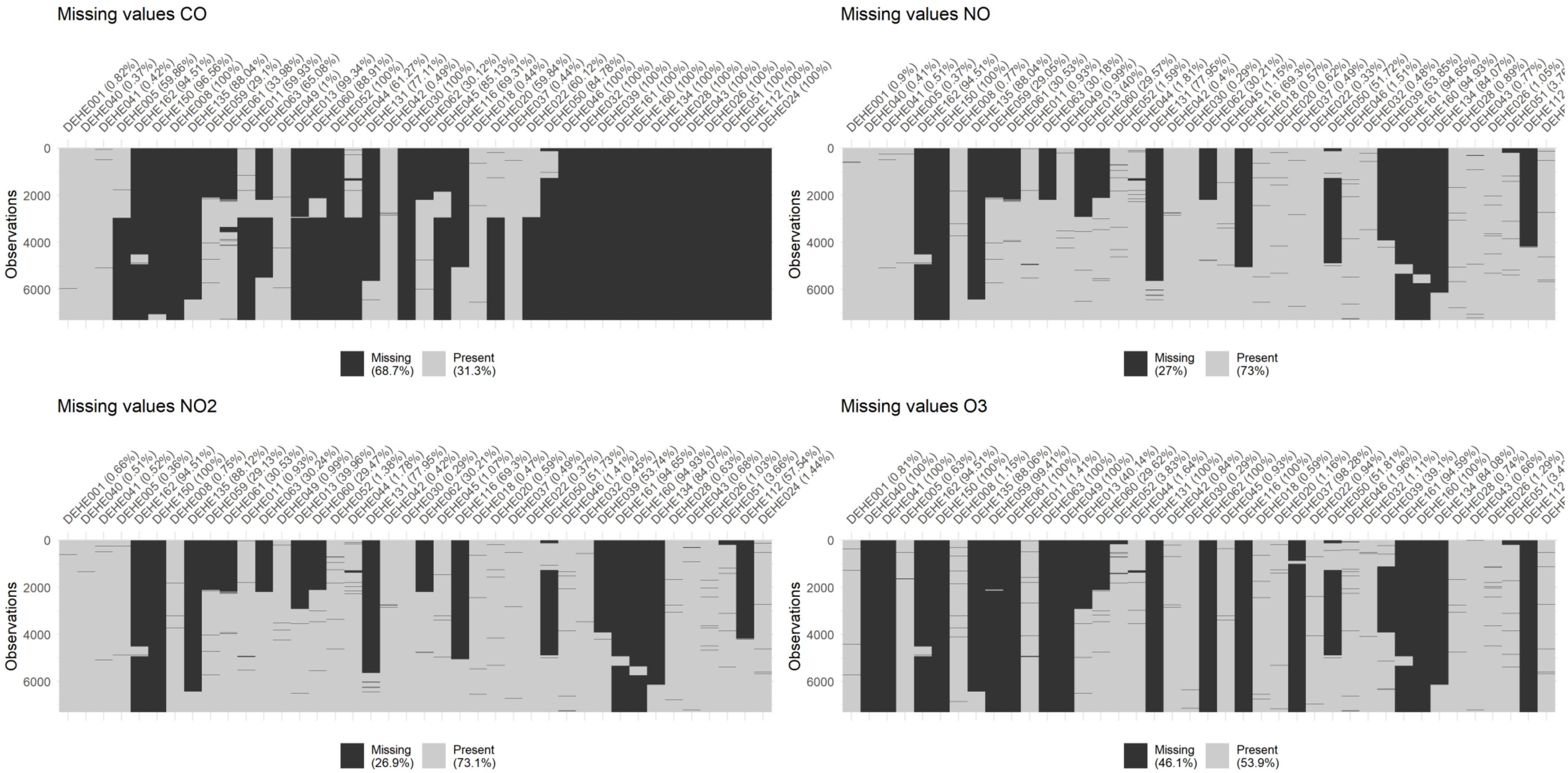
Figure 4-4 – Missing data for the variable mortality

Missing values are illustrated in black, available values in grey. The percentage of missing values for a variable is provided in brackets.

4.3.2. Missing air pollution data

4.3.2.1. Missing air pollution data

Figure 4-5 shows the availability of data for the different air pollution markers broken down according to the stations of the HLNUG for the time period between the 1st of January 2000 until the 31st of December 2019. In descending order the air pollution marker variables contained the following missing percentages: $PM_{2.5}$ (89%), CO (68.7%), SO_2 (64%), O_3 (46%), PM_{10} (36%), NO (27%), NO_2 (27%).


Missing values CO
Observations
Missing (68.7%)
Present (31.3%)
Missing values NO
Observations
Missing (27%)
Present (73%)
Missing values NO2
Observations
Missing (26.9%)
Present (73.1%)
Missing values O3
Observations
Missing (46.1%)
Present (53.9%)

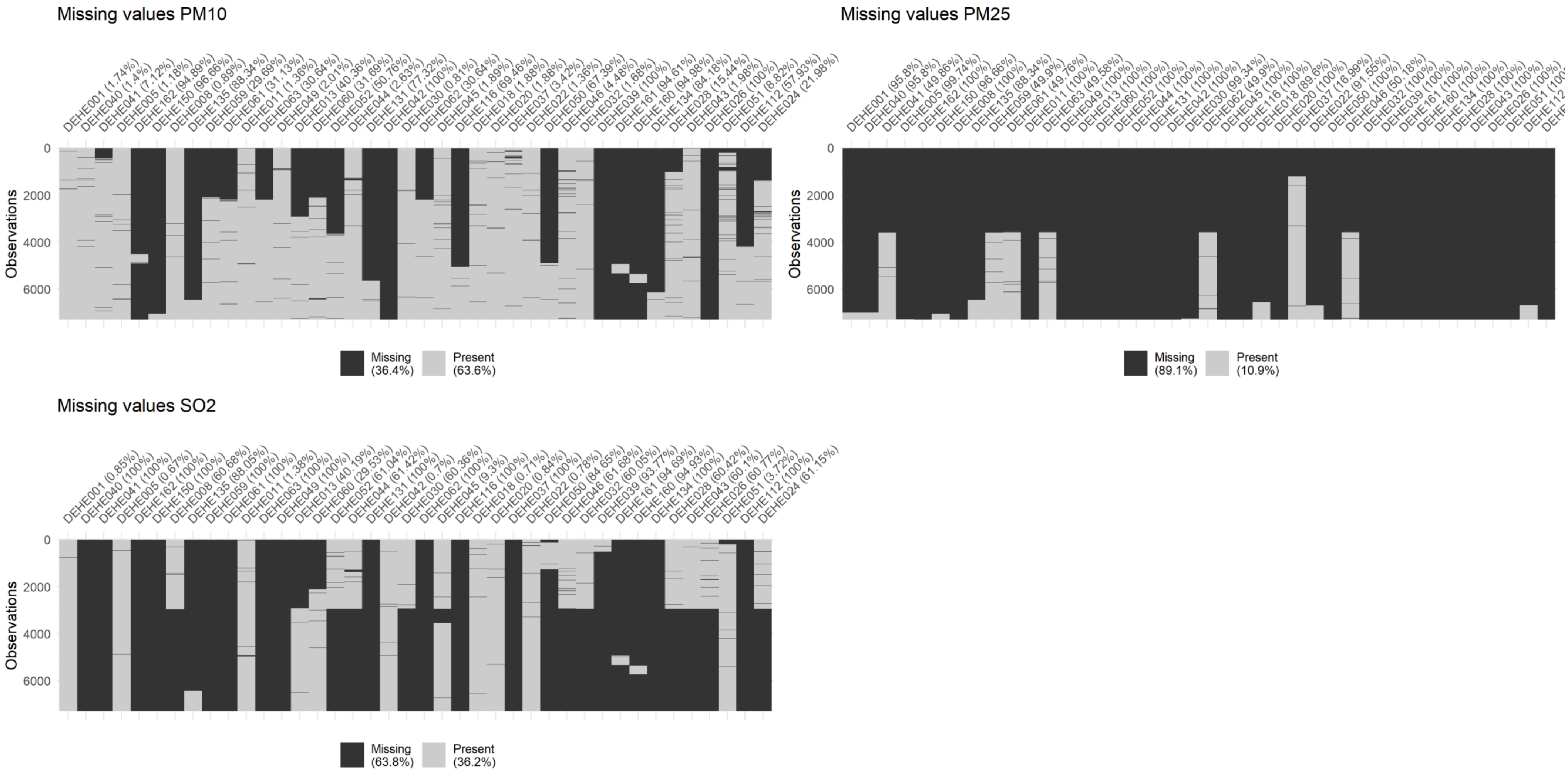


Figure 4-5 missing plots according to air pollutants

Each graph shows the pattern of missing data for one air pollution marker. The columns represent the stations. The missingness of the respective air pollution marker is given in brackets. The rows are days from the 1st January 2000 (row 1) until the 31st December 2019 (row 7305). Missing values are coloured in black, while available values for air pollutants are coloured in grey. Definition of abbreviations: CO = carbon monoxide, NO = nitrogen oxide, NO2 = nitrogen dioxide, O3 = ozone, PM10 = particulate matter with a diameter < 10 µm, PM2.5 = particulate matter with a diameter < 2.5 µm, SO2 = sulphur dioxide

4.3.2.2. *Imputation of air pollution data*

Figure 4-6 shows the graph of the multivariate imputed dataset for CO of the station DEHE001 as an example of the imputation for the air pollution variables. While some values are imputed below 0, the majority of the red dots appear adequate within the blue trends and thus seem plausible. For this example no values were imputed below 0. Air pollution markers with more than 10% missing within the set time period were not imputed.

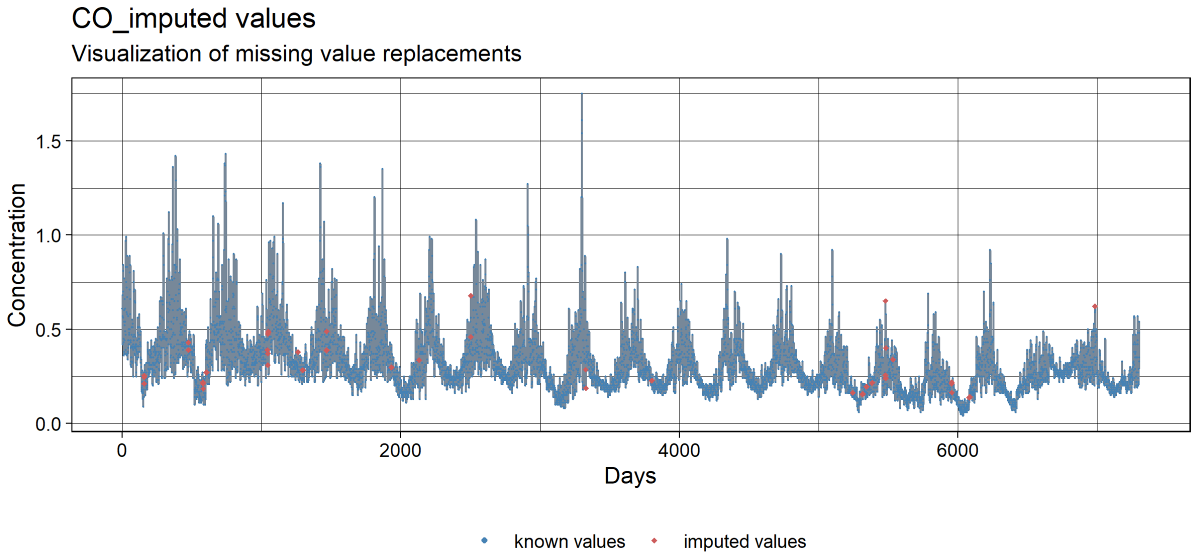


Figure 4-6- CO imputed values station DEHE001

The x-axis depicts the time (01/01/2000 – 31/12/2019) and the y-axis depicts the concentration of CO (carbon monoxide). Known values are shown in blue, imputed values for missing values are shown in red.

4.4. Variation of exposure time periods with the highest prognostic ability for respective air pollution markers

Figure 4-7 shows the ROC curve for the association between ESC SCORE 2 and mortality. The AUC of the ESC SCORE 2 on the day of study enrolment is 0.57. This value represents the measure of the discriminatory ability in its prediction of mortality of the ESC SCORE 2. This AUC was the reference for the comparison with the discriminatory ability of the different air pollution markers.

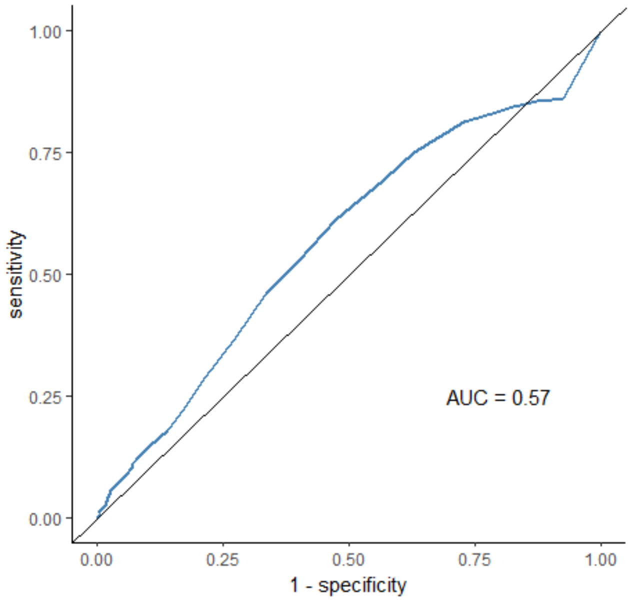


Figure 4-7 - ROC curve for the ESC SCORE 2 on the day of study enrolment

This ROC graph plots the false positive rate (1-specificity) on the x-axis against the true positive rate (sensitivity) on the y-axis for different thresholds of the ESC SCORE 2 in its prediction of mortality. AUC = area under the curve

As explained in the methods section, other than the ROC curve for the ESC SCORE 2 which was calculated for the day of study enrolment, air pollution exposure and its discriminatory ability was measured over time periods within three years. Time periods correspond to mean exposures for all days until study enrolment within 3 years, i.e. -1094 days until study enrolment, and thus 1095 time periods. Figure 4-9 illustrates an overview for the AUCs for different air pollution markers for 1095 time periods. Figure 4-9 provides information that for the exposure to $PM_{2.5}$ and to SO_2 the AUC for a certain time period surpasses the AUC of the ESC SCORE 2 for the day of study enrolment. Thus, the graphs for the exposure to $PM_{2.5}$ and SO_2 are described in more detail below. Plot statistics for the graphs can be found in Table 4-3.

Table 4-3 Overview AUC plot statistics

This table provides an overview of the time periods of air pollution exposure yielding the maximum AUC. The maximum AUC is provided with a 95% CI. The last column contains the mean exposure of the respective air pollution markers for the time period yielding in the maximum AUC value.
Definition of abbreviations: CO = carbon monoxide, NO = nitrogen oxide, NO_2 = nitrogen dioxide, O_3 = ozone, PM_{10} = particulate matter with a diameter < 10 µm, $PM_{2.5}$ = particulate matter with a diameter < 2.5 µm, SO_2 = sulphur dioxide, AUC = area under the curve, CI = confidence interval, sd = standard deviation

Air pollution marker	Time period for AUC max till study enrolment	Maximum AUC (95% CI)	Mean air pollution exposure (sd) [μg/m^3]
CO	3 days	0.52 (0.5, 0.54)	0.40 (0.17)
NO	1077 days	0.54 (0.52, 0.56)	28.98 (17.14)
NO_2	1014 days	0.55 (0.53, 0.57)	33.62 (12.72)
O_3	348 days	0.52 (0.50, 0.54)	45.82 (12.84)
PM_{10}	232 days	0.52 (0.50, 0.54)	20.42 (4.50)
$PM_{2.5}$	402 days	0.59 (0.56, 0.61)	17.37 (1.89)
SO_2	1058 days	0.59 (0.57, 0.60)	1.78 (0.63)

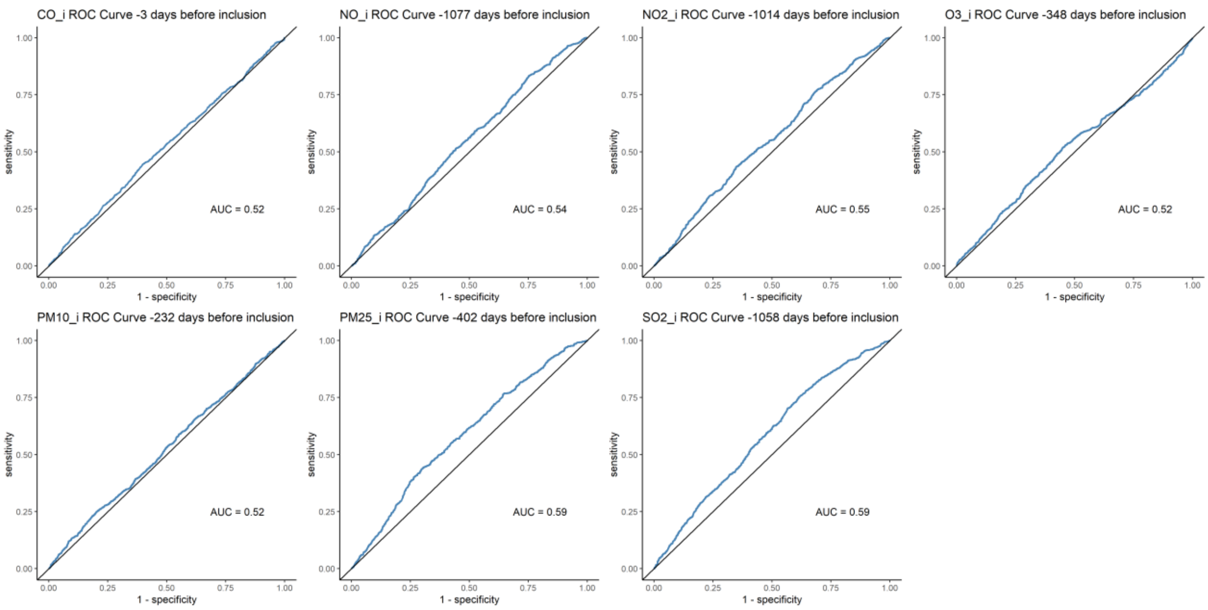
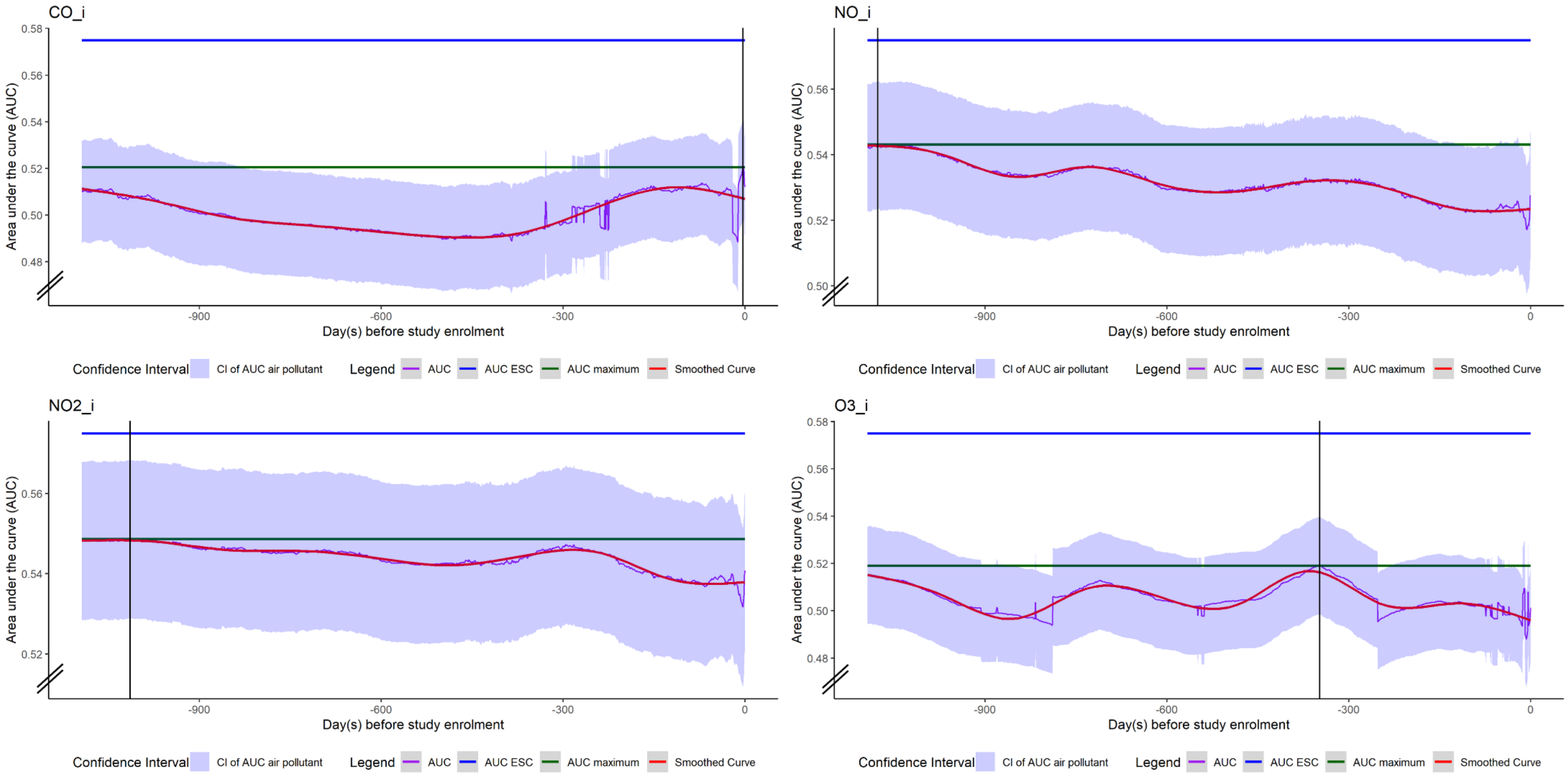


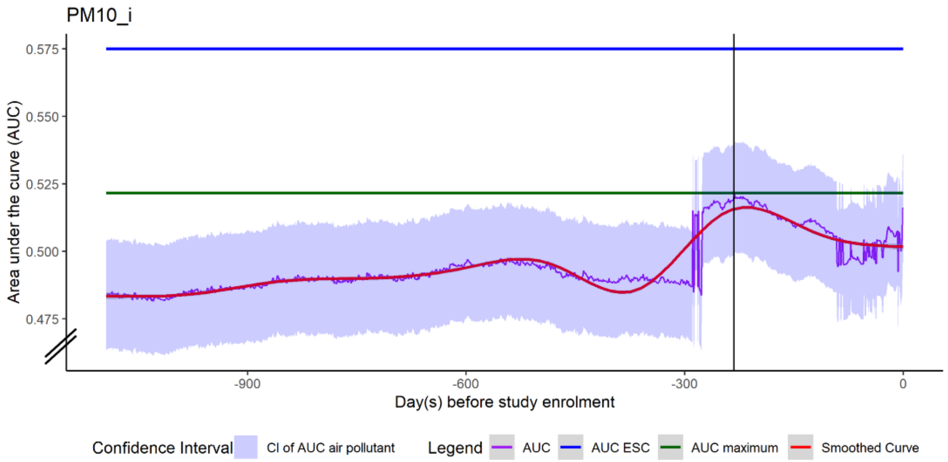
Figure 4-8 Overview of ROC curves for air pollution markers for the time periods yielding in the highest AUC values, respectively

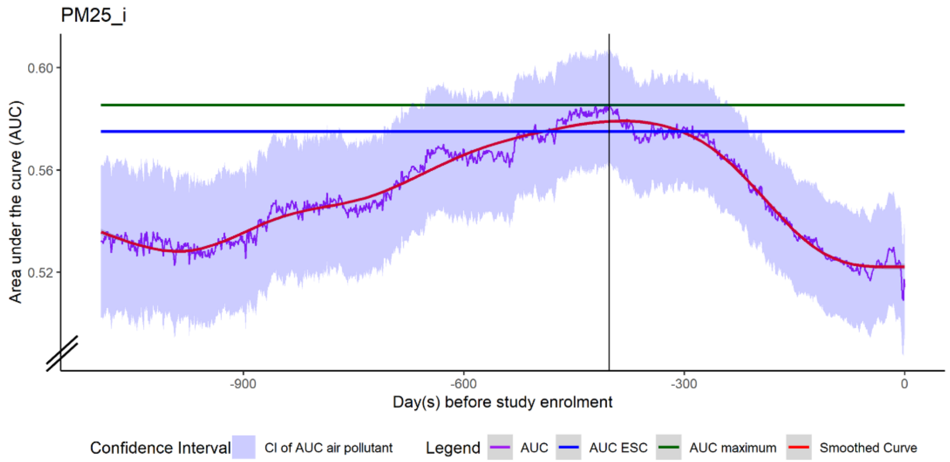
Overview of ROC graphs for air pollution markers for the time periods yielding in the highest AUC values, respectively. The ROC graphs plots the false positive rate (1-specificity) on the x-axis against the true positive rate (sensitivity) on the y-axis for different thresholds of the mean exposure to an air pollution marker for the time period of the maximum AUC value in its prediction of mortality.

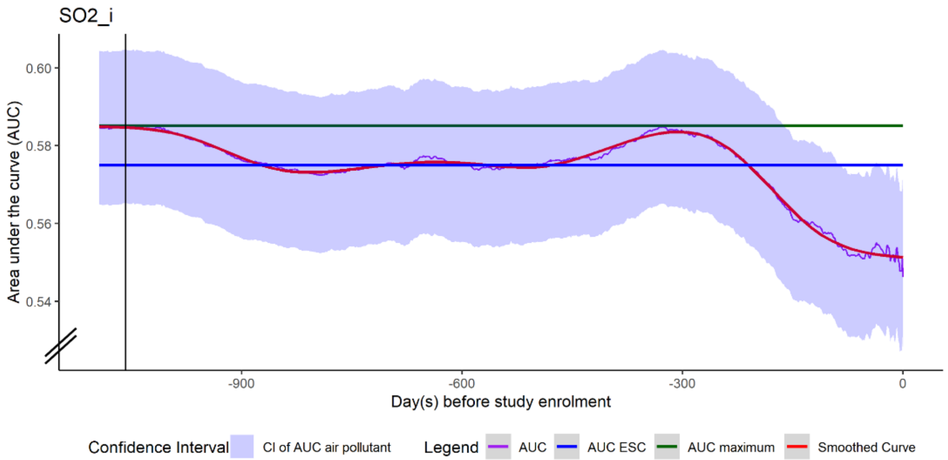
Definition of abbreviations: CO_i = carbon monoxide, NO_i = nitrogen oxide, NO2_i = nitrogen dioxide, O3_i = ozone, PM10_i = particulate matter with a diameter < 10 μm, PM2.5_i = particulate matter with a diameter < 2.5 μm, SO2_i = sulphur dioxide (the "_i" indicates the use of the air pollution data after imputation), AUC = area under the curve



CO_i
NO_i
NO2_i
O3_i
Area under the curve (AUC)
Day(s) before study enrolment
Confidence Interval
CI of AUC air pollutant
Legend
AUC
AUC ESC
AUC maximum
Smoothed Curve







The AUC for the discriminatory ability of the mean exposure to an air pollution marker for a certain time period was plotted on the y-axis against time on the x-axis. The day of enrolment is the very right point on the x-axis (0 days) and time prior to enrolment is plotted on the x- axis in the opposite direction (until – 1094 days). The red line depicts the AUC of the ROC graphs. The green horizontal line shows the maximum AUC of an air pollution marker. The black vertical lines are set at the maximum point of the red curve and thus mark the lower end of the time period until study enrolment. The blue horizontal line resembles the AUC of the ESC SCORE 2 on the day of study enrolment.

Definition of abbreviations: CO_i = carbon monoxide, NO_i = nitrogen oxide, NO2_i = nitrogen dioxide, O3_i = ozone, PM10_i = particulate matter with a diameter < 10 µm, PM2.5_i = particulate matter with a diameter < 2.5 µm, SO2_i = sulphur dioxide (the "_i" indicates the use of the air pollution data after imputation), AUC = area under the curve, CI = confidence interval, ESC = ESC SCORE 2

Figure 4-9- overview of ROC analysis Results

Figure 4-10 shows AUC for 1095 time periods prior to study enrolment for the discriminatory ability of the mean exposure to $PM_{2.5}$ in its prediction of mortality. The red curve has a variance of 0.00038 and a standard deviation of 0.01959. Read from the right to the left: AUC values, and thus the discriminatory ability of $PM_{2.5}$, increase until its maximum for the time period of 402 days until study enrolment (AUC = 0.59). This maximum is marked with the green horizontal line and the black vertical line. The maximum AUC for exposure to $PM_{2.5}$ for the time period of 402 days (AUC = 0.59) surpasses the AUC of the ESC SCORE 2 upon study enrolment (AUC = 0.57). Thus, the maximum discriminatory ability of exposure to $PM_{2.5}$ lies in the time period of approximately 1 year. This maximum predictive time period of 402 days was used as the basis of further analysis.

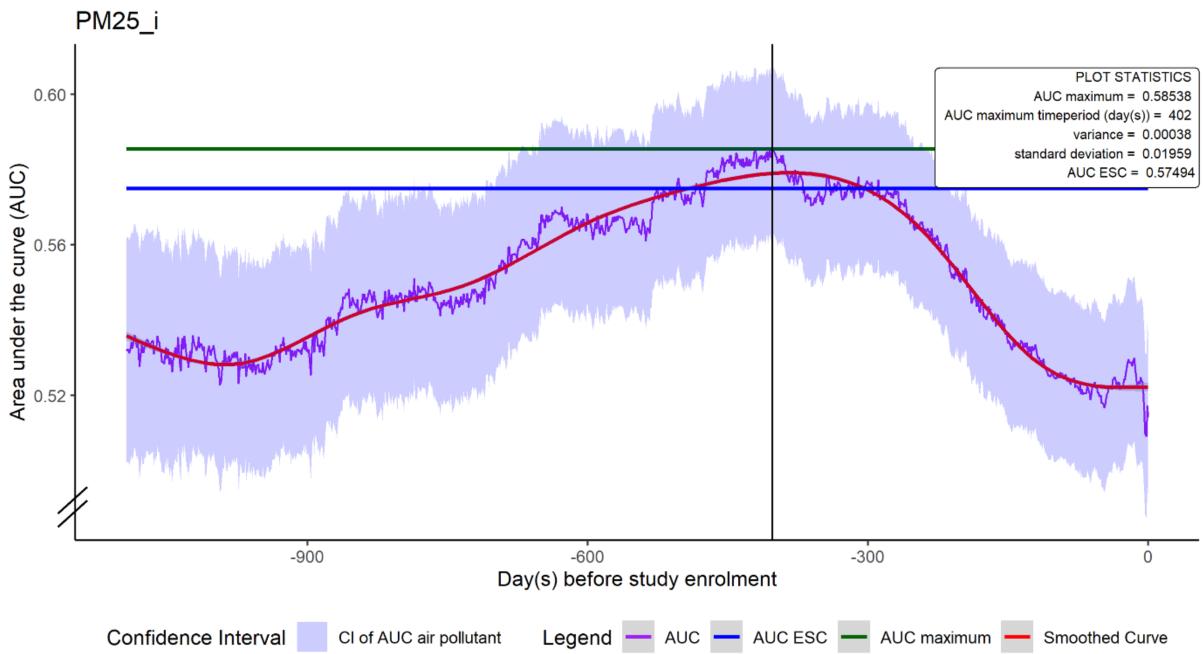


Figure 4-10- AUC curve of $PM_{2.5}$

The 0 to the right of the x-axis represents the day of study enrolment. The blue horizontal line shows the AUC for the ESC SCORE 2 on the day of study enrolment. The red curve illustrates the AUCs for the mean exposure to $PM_{2.5}$ for different time periods for days before study enrolment until study enrolment.

Definition of abbreviations: PM2.5_i = particulate matter with a diameter < 2.5 µm (the "_i" indicates the use of the air pollution data after imputation), AUC = area under the curve, CI = confidence interval, ESC = ESC SCORE 2

The AUC curve for SO_2 (Figure 4-11) can be interpreted like the graphs discussed before. The red curve has a variance of $8*10^{-5}$ and a standard deviation of 0.0089. Read from the right to the left: AUC values, and thus the discriminatory ability of SO_2, have their maximum for the time period of 1058 days until study enrolment. This maximum is marked with the green horizontal line and the black vertical line. As with $PM_{2.5}$, the maximum AUC for exposure to SO_2 for the time period of 1058 days (AUC = 0.59) surpasses the AUC of the ESC SCORE 2 upon study enrolment (AUC = 0.57). Thus, the maximum discriminatory ability of exposure to SO_2 lies in the time period of approximately 3 years.

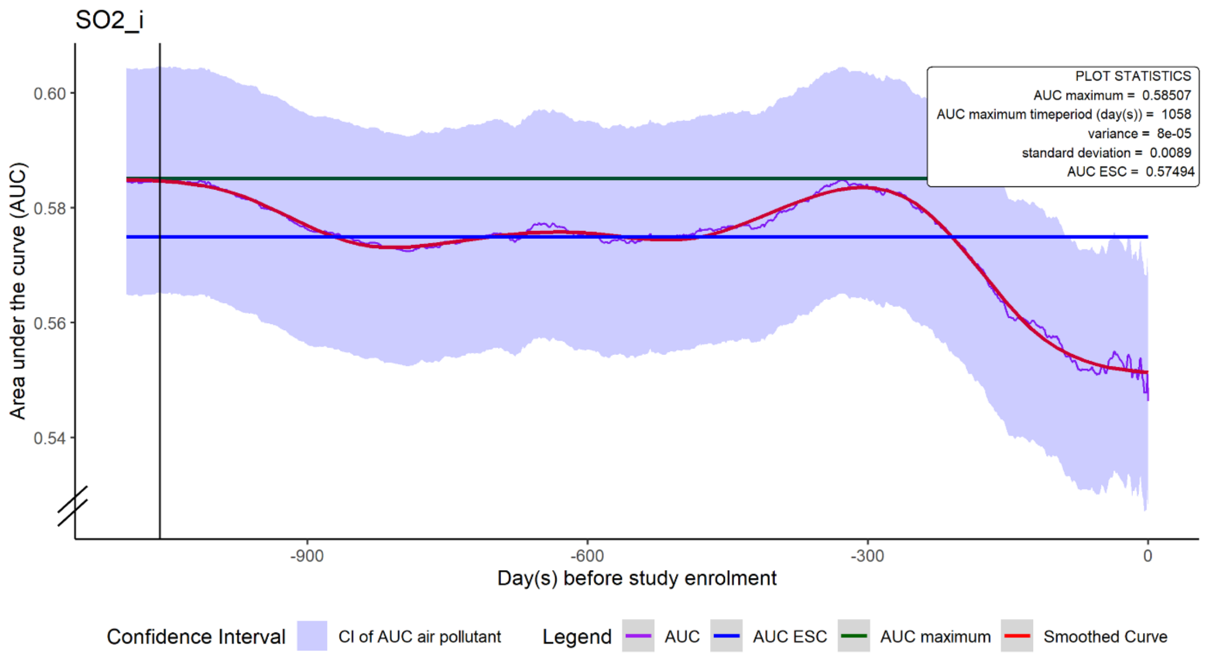


Figure 4-11- AUC curve of SO_2

The 0 to the right of the x-axis represents the day of study enrolment. The blue horizontal line shows the AUC for the ESC SCORE 2 on the day of study enrolment. The red curve illustrates the AUCs for the mean exposure to SO2 for different time periods for days before study enrolment until study enrolment.

Definition of abbreviations: SO2_i = sulphur dioxide (the "_i" indicates the use of the air pollution data after imputation), AUC = area under the curve, CI = confidence interval, ESC = ESC SCORE 2

4.5. Significant benefits in considering air pollution markers compared to patient immanent risk factors only

Different models to predict the probability of the binary outcome mortality were fit. Model one consisted of the input variables of the ESC SCORE 2. Model 2 consisted of the input variables of the ESC SCORE 2 adjusted for the mean air pollution exposure for the time period that yielded the maximum AUC (see Table 4-3). These periods with the maximum AUC had the maximum association with mortality.: The ROC analysis of the exposure to air pollution discussed in 4.4 indicated different time periods to achieve a maximum AUC for different air pollution markers (see chapter 4.4). These results formed the basis for the application of logistic regression of model 2.

Model 3 was then further adjusted for the input variable purchasing power as an indicator of the SES.

Numeric variables were z-transformed. The goodness-of-fit of the respective models was assessed with the Hosmer-Lemeshow test. A $p > 0.05$ indicates a well-fit model. Furthermore, the potential benefit of added variables was analysed by use of the likelihood ratio test, a $p < 0.05$ indicates a significant benefit.

Table 4-4 gives an overview of the logistic regression analysis. All ESC SCORE 2 variables were significantly associated with mortality except non-HDL-cholesterol for the model including CO. In model 2 (mortality predicted by the ESC SCORE 2 variables and mean air pollution concentration), the exposure to NO (OR 1.18; 95% CI: 1.1, 1.27), NO2 (OR 1.24; 95% CI: 1.15, 1.34), PM_{10} (OR 1.11; 95% CI: 1.03, 1.20), $PM_{2.5}$ (OR 1.54; 95% CI: 1.40, 1.68), and SO_2 (OR 1.41; 95% CI: 1.31, 1.52) were positively and significantly associated with mortality, while the association for CO (OR 0.93; 95% CI: 0.86, 1.01) and O_3 (OR 0.99; 95% CI: 0.92, 1.07) was non-significant. Logistic regression analysis results are discussed in more detail below. Forest plots for the ORs of the multivariate second models with all air pollution markers except for $PM_{2.5}$ and SO_2 can be found in the appendix.

Table 4-4 Overview Logistic regression models

This table provides an overview of the logistic regression model analysis. The respective models 1 consist of the input variables of the ESC SCORE 2. Respective models 2 consisted of the input variables of the ESC SCORE 2 adjusted for the mean air pollution exposure for the time period that yielded the maximum AUC and respective models 3 were further adjusted for the input variable purchasing power as an indicator of the SES. The odds ratios were calculated for the air pollution markers in each model. Further likelihood ratio tests were conducted to determine the surplus information gained by adding further input variables and the Hosmer-Lemeshow test was conducted to assess the goodness of fit of respective models. The results of the conducted tests are all provided within this table.

Definition of abbreviations: CO = carbon monoxide, NO = nitrogen oxide, NO_2 = nitrogen dioxide, O_3 = ozone, PM_{10} = particulate matter with a diameter < 10 µm, $PM_{2.5}$ = particulate matter with a diameter < 2.5 µm, SO_2 = sulphur dioxide, OR = odds ratio, HLT = Hosmer-Lemeshow test, LRT = Likelihood ratio test, AUC = area under the curve
*HLT: $p < 0.05$ no good fit, $p > 0.05$ good fit; LRT: $p < 0.05$ significant benefit, $p > 0.05$ no benefit; * the OR cannot be indicated for the air pollution markers in the respective Models 1 as they are not included in these models*

	CO	**NO**	**NO_2**	**O_3**	**PM_{10}**	**$PM_{2.5}$**	**SO_2**
Number of patients (mean air pollution concentration available for time period)							
	3921	4379	4381	4381	4211	3480	4131
Model 1 : Mortality ~ variables of the ESC Score 2							
OR	*	*	*	*	*	*	*
HLT	$\chi^2 = 17.30$ $p < 0.05$	$\chi^2 = 12.14$ $p > 0.05$	$\chi^2 = 12.25$ $p > 0.05$	$\chi^2 = 12.51$ $p > 0.05$	$\chi^2 = 12.78$ $p > 0.05$	$\chi^2 = 18.37$ $p < 0.05$	$\chi^2 = 10.38$ $p > 0.05$
Model 2: Mortality ~ variables of the ESC Score 2 + mean exposure to air pollution marker (time period of max AUC)							
OR (air pollution marker)	0.93 (95% CI: 0.86, 1.01)	1.18 (95% CI: 1.1, 1.27)	1.24 (95% CI: 1.15, 1.34)	0.99 (95% CI: 0.92, 1.07)	1.11 (95% CI: 1.03, 1.20)	1.54 (95% CI: 1.40, 1.68)	1.41 (95% CI: 1.31, 1.52)
LRT (model 1, model 2)	$\chi^2 = 3.45$ $p > 0.05$	$\chi^2 = 19.50$ $p < 0.05$	$X^2 = 32.20$ $p < 0.05$	$\chi^2 = 0.1$ $p > 0.05$	$\chi^2 = 7.3$ $p < 0.05$	$\chi^2 = 93.1$ $p < 0.05$	$\chi^2 = 83.40$ $p < 0.05$
HLT	$\chi^2 = 12.03$ $p > 0.05$	$\chi^2 = 12.88$ $p > 0.05$	$\chi^2 = 14.57$ $p > 0.05$	$\chi^2 = 16.63$ $p < 0.05$	$\chi^2 = 17.90$ $p < 0.05$	$\chi^2 = 13.18$ $p > 0.05$	$\chi^2 = 18.08$ $p < 0.05$
Model 3: Mortality ~ variables of the ESC Score 2 + mean exposure to air pollution marker (time period of max AUC)+ purchasing power							

OR (air pollution marker)	0.93 (95% CI: 0.86, 1.01)	1.18 (95% CI: 1.10, 1.27)	1.24 (95% CI: 1.15, 1.34)	0.99 (95% CI: 0.92, 1.07)	1.12(95% CI: 1.04, 1.21)	1.53 (95% CI: 1.40, 1.68)	1.44 (95% CI: 1.34, 1.56)
LRT (model 2, model 3)	$\chi^2 = 3.08$, $p > 0.05$	$\chi^2 = 1.52$, $p > 0.05$	$\chi^2 = 0.12$, $p > 0.05$	$\chi^2 = 1.99$ $p > 0.05$	$\chi^2 = 2.03$ $p > 0.05$	$\chi^2 = 0.22$ $p > 0.05$	$\chi^2 = 2.89$ $p > 0.05$
HLT	$\chi^2 = 12.18$ $p > 0.05$	$\chi^2 = 13.51$ $p > 0.05$	$\chi^2 = 14.33$ $p > 0.05$	$\chi^2 = 11.97$ $p > 0.05$	$\chi^2 = 18.83$ $p < 0.05$	$\chi^2 = 15.72$ $p < 0.05$	$\chi^2 = 23.09$ $p < 0.05$

4.5.1. Multivariate model: Patient immanent risk factors of the ESC SCORE 2 with $PM_{2.5}$

The likelihood ratio test comparing model 1 and model 2 demonstrated a significant prognostic contribution ability for the inclusion of the mean $PM_{2.5}$ values for its maximum prediction time period (Likelihood ratio test, χ^2 = 93.1 degrees of freedom, $p < 0.05$). Adding purchasing power as a proxy for the SES to model 2 did not improve the models fit significantly (Likelihood ratio test, χ^2 = 0.22, $p > 0.05$). Thus, the final model consisted of the ESC SCORE 2 variables and the exposure to $PM_{2.5}$ over the time period -402 days till enrolment (Model 2). Model 2 performed well in the goodness-of-fit test (Hosmer- Lemeshow test, χ^2 = 13.18, $p > 0.05$). More details can be found in Table 4-4. Figure 4-12 illustrates the multivariate logistic regression model of mortality predicted by the ESC SCORE 2 risk factors and the mean exposure to $PM_{2.5}$. All coefficients in model 2 were significantly associated with mortality ($p < 0.05$). The risk factors relating to mortality were $PM_{2.5}$ (OR: 1.54, CI: 1.40, 1.68), non-HDL cholesterol (OR: 0.89, CI: 0.81, 0.97), sex[male](OR: 1.37, CI: 1.14, 1.65), age (OR: 2.80, CI: 2.49, 3.12), systolic blood pressure (OR: 0.86, CI: 0.79, 0.93), and smoking (OR: 2.01, CI: 1.60, 2.54, $p < 0.05$ for all).

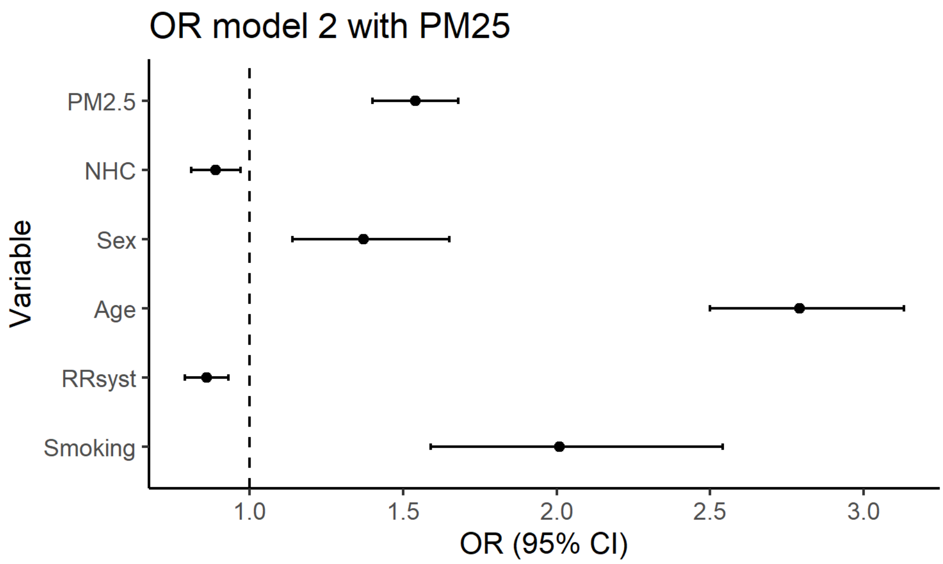


Figure 4-12 OR model 2 with $PM_{2.5}$

ORs of a multivariate logistic regression model of mortality predicted by the ESC SCORE 2 risk factors and the mean exposure to $PM_{2.5}$
Definition of abbreviations: OR = odds ratio, model 2 = Mortality ~ variables of the ESC Score 2 + mean exposure to air pollution marker (time period of max AUC, $PM_{2.5}$ = particulate matter with a diameter < 2.5 µm, sex = sex [male], RRsyst = systolic blood pressure, CI = confidence interval, NHC = non-HDL cholesterol

4.5.2. Multivariate model: Patient immanent risk factors of the ESC SCORE 2 with SO_2

The likelihood ratio test comparing model 1 and model 2 demonstrated a significant prognostic contribution ability for the inclusion of the mean SO_2 values for its maximum prediction time period (Likelihood ratio test, $\chi 2 = 83.40$, $p < 0.05$). Adding purchasing power as a proxy for the SES to model 2 did not improve the models fit signific. antly (Likelihood ratio test, $\chi 2 = 2.89$, $p > 0.05$). Thus, the final model consisted of the ESC SCORE 2 variables and the exposure to SO_2 over the time period -1058 days until en-rolment (Model 2). Model 2 did not perform well in the goodness-of-fit test (Hosmer-Lemeshow test, $\chi 2 = 18.08$, $p < 0.05$). Figure 4-13 illustrates the multivariate logistic regression model of mortality predicted by the ESC SCORE 2 risk factors and the mean exposure to SO_2. All coefficients in model 2 were significantly associated with mortality ($p < 0.05$)The risk factors relating to mortality were SO_2 (OR: 1.41, CI: 1.31, 1.52), non-hdl-cholesterol (OR: 0.88, CI: 0.81, 0.96), sex[male](OR: 1.34, CI: 1.14, 1.59), age (OR: 2.90, CI: 2.61, 3.21), systolic blood pressure (OR: 0.87, CI: 0.81, 0.94), and smoking (OR: 1.93, CI: 1.56, 2.39).

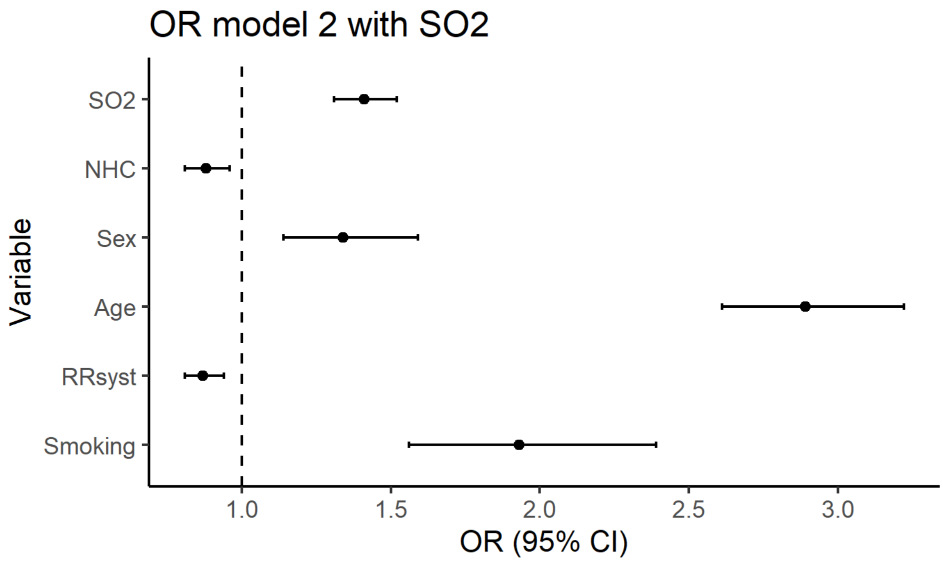


Figure 4-13 OR model with SO_2

ORs of a multivariate logistic regression model of mortality predicted by the ESC SCORE 2 risk factors and the mean exposure to SO_2
Definition of abbreviations: OR = odds ratio, model 2 = Mortality ~ variables of the ESC Score 2 + mean exposure to air pollution marker (time period of max AUC, SO_2 = sulphur dioxide, sex = sex [male], RRsyst = systolic blood pressure, CI = confidence interval, NHC = non-HDL cholesterol

After having presented the results of this dissertation, they are discussed in the next chapter.

5. Discussion

The present study showed that the power of patient intrinsic variables to predict mortality can be improved significantly by considering extrinsic air pollution markers (NO, NO_2, PM_{10}, $PM_{2.5}$, SO_2) in single air pollutant models. The most relevant exposure time differed between different air pollution markers. To my knowledge this is the first study reviewing the time-dependent association of air pollution markers and mortality applying ROC curve analysis. ROC curve analysis is a well-known and established statistical analysis tool in medical research to predict mortality and cardiovascular diseases (see e.g., Keller *et al.*, 2011; O'Connor *et al.*, 1992; Zou, O'Malley and Mauri, 2007). In this chapter, the study setting, design and results in relation to existing literature are discussed.

5.1. Study setting and design

5.1.1. Study population, patient- level characteristics

Most studies on the association between air pollution and mortality were conducted in the general population and not specifically in vulnerable subgroups with cardiovascular burden (see e.g., Beelen *et al.*, 2014; Burnett *et al.*, 2004; Carey *et al.*, 2013; Cesaroni *et al.*, 2014; Chen *et al.*, 2021; Chuang *et al.*, 2011; Crouse *et al.*, 2012; Dockery *et al.*, 1993; Filleul *et al.*, 2005; Gehring *et al.*, 2006; Heinrich *et al.*, 2013; Kim *et al.*, 2020; Laden *et al.*, 2006; Lipfert *et al.*, 2006; Liu *et al.*, 2018; Meng *et al.*, 2021; Miller *et al.*, 2007; Mirowsky *et al.*, 2015; Peters *et al.*, 2001; Pope *et al.*, 1995; Pope *et al.*, 2002; Pope *et al.*, 2004; Puett *et al.*, 2008; Raaschou-Nielsen *et al.*, 2012; Samet *et al.*, 2000; Samoli *et al.*, 2007; Stafoggia *et al.*, 2014; Zeger *et al.*, 2008). Of these, some analysed their results in regard to preexisting risk factors or disease (see e.g., Kim *et al.*, 2020). Studies showing particular risk of mortality or particular diseases for patients with pre-existing cardiovascular risk factors or conditions exist but are rare (see e.g., Goldberg *et al.*, 2013; Mann *et al.*, 2002; Pope *et al.*, 2006; Pope and Muhlestein *et al.*, 2015). Brook *et al.* (2010) thus call on more studies in this regard.

In individuals with pre-existing cardiovascular risk factors encompassing vascular disease or diabetes, exposure to air pollution bears a high risk for mortality (Brook, Brook and Rajagopalan, 2003). Diverse cardiovascular effects exacerbate comorbidities (Miller and Newby, 2020) and individuals with pre-existing cardiovascular risk factors or disease are likely to be more susceptible to air pollution exposure (Brook *et al.*, 2010). Thus, the focus on individuals with pre-existing cardiovascular burden for this study allowed for a potentially more clear-cut examination of the influence of air pollution for particularly vulnerable groups on mortality.

The study population of this dissertation is a comparatively homogenous group differing from the general German population in respect to the presence of cardiovascular diseases. In general, the lifetime prevalence of coronary artery disease (CAD) among individuals aged 40-79 in Germany is estimated to be about 4.7% (Gößwald *et al.*, 2013) to 12% (Dornquast *et al.*, 2016) while the 12 month prevalence of CAD was found to be 4.8% (Busch and Kuhnert, 2017). Within the cohort the prevalence of coronary artery disease was 5-10 times as high (51.5%). In addition, cardiovascular risk factors were highly prevalent. Thus, this study deals with a specifically vulnerable group thereby adding to the field of air pollution marker research.

5.1.2. Investigation period

Within this study the median follow-up period for mortality was approximately 10 years. This time period complies with the forecast period of the ESC SCORE 2 to predict fatal and non-fatal cardiovascular events (SCORE2 working group and ESC Cardiovascular risk collaboration, 2021). The variables included in the ESC SCORE 2 are used as traditional risk factors within this dissertation.

The procedure chosen in this dissertation is novel with regard to the chosen time period of air pollution exposure analysis. Existing studies investigate the exposure to air pollution up until a certain event (e.g. total mortality, cardiovascular mortality or respiratory mortality). In contrast, air pollution exposure in this dissertation was assessed for time periods within the three years prior to study enrolment. This was done, to identify the best time period to predict future mortality. Thus, seamless comparison with the existing literature may be restricted.

This approach additionally differed from the existing literature where of set time periods of days, weeks, months or years were assessed (see for e.g.,Stockfelt *et al.*, 2015). Within this dissertation, the exposure time period that yielded a maximum AUC was approximately within the first year for the majority of air pollution markers (CO [3 days], O_3 [348 days], PM_{10} [232 days], PM2.5 [402 days]). Nevertheless, for other air pollution markers (NO [1077 days], NO_2 [1014 days], SO_2 [1058 days]) the time period with the highest AUC was close to the maximum time looked at. Hence, there might be unidentified trends and future studies should investigate a larger time period.

5.1.3. Study location and geographic spread

There is a host of literature on the impact of air pollution on health from other parts of Germany such as Bavaria (Breitner *et al.*, 2014; Chen *et al.*, 2020; Ibald-Mulli *et al.*, 2001; Peters *et al.*, 2001; Peters *et al.*, 2004; Wolf *et al.*, 2015), North-Rhine Westphalia (Bauer *et al.*, 2010; Gehring *et al.*, 2006; Heinrich *et al.*, 2013; Hoffmann *et al.*, 2007; Hoffmann *et al.*, 2009; Katsouyanni *et al.*, 1997) , Thuringia (Hildebrandt *et al.*, 2009; Katsouyanni *et al.*, 2001; Rückerl *et al.*, 2006). To my knowledge, this is the first study dealing with the association between air pollution markers and mortality within the region of Central Hesse. Thus, this study opens up a new geographic region in the field of air pollution and cardiovascular health research.

With regard to the more technical decision of a reasonable distance between an individual's residing area and air quality monitoring stations a clearly optimal distance for explanatory power of air pollution exposure appears to be non-existent (Li *et al.*, 2019). If at all clarified, distances to air monitoring sites in epidemiological studies of air pollution and health vary between 2.3km (Filleul *et al.*, 2005), 10km (Yoon *et al.*, 2020), 16 km (Miller *et al.*, 2007), 20 km (Li *et al.*, 2019; Padula *et al.*, 2014), 30km (Gaio *et al.*, 2022; Lanzinger *et al.*, 2014; Schneider *et al.*, 2008; Schneider *et al.*, 2010; Schneider *et al.*, 2011), 8 km and 30 km (Ostro *et al.*, 2010), 40 km (Rage *et al.*, 2009), 50 km (Milojevic *et al.*, 2014; Spencer-Hwang *et al.*, 2011), up to 80 km (Schildcrout *et al.*, 2006; Schwartz *et al.*, 2008). Among these studies, some applied inverse distance weighted interpolation, while others applied sensitivity analysis for different distances and yet others used the data as they were.

In regard to the scope of this dissertation project, air pollution data were used without modification. The distance of 35 km was chosen for practical reasons. This was due to the fact that a distance of 35 km graphically allowed an assignment of the majority of the patients to an air pollution station.

5.2. Air pollution markers and adverse health effects

In this section, the ORs of each model 2, respectively (mortality predicted by the ESC SCORE 2 variables and the mean air pollution concentration for the best predictive time period) are analysed in regard to existing literature. Thus, the association of the ESC SCORE 2 variables (non-HDL-cholesterol, sex, age, systolic blood pressure, and smoking) with mortality is adjusted for air pollution exposure. As discussed earlier, a direct comparison with the existing literature is limited because of the time period of exposure analysis. While this study assesses the impact of air pollution before study enrolment to predict future mortality events, the existing literature evaluates exposure to air pollution before a certain event. Nevertheless, the trend of influence of air pollution markers is very likely to be comparable in either method.

CO

Overall, a no significant relationship between CO and mortality was found. However, the existing literature is conflicting in regard to the association of CO exposure with mortality. The findings of this dissertation are contradicting the findings of a large global time-series studies in 337 cities by Chen *et al.* (2021). They found that a 1 mg/m³ increase in the mean CO concentration of the previous day was associated with a 0.91% (95% CI: 0.32–1.50) increase in daily total mortality (Chen et al., 2021). Nevertheless, they also identified regional differences, with no or unfavourable (not significant associations) for countries such as the UK, Sweden, or Taiwan. The authors of this study attribute this to differences in air pollution levels, exposure assessment accuracies, population susceptibility, basic health status, and different sample sizes (ibid.). These results are comparable to the findings of Liu *et al.* (2018) from 272 cities in China. They observed significant increments in mortality 1.12% (95% posterior interval [PI] 0.42, 1.83) for a 1 mg/m³ increase in average CO concentrations on the present day and previous day (Liu *et al.*, 2018).

A study from 19 European cities participating in the APHEA-2 Project (Air Pollution and Health: A European Approach) indicates a significant association of total cardiovascular mortality and CO: The authors identified an association between 1 mg/m^3 increase in the 2-day average CO levels with a 1.20% [95% CI: 0.63, 1.77%] increase in total mortality and a 1.25% (95% CI: 0.30, 2.21%) increase in cardiovascular mortality (Samoli *et al.*, 2007). However, Samoli *et al.* (2007) found that CO effects varied with the region. For eastern European cities they found no CO effect on mortality (Samoli *et al.*, 2007). In their study, they assume that higher standardised mortality rates and older cohorts in these cities may be an explanation for this due to risks competing mortality (ibid.). This could also be the case for the results of this dissertation including an older cohort.

Furthermore, a meta-analysis of time series studies by Stieb, Judek and Burnett (2002) observes random effects pooled estimates of excess all-cause mortality (single-pollutant models) associated with a 1.7% (1.2, 2.2%) per 1.1 ppm change in CO concentration. Elevated levels were mostly significantly associated with a higher risk of mortality but some were also non-significantly related with mortality (ibid.). The results for the association between CO exposure and mortality within this dissertation are generally consistent with long-term air pollution exposure study in a Californian teachers cohort by Lipsett *et al.* (2011). The HR for all-cause mortality found in their study was 0.93 [95% CI: 0.84 , 1.02] for CO interquartile range increases and therefore non-significant (ibid.). Interestingly, a study by Atkinson and Analitis *et al.* (2016) observed a direct relationship between higher CO exposure and respiratory mortality, but inverse associations with cardiovascular and total mortality. However, the authors investigated the association with daily mortality and therefore their statistical measures differ substantially from those applied within this dissertation. Overall, the literature in regard to CO exposure and mortality appears to be inconsistent. In this study, no significant relationship between CO exposure and mortality was identified. The inconsistency with some studies might be explained by differences in the study design such as the age and the comorbidity characteristics of this study population.

NO and NO_2

In this dissertation, a significant association for higher NO exposure with increased mortality (OR 1.18; 95% CI: 1.1, 1.27) and NO_2 (OR 1.24; 95% CI: 1.15, 1.34) was identified. With regard to the existing literature, most studies focus on the oxides of nitrogen (NOx, i.e. NO and NO_2) or NO_2. For this reason, the discussion of existing literature for them is combined. While a short-term study on the exposure to traffic-related air pollution and daily mortality found no evidence for associations between NOx and total or cause-specific mortality (Atkinson and Analitis *et al.*, 2016), the results of this dissertation are generally consistent with the majority of the identified epidemiological literature: higher NO_2 levels are significantly associated with a higher risk of all-cause mortality Canada's 12 largest cities (Burnett *et al.*, 2004), in a large English national cohort (Carey *et al.*, 2013), seven French cities (Filleul *et al.*, 2005), North-Rhine-Westphalia (Gehring *et al.*, 2006; Heinrich *et al.*, 2013), Belgium (Collart *et al.*, 2018), Denmark (Raaschou-Nielsen *et al.*, 2012), and a 398 city study in low to high income countries by Meng *et al.* (2021).

In the before mentioned study with Californian teachers, Lipsett *et al.* (2011) found significant evidence for the direct association of higher NOx, PM_{10}, and $PM_{2.5}$ exposure with increased mortality from ischemic heart disease but not for all-cause mortality. Further, a French study that analysed the impacts of air pollution on mortality over a time span of 25 years linked increased values of NO and NO_2 to higher non-accidental mortality rates (Filleul *et al.*, 2005). Importantly, in the aforementioned meta-analysis by Stieb, Judek and Burnett (2002) random effects pooled estimates of excess all-cause mortality (single-pollutant models) were 2.8% (95% CI: 2.1, 3.5%) per 24.0 ppb change in NO_2. These results were reinforced by a systematic review of the literature by Huangfu and Atkinson (2020) who identified that an elevated exposure to NO_2 is significantly linked with higher all-cause mortality and with a random- effects summary relative risk of 1.02 (95% CI: 1.01, 1.04).

In summary, the results of this dissertation of a significant relationship between heightened levels of NO and NO_2 and increased mortality are in accordance with the major part of the identified literature. Differences to existing evidence are plausibly attributable to divergences in the study design, for instance the time period investigated.

O_3

In this dissertation, a non-significant association of O_3 exposure with mortality OR 0.99 (95% CI: 0.92, 1.07) was observed. Previous studies analysing the association between ozone and mortality have produced inconclusive results.

Some of the existing literature revels an association between elevated exposure to ozone and an increment in mortality. A review, meta-analysis and multi- city time series study by Ito, Leon and Lippmann (2005) supports that increased short-term exposure to O_3 is predictive of higher mortality. Nevertheless, the authors found inconsistent estimates for distinct cities which they contribute to unique characteristics of precisely these cities (ibid.). The majority of the cohort within this dissertation is ubicated in rural areas, hence this characteristic varies substantially from the study by Ito, Leon and Lippmann (2005).

Furthermore, a meta-analysis of 39 time-series studies in comparison to a large 95 city time-series study in urban areas within the U.S. concluded a strong evidence for the association between O_3 exposure and mortality (Bell, Dominici and Samet, 2005). Nevertheless, they also identified publication bias leading them to conclude a possible overestimation of the true relationship between O_3 exposure and mortality (ibid.). In contrast, one national English cohort study by Carey *et al.* (2013) produced an inverse significant HR of 0.86 (95% CI, 0.78, 0.94) for all-cause mortality associated with residential O_3 concentrations. In a Canadian cohort of deceased individuals > 65 years residing in Montreal between 1990 and 2003, Goldberg *et al.* (2013) could not identify associations between O_3 exposure and daily non-accidental mortality.

Moreover, a study linking data of the American Cancer Society Prevention study II with air pollution exposure found an association per 10-ppb increment ozone with increased cardiovascular and respiratory mortality but no association with all-cause mortality RR: 1.001(95% CI: 0.996,1.007) (Jerrett *et al.*, 2009). Nevertheless, no effect of ozone for cardiovascular death could be found when $PM_{2.5}$ was considered (ibid.). A cohort study in 20 of the largest US cities by Samet *et al.* (2000) also found no effect of ozone on all-cause, cardiovascular and respiratory mortality. Additionally, a recent systematic review and meta-analysis by Atkinson and Butland *et al.* (2016) and Huangfu and Atkinson (2020) found no association between long-term annual O_3 exposure and all-cause mortality. In conclusion, the available literature is ambiguous about the association between O_3 exposure and mortality, the results of this dissertation indicate no association with mortality. Outcome deviations could be explained by the characteristics of the study design such as the rural setting of the study setting in comparison to urban settings with some identified studies.

PM_{10} and $PM_{2.5}$

In this dissertation, significant associations for the increased exposure to $PM_{2.5}$ with higher mortality (OR 1.54 (95% CI: 1.40, 1.68)), and PM_{10} (OR 1.11 (95% CI: 1.03, 1.20)) and higher mortality were identified, respectively. Owing to the frequent analysis of $PM_{2.5}$ and PM_{10} as air pollution markers within the same studies, both are discussed together in view of the existing literature. The results observed within this dissertation are in accordance with previous studies: Elevated concentrations in short- and long-term exposure of $PM_{2.5}$ and PM_{10} of are consistently linked to elevated risk of all-cause, cardiovascular, and respiratory mortality.

Groundbreaking cohort studies like the Harvard 6 cities study by Dockery *et al.* (1993) and the American Cancer Society Cancer Prevention study II (ACS CPS II) by Pope *et al.* (1995) illustrated that excess all-cause mortality is associated with long-term exposure to fine particulate air pollution. Ever since, evidence for this association has accumulated. The results have been confirmed and extended by successive studies of the ACS CPS II cohort (Jerrett *et al.*, 2009; Pope *et al.*, 2002; Pope *et al.*, 2004) and the Harvard 6 cities study by Laden *et al.* (2006). The study by Laden *et al.* (2006) outlined an overall mortality increase associated with each 10 µg/ m^3 increase in average $PM_{2.5}$ exposure RR 1.16 (95% CI: 1.07, 1.26) and. In the study of a representative national English cohort, a 10 µg/ m^3 increase of PM_{10} and $PM_{2.5}$ produced an HR of 1.07 (95% CI, 0.99, 1.16) and 1.13 (95% CI, 1.00–1.27) for all-cause mortality, respectively (Carey *et al.*, 2013).

As part of the Air Pollution and Health: A European Approach 2 (APHEA 2) project an increase in daily mortality was associated with short-term exposure to PM_{10} in 29 European cities (Katsouyanni *et al.*, 1997). Additionally, several German population based cohort studies also found an association the exposure to PM_{10} and between all-cause mortality (Gehring *et al.*, 2006; Heinrich *et al.*, 2013). Other US cohort studies also observed significant associations between the exposure of $PM_{2.5}$ and PM_{10} and mortality (Lipfert *et al.*, 2006; Lipsett *et al.*, 2011; Puett *et al.*, 2008; Samet *et al.*, 2000). Moreover, a study by Pope and Muhlestein *et al.* (2015) evaluated the impact of a 10 µg/ m^3 $PM_{2.5}$ increase on the same day. They linked same-day $PM_{2.5}$ increase with elevated significant ORs for ACS , non-ST-/ ST-segment elevation myocardial infarction, unstable angina, and non–ST-segment elevation ACS in patients with pre-existing cardiovascular conditions.

According to a meta-analysis by Stieb, Judek and Burnett (2002), random effects pooled estimates of excess all-cause mortality (single-pollutant models) were 2% (95% CI: 1.5% , 2.4%) per 31.3 µg/ m^3 PM_{10}. A systematic review by Chen, Goldberg and Villeneuve (2008) showed a 6 % increase in non-accidental mortality for the long-term exposure to $PM_{2.5}$ per 10 µg/ m^3 increase. A later conducted systematic review and meta-analysis by Atkinson *et al.* (2014) also found significant association for the short-term exposure to $PM_{2.5}$. In their study, a rise of 10 µg/ m^3 in the exposure to $PM_{2.5}$ was associated with a 1.04% (95% CI 0.52%, 1.56%) increase in mortality risk (ibid.). Nevertheless, Atkinson *et al.* (2014) pointed out that heterogeneity between studies and small study bias should be borne in mind. These results were confirmed by a more recent systematic review and meta-analysis by Chen and Hoek (2020). In their analysis of 107 studies the Risk Ratio of the exposure to $PM_{2.5}$ and PM_{10} with non-accidental all-cause mortality was 1.08 (95% CI: 1.06, 1.09) and 1.04 (95% CI: 1.03, 1.06) per 10 µg/ m^3 increase, respectively (ibid.).

Generally, the results of this dissertation are in accordance with the majority of the extensive existing evidence showing that increased exposure to $PM_{2.5}$ and PM_{10} is significantly linked to increased all- cause mortality.

SO_2

In this study elevated SO_2 values were significantly predictive of higher mortality yielding an OR of 1.41 (95% CI: 1.31, 1.52) in the single-air pollutant model. There is inconsistent evidence for the association of the exposure to SO_2 and all-cause mortality. A French population-based cohort study of 14284 individuals residing in 7 French cities found a non-significant association between exposure to SO_2 and non-accidental mortality RR 1.01 (95% CI: 0.99 – 1.03) per 10 µg/ m^3 SO_2 increase (Filleul *et al.*, 2005). Interestingly, a time-series study analysing data from 12 European cities within the APHEA project linked daily mortality to a 50 µg/ m^3 SO_2 increase of 3% (CI 95%: 2%, 4%) in western European cities and of 0.8% (− 0.1% to 2.4%) in central eastern European cities (Katsouyanni *et al.*, 1997). According to Katsouyanni *et al.* (1997) this e.g. may be due to differences in populations and in air pollution sources. The findings of this study are in line with the representative English national cohort study by Carey *et al.* (2013) which linked excess all-cause mortality to a 10 µg/ m^3 increment, adjusted for area income (HR 1.20; 95% CI: 1.12,1.28). In the long-term population based study of a Californian teachers' cohort, all-cause mortality was significantly associated with SO_2

(HR, 1.11; 95% CI, 1.00, 1.23) (Lipsett *et al.*, 2011). Nevertheless, the analysis of Lipsett *et al.* (2011) argues that these findings rested on few cases. Goldberg *et al.* (2013) observed an increase of daily mortality in individuals with pre-existing cardiovascular diseases.

Recently, a multicounty analysis of the short-term association between SO_2 exposure and mortality in 399 cities observed excess mortality (RR 1.0045, 95% CI: 1.0019, 1.0070) per 10 µg/ m^3 SO_2 increase (O'Brien *et al.*, 2023). In addition, the previously introduced meta-analysis by Stieb, Judek and Burnett (2002) showed random effects pooled estimates for the association between SO_2 exposure with excess all-cause mortality (single-pollutant models) of 0.9% (95% CI: 0.7, 1.2%) per 9.4 ppb change.

Considering all this, the findings of this dissertation indicate a significant relationship between increased exposure to SO_2 and heightened mortality rates while the existing literature is inconsistent. As with the other air pollution markers, this contradictoriness is plausibly attributable to differences in the study setting and the study population characteristics such as the comorbidities of the included populations.

5.3. Considering the air pollution exposure in combination with ESC SCORE 2 variables improves the prediction of mortality

The ESC SCORE 2 variables were significantly associated with mortality except non-HDL-cholesterol in the multivariate model with CO. These associations remained significant after the adjustment for air pollution markers. As discussed in the previous subsection, increased exposure to NO, NO_2, PM_{10}, $PM_{2.5}$, and SO_2 was significantly associated with higher mortality in the multivariate models (model 2: mortality predicted by the ESC SCORE 2 variables and the mean air pollution concentration for the best predictive time period).

This dissertation is distinct to the existing literature because it is one of the few studies assessing the association between air pollution exposure and mortality in addition to patient-level clinical data such as non-HDL cholesterol and systolic blood pressure. Additionally, to my knowledge, this is the first study analysing the effects of air pollution exposure in connection to the ESC SCORE 2 and its variables. This allowed to analyse the improvement of the predictive power of the ESC SCORE 2 and its variables when the mean exposure to an air pollution marker was added. Some of the existing literature on the association between air pollution exposure and mortality controls for covariates including sex, smoking, body mass index, income, education, current symptoms, physical activity, dietary consumption, previous disease, medication, occupational exposure, and alcohol consumption. In general, the results of this dissertation are consistent with the existing literature in regard to these patient-level covariates: A cohort study by Jerrett *et al.* (2009) found independent associations between air pollution exposure and mortality when adjusting for 20 individual person risk variables and 7 ecological variables.

After adjusting for covariates, two German studies observed of air pollution on mortality that remained increased but smaller (Gehring *et al.*, 2006) or changed only marginally and stayed significant after adjusting for such covariates (Heinrich *et al.*, 2013). This was also the case for study of the Korean National Health Insurance Service (NHIS)-based National Sample Cohort (NSC) database (Kim *et al.*, 2020). The authors analysed their data using a multivariate model that adjusted for patient-level risk factors and found reduced but significantly increased associations (ibid.).

In addition, other population based cohort studies for instance by Carey *et al.* (2013), Laden *et al.* (2006), Miller *et al.* (2007), Pope *et al.* (2002), Puett *et al.* (2008) also observed robust associations between exposure to air pollution and all-cause mortality with no significant changes after adjustment for covariates. Overall, despite substantial differences in the study design, the findings of significant associations between NO, NO_2, PM_{10}, $PM_{2.5}$, and SO_2 and all-cause mortality in multivariate models within this dissertation are consistent with the majority of the identified literature in regard to covariate analysis.

5.4. No benefit for the mortality model precision by considering the socioeconomic status

In their review of the literature Miller and Newby (2020) identified considerable socioeconomic disparities in exposure to air pollution. Poor areas are often those areas where exposure to air pollution is highest (ibid.). Thus, it was decided to analyse the socioeconomic status in regard to potential predictive benefits and possible adjustment of the predictive power of the air pollutants within this dissertation. In their study on different indicators of the SES, Darin-Mattsson, Fors and Kåreholt (2017) found no relevant differences of using income instead of more complex SES measures. Importantly, they also observed that while income as a single indicator could be used as a socioeconomic proxy variable, it was less useful to assess distinct facets of health inequities.

Within this dissertation, purchasing power was the closest variable to income available, which was only available on postcode level and many other studies also used area-level SES data (see e.g., Carey *et al.*, 2013; Eftim *et al.*, 2008; Samet *et al.*, 2000). Hence, purchasing power on a postcode level was used as a proxy for the SES.

In this dissertation, there was no significant benefit of adding purchasing power as an explanatory variable to a model of ESC SCORE 2 variables and air pollution in the prediction of mortality. Thus, it appears that purchasing power does not increase the predictive power of air pollution markers. From the graphical analysis (Figure 4-1), it is possible that the geographical spread with marginally varying purchasing power postal code areas was too homogenous.

A Canadian study by Jerrett *et al.* (2004) observed an attenuation of the effect of air pollution on mortality when socioeconomic characteristics were included. Further, Carey *et al.* (2013) in their population based cohort study also used neighbourhood indicators (area income) of socioeconomic deprivation. However, in their model, SES adjustment attenuated the association of air pollution markers with mortality (Carey *et al.*, 2013). Hence, income is possibly more accurate in capturing the SES than purchasing power. Observations of this dissertation are comparable to the findings of Samet *et al.* (2000) who also did not use individual indicators for the SES but county-level data. They found no evidence of SES adjustment in the association between PM_{10} and mortality (ibid.). They argue that the socioeconomic indicators by the U.S census might be too imprecise and might not adequately reflect medical conditions and poor health due to intercounty differences (ibid.). This could also be the case within this dissertation.

A global review by Hajat, Hsia and O'Neill (2015) indicates that findings from European studies in regard to higher concentrations of air pollutants are mixed. Interestingly, some studies they identified also observed higher exposure to air pollution in areas of higher SES. However, they discussed that in higher SES areas individuals may be preferably able to flee high exposure conditions through costly mitigation tactics such as private transportation (ibid.). In addition, they argued that there are considerable limitations in using only one indicator as a SES proxy given that one indicator may be limited in capturing SES complexity (ibid.). Thus, an indicator capturing more of the complexity of social disparities may possibly lead to different results. This might need rigorous future research.

To summarize, the non-significant effects of the socioeconomic status on mortality within this dissertation might be attributable to the insufficiency of purchasing power to capture the socioeconomic status.

5.5. Strengths and limitations

This dissertation has many strengths. Patient-level clinical characteristics and medical history were raised within a certified centre of the German centre for cardiovascular research. Data collection occurred as part of an ongoing prospective patient cohort through standardised data collection by medical personnel.

Furthermore, to my knowledge this is the first study assessing the association between air pollution exposure and mortality in central Hesse as well as in connection to the variables of the ESC SCORE 2. The data allowed to analyse the effects of different air pollution exposure time frames. Hence, by applying ROC analysis over time, time frames of maximum exposure interest could be identified. The methodology applied within this dissertation is innovative - to my knowledge, this is the first study that applies ROC analysis to identify the most important time period of air pollution exposure for its association with mortality. It therefore commences a new area within the field of air pollution and health research.

Nevertheless, the results of this study should be dealt with caution due to several limitations. There is the possibility of exposure misclassification bias: The possible distance of an air monitoring station was set to 35km because this distance was feasible, allowed the assignment of the majority of the patients, and a clear distance standard without loss of explanatory power was absent. Furthermore, personal exposure may vary substantially even within the same neighbourhood (Brook *et al.*, 2010; Sarnat *et al.*, 2007). This variation arises based on differences in personal exposure (working conditions, geographic mobility, commuting vehicle, information about time spent indoors, distance to major road, etc.). However, according to Brook *et al.* (2010) the effect of more precise measurement of personal exposure on misclassification is yet to be clarified.

Furthermore, many studies have considered meteorological factors, seasonality, and multipollutant models in their studies (see e.g., Atkinson and Analitis *et al.*, 2016; Collart *et al.*, 2018; Faustini, Rapp and Forastiere, 2014; Huang *et al.*, 2017; Huangfu and Atkinson, 2020; Klompmaker *et al.*, 2021; Samoli *et al.*, 2001; Samoli *et al.*, 2007; Stieb, Judek and Burnett, 2002). The consideration of these factors was out of the scope of this dissertation but appears to be relevant and should be further assessed in the future.

In addition, smoking was generalised into yes and no without taking account of the pack years. Moreover, former smokers were classified as non-smokers because the ESC SCORE 2 uses the variable current smoking habits and no information was available about household second hand smoking. Mortality information was retrieved from the national post office in September 2022 – for this reason the exact date and the exact cause of death are not available. This may be considered in future studies and follow up studies.

The aim of this dissertation is to improve the prediction of mortality. However, the ESC SCORE 2 was developed to predict fatal and non-fatal cardiovascular events. It was yet decided to use the ESC SCORE 2 and its variables because this score has a class I recommendation within the “ESC guideline on cardiovascular diseases prevention in clinical practice” by Visseren *et al.* (2021) and is hence assumed to be a well-established score in clinical practice and therefore including the risk variables of interest.

The risk of the burden of cardiovascular diseases associated with air pollution may be underestimated when only fatal events are analysed. For this reason, future studies should include both endpoints. While all analysis were conducted rigorously, caution should be exercised in interpreting the results of this dissertation. Further studies are needed to confirm the possible beneficial predictive power of considering air pollution exposure in mortality risk prediction.

6. Conclusion

The aim of this dissertation was to explore the potential prognostic benefit of the extrinsic risk factor air pollution in comparison to the well-established ESC SCORE 2 risk factors in a population with pre-existing cardiovascular risk factors in Hesse. To that end, ROC and logistic regression analyses were conducted.

The methods applied within this study are novel in the field of air pollution and health research. While the existing literature evaluates exposure to air pollution before a certain event, this dissertation assessed the impact of air pollution before study enrolment to predict future mortality events. For each air pollution marker, the mean exposure time periods with the highest predictive power within the range of three years prior to study enrolment were identified. For the majority of air pollution markers, this exposure time period was approximately within the first year (CO [3 days], O_3 [348 days], PM_{10} [232 days], PM2.5 [402 days]). Nevertheless, for other air pollution markers (NO [1077 days], NO_2 [1014 days], SO_2 [1058 days]) the time period was close to earliest time period looked at. Future studies, should therefore cover larger time periods of exposure analysis. These mean air pollution exposure concentrations for the time periods of maximum predictive power were subsequently each analysed in single air pollution marker and multivariate logistic regression models, respectively.

The overall findings of this dissertation are remarkable: it was observed that the power of ESC SCORE 2 variables to predict mortality can be improved significantly by including air pollution markers (NO, NO_2, PM_{10}, $PM_{2.5}$, SO_2) in single air pollutant models. This was not the case for CO and O_3. In this research, purchasing power as a proxy of the SES did not attenuate the predictive power of air pollution markers. An indicator accounting for more of the complexity of social disparities may possibly lead to another outcome.

To the best of my knowledge, this is the first study assessing the association between air pollution and mortality in Hesse, Germany and one of the few studies specifically looking at a population with preexisting cardiovascular risk factors.

In brief, in comparison to traditionally used risk variables like those used within the ESC SCORE 2, mortality was significantly and positively associated with time dependent exposure to air pollution markers such as NO, NO_2, PM_{10}, $PM_{2.5}$, SO_2. The results of this dissertation indicate that including air pollution exposure in a score assessing the mortality risk of individuals with cardiovascular burden is likely to enhance mortality risk estimation.

Further evaluation of the prognostic impact of a potential extension of scores like the ESC SCORE 2 with a variable capturing the exposure to air pollution markers should be subject to future studies.

Summary

Background and Aim: Non-communicable diseases, e.g. cardiovascular diseases, are among the leading causes of mortality worldwide. Well-established risk scores, such as the ESC SCORE 2 (European Society of Cardiology Systematic COronary Risk Evaluation) are used to predict the onset of cardiovascular disease and mortality. However, they consider patient intrinsic risk factors, e.g. smoking but not extrinsic risk factors like air pollution. Yet, there is extensive evidence for an unfavourable association between air pollution exposure and cardiovascular and other mortality. This dissertation aims to explore the time-dependent association between the extrinsic risk factor air pollution and mortality in a population with pre-existing cardiovascular risk factors in Hesse, Germany. Moreover, it explores the utility of considering air pollution and socio-economic status in mortality risk prediction, compared to existing risk scores.

Methods and results: Between 2010 and 2019, patients (N = 4610, 32% female, median age 69 years) scheduled for coronary angiography were enrolled in a prospective registry cohort at a certified German centre for cardiovascular research. Mortality was the outcome variable (1122 registered deaths). Air pollution data were retrieved from the Hessian State Agency for Nature Conservation, Environment and Geology (HLNUG) and assigned to patients according to postcode information. Air pollution markers of interest were: carbon monoxide (CO), nitrogen monoxide (NO), nitrogen dioxide (NO_2), ozone (O_3), particulate matter (PM_{10} and $PM_{2.5}$ with an aerodynamic diameter of <10µm and <2.5µm, respectively) and sulphur dioxide (SO_2). Receiver Operating Characteristic (ROC) analysis was conducted to identify the time period with the highest prognostic importance of air pollution exposure for its association with mortality within the range of 3 years prior to study enrolment. Exposure time periods with the highest prognostic ability varied for respective air pollution markers (CO [3 days], O_3 [348 days], PM_{10} [232 days], $PM_{2.5}$ [402 days], NO [1077 days], NO_2 [1014 days], SO_2 [1058 days]). Mean air pollution exposure concentrations for time periods of maximum predictive power were then each analysed in multivariate logistic regression models, with single air pollution markers. In these single air pollution marker models, adding NO, NO_2, PM_{10}, $PM_{2.5}$, SO_2 but not CO and O_3 to a model including the ESC SCORE 2 variables, could significantly improve a models' mortality prediction power. There was no significant benefit of adding purchasing power as an explanatory variable.

Conclusion: The findings of this dissertation indicate that considering air pollution exposure in mortality risk prediction for individuals with cardiovascular burden is likely to enhance mortality risk estimation and should be subject to further research.

Zusammenfassung

Hintergrund und Absicht: Nicht-übertragbare Krankheiten, wie koronare Herzkrankheiten, zählen weltweit zu den führenden Todesursachen. Etablierte Risiko-Scores wie der ESC SCORE 2 (Europäische Gesellschaft für Kardiologie Systematische koronare Risiko Einschätzung) beinhalten intrinsische Risikofaktoren z.B. das Rauchverhalten, extrinsische Risikofaktoren wie Luftverschmutzung bleiben unberücksichtigt. Nichtsdestotrotz existiert eine umfassende Datengrundlage über den ungünstigen Zusammenhang zwischen Luftverschmutzung und kardiovaskulärer und anderer Mortalität. Ziel der Dissertation ist die Untersuchung des zeitabhängigen Zusammen-hang zwischen dem extrinsischen Risikofaktor Luftverschmutzung und Mortalität in einer hessischen Kohorte mit bestehenden kardiovaskulären Risikofaktoren. Zudem sollen prognostische Vorteile der Berücksichtigung von Luftverschmutzung gegenüber etablierter Risikofaktoren und des sozioökonomischen Status exploriert werden.

Methoden und Ergebnisse: Zwischen 2010 und 2019 wurden Patienten mit durchgeführter Koronarangiografie (N=4610, 32% weiblich, medianes Alter 69 Jahre), in ein prospektives Kohortenregister in einem zertifizierten deutschen Zentrum für kardiovaskuläre Forschung eingetragen. Die Zielvariable war Mortalität (1122 registrierte Sterbefälle). Luftverschmutzungsdaten wurden durch das Hessisches Landesamt für Naturschutz, Umwelt und Geologie (HLNUG) bereitgestellt und Patienten gemäß der Postleitzahl ihres Wohnorts zugeordnet. Berücksichtigt wurden: Kohlenstoffmonoxid (CO), Stickstoffmonoxid (NO), Stickstoffdioxid (NO_2), Ozon (O_3), Feinstaub < 10 µm (PM_{10}), Feinstaub <2,5 µm ($PM_{2,5}$) und Schwefeldioxid (SO_2). Über ROC (Receiver-Operating Characteristic)-Kurvenanalysen wurde die Zeitspanne mit dem höchsten prognostischen Wert für den Zusammenhang zwischen der Exposition gegenüber Luftverschmutzung und Mortalität innerhalb eines Zeitraums von drei Jahren vor Studieneinschluss eruiert. Die Zeiträume mit der höchsten prognostischen Aussagekraft waren unterschiedlich je Luftverschmutzungsmarker (CO [3 Tage], O_3 [348 Tage], PM_{10} [232 Tage], PM2.5 [402 Tage], NO [1077 Tage], NO_2 [1014 Tage] und SO_2 [1058 Tage]). In multivariaten logistischen Regressionsanalysen wurden mittlere Expositionswerte der Luftverschmutzungsmarker über diese Zeiträume untersucht. Durch das Ergänzen des Modells um einzelne Luftverschmutzungsmarker (NO, NO_2, PM_{10}, $PM_{2.5}$ oder SO_2) konnte die prognostische Fähigkeit des jeweiligen Modells signifikant verbessert werden. CO und O_3 hatten keine prognostische Relevanz. Auch die Berücksichtigung des sozioökonomischen Status hatte keinen signifikanten Einfluss auf die Risikoprädiktion.

Fazit: Die Ergebnisse dieser Dissertation weisen darauf hin, dass die Berücksichtigung der Exposition gegenüber Luftschadstoffen die Mortalitätsrisikoprädiktion verbessern könnte und sollte deswegen Gegenstand weiterer Forschung sein.

List of figures, tables and abbreviations

Table of Figures

Table of tables

Table of equations

Abbreviations

ACS	Acute coronary syndrome
AUC	Area under the curve
BMI	Body mass index
CABG	Coronary artery bypass grafting
CAD	Coronary artery disease
CI	Confidence intervall
CO	Carbon monoxide

DDS	Digital Data Services
EPSG	European Petroleum Survey Group Geodesy
ESC	European Society of cardiology
FN	False negative
FP	False positive
FPR	False positive rate
GmbH	Gesellschaft mit beschränkter Haftung (company with limited liability)
H_0	Null hypothesis
H_1	Alternative hypothesis
HLNUG	Hessisches Landesamt für Naturschutz, Umwelt und Geologie (Hessian State Agency for Nature Conservation, Environment and Geology)
HR	Hazard ratio
IQR	Interquartile range
MAR	missing at random
MCAR	missing completely at random
MNAR	missing not at random
NCDs	Non-communicable diseases
NHIS	National Health Insurance Service
NO	Nitrogen oxide
NO_2	Nitrogen dioxide
NSC	National Sample Cohort
O_3	Ozone
OR	Odds ratio
PCI	Percutaneous coronary intervention
PM_{10}	Particulate matter with an aerodynamic diameter of <10µm
PM_{25}	Particulate matter with an aerodynamic diameter of <2.5µm
ROC	Receiver operator characteristic
RR	Rate ratio
SD	Standard deviation
SES	Socioeconomic status
SCORE	Systematic COronary Risk Evaluation
SO_2	Sulphur dioxide
TN	True negative
TP	True positive
TPR	True positive rate
WHO	World Health Organisation

Reference list

Altman, D.G. and Bland, J.M. (1994) 'Diagnostic tests 3: receiver operating characteristic plots', *BMJ (Clinical Research Ed.)*, 309(6948), p. 188. doi: 10.1136/bmj.309.6948.188

American Psychological Association 'Socioeconomic status'. Available at: https://www.apa.org/topics/socioeconomic-status (Accessed: 20 September 2023).

Andersson, C. *et al.* (2021) 'Framingham Heart Study: JACC Focus Seminar, 1/8', *Journal of the American College of Cardiology*, 77(21), pp. 2680–2692. doi: 10.1016/j.jacc.2021.01.059

Atkinson, R.W. *et al.* (2014) 'Epidemiological time series studies of PM2.5 and daily mortality and hospital admissions: a systematic review and meta-analysis', *Thorax*, 69(7), pp. 660–665. doi: 10.1136/thoraxjnl-2013-204492

Atkinson, R.W. *et al.* (2016) 'Long-term exposure to ambient ozone and mortality: a quantitative systematic review and meta-analysis of evidence from cohort studies', *BMJ Open*, 6(2), e009493. doi: 10.1136/bmjopen-2015-009493

Atkinson, R.W. *et al.* (2016) 'Short-term exposure to traffic-related air pollution and daily mortality in London, UK', *Journal of Exposure Science & Environmental Epidemiology*, 26(2), pp. 125–132. doi: 10.1038/jes.2015.65

Auguie, B. (2017) *_gridExtra: Miscellaneous Functions for "Grid" Graphics_.: R package version 2.3*. Available at: https://cran.r-project.org/package=gridExtra.

Baker, E.H. (2014) 'Socioeconomic Status, Definition', in Cockerham, W.C., Dingwall, R. and Quah, S.R. (eds.) *The Wiley Blackwell encyclopedia of health, illness, behavior, and society.* Chichester, West Sussex: Wiley Blackwell, pp. 2210–2214.

Bartlett, J.W., Harel, O. and Carpenter, J.R. (2015) 'Asymptotically Unbiased Estimation of Exposure Odds Ratios in Complete Records Logistic Regression', *American Journal of Epidemiology*, 182(8), pp. 730–736. doi: 10.1093/aje/kwv114

Bauer, M. *et al.* (2010) 'Urban particulate matter air pollution is associated with subclinical atherosclerosis: results from the HNR (Heinz Nixdorf Recall) study', *Journal of the American College of Cardiology*, 56(22), pp. 1803–1808. doi: 10.1016/j.jacc.2010.04.065

Beelen, R. *et al.* (2014) 'Long-term exposure to air pollution and cardiovascular mortality: an analysis of 22 European cohorts', *Epidemiology (Cambridge, Mass.)*, 25(3), pp. 368–378. doi: 10.1097/EDE.0000000000000076

Bell, M.L., Dominici, F. and Samet, J.M. (2005) 'A meta-analysis of time-series studies of ozone and mortality with comparison to the national morbidity, mortality, and air pollution study', *Epidemiology (Cambridge, Mass.)*, 16(4), pp. 436–445. doi: 10.1097/01.ede.0000165817.40152.85

Breitner, S. *et al.* (2014) 'Short-term effects of air temperature on cause-specific cardiovascular mortality in Bavaria, Germany', *Heart (British Cardiac Society)*, 100(16), pp. 1272–1280. doi: 10.1136/heartjnl-2014-305578

Brook, R…, Brook, J.R. and Rajagopalan, S. (2003) 'Air Pollution: The "Heart" of the Problem', *Current Science Inc*, 5, pp. 32–39.

Brook, R.D. *et al.* (2004) 'Air pollution and cardiovascular disease: a statement for healthcare professionals from the Expert Panel on Population and Prevention Science of the American Heart Association', *Circulation*, 109(21), pp. 2655–2671. doi: 10.1161/01.CIR.0000128587.30041.C8

Brook, R.D. *et al.* (2010) 'Particulate matter air pollution and cardiovascular disease: An update to the scientific statement from the American Heart Association', *Circulation*, 121(21), pp. 2331–2378. doi: 10.1161/CIR.0b013e3181dbece1

Brook, R.D., Newby, D.E. and Rajagopalan, S. (2017) 'Air Pollution and Cardiometabolic Disease: An Update and Call for Clinical Trials', *American Journal of Hypertension*, 31(1), pp. 1–10. doi: 10.1093/ajh/hpx109

Brook, R.D. and Rajagopalan, S. (2010) 'Particulate matter air pollution and atherosclerosis', *Current Atherosclerosis Reports*, 12(5), pp. 291–300. doi: 10.1007/s11883-010-0122-7

Bundesamt für Kartographie und Geodäsie (2019) *WMS NUTS-Gebiete 1:250 000*. Available at: https://gdz.bkg.bund.de/index.php/default/open-data/wms-nuts-1-250-000-stand-31-12-wms-nuts250-1231.html (Accessed: 2019).

Burnett, R.T. *et al.* (2004) 'Associations between short-term changes in nitrogen dioxide and mortality in Canadian cities', *Archives of Environmental Health*, 59(5), pp. 228–236. doi: 10.3200/AEOH.59.5.228-236

Busch, M.A. and Kuhnert, R. (2017) '12-Month prevalence of coronary heart disease in Germany', *Journal of Health Monitoring*, 2(1), pp. 58–63. doi: 10.17886/RKI-GBE-2017-018

Carey, I.M. *et al.* (2013) 'Mortality associations with long-term exposure to outdoor air pollution in a national English cohort', *American Journal of Respiratory and Critical Care Medicine*, 187(11), pp. 1226–1233. doi: 10.1164/rccm.201210-1758OC

Cesaroni, G. *et al.* (2014) 'Long term exposure to ambient air pollution and incidence of acute coronary events: prospective cohort study and meta-analysis in 11 European cohorts from the ESCAPE Project', *BMJ (Clinical Research Ed.)*, 348, f7412. doi: 10.1136/bmj.f7412

Chatterjee, S. and Simonoff, J.S. (2013) *Handbook of regression analysis*. (Wiley handbooks in applied statistics). Hoboken, New Jersey: Wiley.

Chen, H., Goldberg, M.S. and Villeneuve, P.J. (2008) 'A systematic review of the relation between long-term exposure to ambient air pollution and chronic diseases', *Reviews on Environmental Health*, 23(4), pp. 243–297. doi: 10.1515/reveh.2008.23.4.243

Chen, J. and Hoek, G. (2020) 'Long-term exposure to PM and all-cause and cause-specific mortality: A systematic review and meta-analysis', *Environment International*, 143, p. 105974. doi: 10.1016/j.envint.2020.105974

Chen, K. *et al.* (2020) 'Hourly Exposure to Ultrafine Particle Metrics and the Onset of Myocardial Infarction in Augsburg, Germany', *Environmental Health Perspectives*, 128(1), p. 17003. doi: 10.1289/EHP5478

Chen, K. *et al.* (2021) 'Ambient carbon monoxide and daily mortality: a global time-series study in 337 cities', *The Lancet Planetary Health*, 5(4), e191-e199. doi: 10.1016/S2542-5196(21)00026-7

Chuang, K.-J. *et al.* (2011) 'Long-term air pollution exposure and risk factors for cardiovascular diseases among the elderly in Taiwan', *Occupational and Environmental Medicine*, 68(1), pp. 64–68. doi: 10.1136/oem.2009.052704

Collart, P. *et al.* (2018) ‘Short-term effects of nitrogen dioxide on hospital admissions for cardiovascular disease in Wallonia, Belgium’, *International Journal of Cardiology*, 255, pp. 231–236. doi: 10.1016/j.ijcard.2017.12.058

Conroy, R.M. *et al.* (2003) ‘Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project’, *European Heart Journal*, 24(11), pp. 987–1003. doi: 10.1016/s0195-668x(03)00114-3

Crouse, D.L. *et al.* (2012) ‘Risk of nonaccidental and cardiovascular mortality in relation to long-term exposure to low concentrations of fine particulate matter: a Canadian national-level cohort study’, *Environmental Health Perspectives*, 120(5), pp. 708–714. doi: 10.1289/ehp.1104049

Darin-Mattsson, A., Fors, S. and Kåreholt, I. (2017) ‘Different indicators of socioeconomic status and their relative importance as determinants of health in old age’, *International Journal for Equity in Health*, 16(1), p. 173. doi: 10.1186/s12939-017-0670-3

Dawber, T.R., Moore, F.E. and Mann, G.V. (1957) ‘Coronary heart disease in the Framingham study’, *American Journal of Public Health and the Nation's Health*, 47(4 Pt 2), pp. 4–24. doi: 10.2105/ajph.47.4_pt_2.4

Diouf, K. *et al.* (2019) ‘Anti-citrullinated protein antibodies are not associated with extent of disease or prognosis in patients with coronary artery disease’, *Clinical Chemistry and Laboratory Medicine*, 57(7), e159-e161. doi: 10.1515/cclm-2018-1097

Dockery, D.W. *et al.* (1993) ‘An association between air pollution and mortality in six U.S. cities’, *The New England Journal of Medicine*, 329(24), pp. 1753–1759. doi: 10.1056/NEJM199312093292401

Dong, Y. and Peng, C.-Y.J. (2013) ‘Principled missing data methods for researchers’, *SpringerPlus*, 2(1), p. 222. doi: 10.1186/2193-1801-2-222

Dornquast, C. *et al.* (2016) ‘Regional Differences in the Prevalence of Cardiovascular Disease’, *Deutsches Arzteblatt International*, 113(42), pp. 704–711. doi: 10.3238/arztebl.2016.0704

Eftim, S.E. *et al.* (2008) ‘Fine particulate matter and mortality: a comparison of the six cities and American Cancer Society cohorts with a medicare cohort’, *Epidemiology (Cambridge, Mass.)*, 19(2), pp. 209–216. doi: 10.1097/EDE.0b013e3181632c09

Elsner, L.K. *et al.* (2020) ‘Prognostic performance of the ESC SCORE and its German recalibrated versions in primary and secondary prevention’, *European Journal of Preventive Cardiology*, 27(19), pp. 2166–2169. doi: 10.1177/2047487319868034

Evans, G.W. and Kantrowitz, E. (2002) ‘Socioeconomic status and health: the potential role of environmental risk exposure’, *Annual Review of Public Health*, 23, pp. 303–331. doi: 10.1146/annurev.publhealth.23.112001.112349

Faustini, A., Rapp, R. and Forastiere, F. (2014) ‘Nitrogen dioxide and mortality: review and meta-analysis of long-term studies’, *European Respiratory Journal*, 44(3), pp. 744–753. doi: 10.1183/09031936.00114713

Fawcett, T. (2006) ‘An introduction to ROC analysis’, *Pattern Recognition Letters*, 27(8), pp. 861–874. doi: 10.1016/j.patrec.2005.10.010

Filleul, L. *et al.* (2005) ‘Twenty five year mortality and air pollution: results from the French PAARC survey’, *Occupational and Environmental Medicine*, 62(7), pp. 453–460. doi: 10.1136/oem.2004.014746

Franco, P. *et al.* (2020) ‘Air Pollution and Emergency Hospital Admissions—Evidences from Lisbon Metropolitan Area, Portugal’, *Applied Sciences*, 10(22), p. 7997. doi: 10.3390/app10227997

Gaio, V. *et al.* (2022) ‘PM10 exposure interacts with abdominal obesity to increase blood triglycerides: a cross-sectional linkage study’, *European Journal of Public Health*, 32(2), pp. 281–288. doi: 10.1093/eurpub/ckab190

García-Portugués, E. (2023) *Notes for Predictive Modeling*. 5th edn. Available at: https://bookdown.org/egarpor/PM-UC3M/.

Gehring, U. *et al.* (2006) ‘Long-Term Exposure to Ambient Air Pollution and Cardiopulmonary Mortality in Women’, *Epidemiology*, 17(5), pp. 545–551.

Goldberg, M.S. *et al.* (2013) ‘Associations between ambient air pollution and daily mortality among elderly persons in Montreal, Quebec’, *The Science of the Total Environment*, 463-464, pp. 931–942. doi: 10.1016/j.scitotenv.2013.06.095

Gómez-Carracedo, M.P. *et al.* (2014) ‘A practical comparison of single and multiple imputation methods to handle complex missing data in air quality datasets’, *Chemometrics and Intelligent Laboratory Systems*, 134, pp. 23–33. doi: 10.1016/j.chemolab.2014.02.007

Gößwald, A. *et al.* (2013) ‘Prävalenz von Herzinfarkt und koronarer Herzkrankheit bei Erwachsenen im Alter von 40 bis 79 Jahren in Deutschland: Ergebnisse der Studie zur Gesundheit Erwachsener in Deutschland (DEGS1)’, *Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz*, 56(5-6), pp. 650–655. doi: 10.1007/s00103-013-1666-9

Graham, J.W. (2009) ‘Missing data analysis: making it work in the real world’, *Annual Review of Psychology*, 60, pp. 549–576. doi: 10.1146/annurev.psych.58.110405.085530

Grolemund, G. and Wickham, H. (2011) ‘Dates and Times Made Easy with lubridate’, *Journal of Statistical Software*, 40(3), pp. 1–25. doi: 10.18637/jss.v040.i03

Grün, D. (2023) *estimateCVRisk,* 4 December. Available at: https://github.com/DGruen89/estimateCVRisk (Accessed: 4 December 2023).

Hajat, A., Hsia, C. and O'Neill, M.S. (2015) ‘Socioeconomic Disparities and Air Pollution Exposure: a Global Review’, *Current Environmental Health Reports*, 2(4), pp. 440–450. doi: 10.1007/s40572-015-0069-5

Hajian-Tilaki, K. (2013) ‘Receiver Operating Characteristic (ROC) Curve Analysis for Medical Diagnostic Test Evaluation’, *Caspian Journal of Internal Medicine*, 4(2), pp. 627–635. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3755824/.

Harrell, F.E. (2022) *Hmisc: Harrell Miscellaneous: R package version 4.7-0*. Available at: https://cran.r-project.org/package=Hmisc (Accessed: 9 May 2023).

Harris, P.A. *et al.* (2009) ‘Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support’, *Journal of Biomedical Informatics*, 42(2), pp. 377–381. doi: 10.1016/j.jbi.2008.08.010

Harris, P.A. *et al.* (2019) ‘The REDCap consortium: Building an international community of software platform partners’, *Journal of Biomedical Informatics*, 95, p. 103208. doi: 10.1016/j.jbi.2019.103208

Harrison, E., Drake, T. and Ots, R. (2023) *finalfit: Quickly Create Elegant Regression Results Tables and Plots when Modelling: R package version 1.0.6*. Available at: https://github.com/ewenharrison/finalfit.

Heinrich, J. *et al.* (2013) 'Long-term exposure to NO2 and PM10 and all-cause and cause-specific mortality in a prospective cohort of women', *Occupational and Environmental Medicine*, 70(3), pp. 179–186. doi: 10.1136/oemed-2012-100876

Hijmans, R. (2023) *raster: Geographic Data Analysis and Modeling: R package version 3.6-21*. Available at: https://cran.r-project.org/web/packages/raster/raster.pdf (Accessed: 14 June 2023).

Hilbe, J.M. (2015) *Practical guide to logistic regression*. Boca Raton: Chapman & Hall/CRC.

Hildebrandt, K. *et al.* (2009) 'Short-term effects of air pollution: a panel study of blood markers in patients with chronic pulmonary disease', *Particle and Fibre Toxicology*, 6(1), p. 25. doi: 10.1186/1743-8977-6-25

Hippel, P.T. von (2013) 'Should a Normal Imputation Model be Modified to Impute Skewed Variables?' *Sociological Methods & Research*, 42(1), pp. 105–138. doi: 10.1177/0049124112464866

HLNUG (2022a) *Messtechnik | Hessian State Office for Nature Conservation, Environment and Geology,* 16 August. Available at: https://www.hlnug.de/themen/luft/luftqualitaet/luftmessnetz/messtechnik (Accessed: 16 August 2022).

HLNUG (2022b) *Luftmessnetz,* 2 September. Available at: https://www.hlnug.de/themen/luft/luftqualitaet/luftmessnetz (Accessed: 2 September 2022).

HLNUG (2024) *Messdatenportal,* 1 July. Available at: https://www.hlnug.de/messwerte/datenportal (Accessed: 1 July 2024).

Hoffmann, B. *et al.* (2007) 'Residential exposure to traffic is associated with coronary atherosclerosis', *Circulation*, 116(5), pp. 489–496. doi: 10.1161/CIRCULATIONAHA.107.693622

Hoffmann, B. *et al.* (2009) 'Residential exposure to urban air pollution, ankle-brachial index, and peripheral arterial disease', *Epidemiology (Cambridge, Mass.)*, 20(2), pp. 280–288. doi: 10.1097/EDE.0b013e3181961ac2

Hoo, Z.H., Candlish, J. and Teare, D. (2017) 'What is an ROC curve?' *Emergency Medicine Journal*, 34(6), pp. 357–359. doi: 10.1136/emermed-2017-206735

Horton, N.J., Lipsitz, S.R. and Parzen, M. (2003) 'A Potential for Bias When Rounding in Multiple Imputation', *The American Statistician*, 57(4), pp. 229–232. doi: 10.1198/0003130032314

Hosmer, D.W. and Lemeshow, S. (2000) *Applied Logistic Regression*. Hoboken, NJ, USA: John Wiley & Sons, Inc.

Hosmer, D.W., Lemeshow, S. and Sturdivant, R.X. (2013) *Applied logistic regression*. 3rd edn. (Wiley series in probability and statistics). Hoboken: Wiley.

Huang, C.-H. *et al.* (2017) 'The Interaction Effects of Meteorological Factors and Air Pollution on the Development of Acute Coronary Syndrome', *Scientific Reports*, 7, p. 44004. doi: 10.1038/srep44004

Huangfu, P. and Atkinson, R. (2020) ‘Long-term exposure to NO2 and O3 and all-cause and respiratory mortality: A systematic review and meta-analysis’, *Environment International*, 144, p. 105998. doi: 10.1016/j.envint.2020.105998

Ibald-Mulli, A. *et al.* (2001) ‘Effects of air pollution on blood pressure: a population-based approach’, *American Journal of Public Health*, 91(4), pp. 571–577. doi: 10.2105/ajph.91.4.571

Ito, K., Leon, S.F. de and Lippmann, M. (2005) ‘Associations between ozone and daily mortality: analysis and meta-analysis’, *Epidemiology (Cambridge, Mass.)*, 16(4), pp. 446–457. doi: 10.1097/01.ede.0000165821.90114.7f

James, G. *et al.* (2021) *An introduction to statistical learning: With applications in R*. 2nd edn. (Springer texts in statistics). New York, New York: Springer. Available at: https://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=2985424.

Jamshidian, M. and Mata, M. (2007) ‘2 - Advances in Analysis of Mean and Covariance Structure when Data are Incomplete’, in Lee, S.-Y. (ed.) *Handbook of Latent Variable and Related Models : Handbook of Computing and Statistics with Applications.* Amsterdam: North-Holland, pp. 21–44. Available at: https://www.sciencedirect.com/science/article/pii/B9780444520449500057.

Jerrett, M. *et al.* (2004) ‘Do socioeconomic characteristics modify the short term association between air pollution and mortality? Evidence from a zonal time series in Hamilton, Canada’, *Journal of Epidemiology and Community Health*, 58(1), pp. 31–40. doi: 10.1136/jech.58.1.31

Jerrett, M. *et al.* (2009) ‘Long-term ozone exposure and mortality’, *The New England Journal of Medicine*, 360(11), pp. 1085–1095. doi: 10.1056/NEJMoa0803894

Junger, W. and Leon, A.P. de (2018) *mtsdi: Multivariate Time Series Data Imputation: R package version 0.3.5*. Available at: https://cran.r-project.org/package=mtsdi (Accessed: 9 September 2022).

Junger, W.L. and Ponce de Leon, A. (2015) ‘Imputation of missing data in time series for air pollutants’, *Atmospheric Environment*, 102, pp. 96–104. doi: 10.1016/j.atmosenv.2014.11.049

Junninen, H. *et al.* (2004) ‘Methods for imputation of missing values in air quality data sets’, *Atmospheric Environment*, 38(18), pp. 2895–2907. doi: 10.1016/j.atmosenv.2004.02.026

Katsouyanni, K. *et al.* (1997) ‘Short-term effects of ambient sulphur dioxide and particulate matter on mortality in 12 European cities: results from time series data from the APHEA project. Air Pollution and Health: a European Approach’, *BMJ (Clinical Research Ed.)*, 314(7095), pp. 1658–1663. doi: 10.1136/bmj.314.7095.1658

Katsouyanni, K. *et al.* (2001) ‘Confounding and effect modification in the short-term effects of ambient particles on total mortality: results from 29 European cities within the APHEA2 project’, *Epidemiology (Cambridge, Mass.)*, 12(5), pp. 521–531. doi: 10.1097/00001648-200109000-00011

Keller, T. *et al.* (2011) ‘Serial changes in highly sensitive troponin I assay and early diagnosis of myocardial infarction’, *JAMA*, 306(24), pp. 2684–2693. doi: 10.1001/jama.2011.1896

Kim, I.-S. *et al.* (2020) ‘Long-term fine particulate matter exposure and cardiovascular mortality in the general population: a nationwide cohort study’, *Journal of Cardiology*, 75(5), pp. 549–558. doi: 10.1016/j.jjcc.2019.11.004

Klompmaker, J.O. *et al.* (2021) ‘Air pollution and cardiovascular disease hospitalization - Are associations modified by greenness, temperature and humidity?’ *Environment International*, 156, p. 106715. doi: 10.1016/j.envint.2021.106715

Laden, F. *et al.* (2006) ‘Reduction in fine particulate air pollution and mortality: Extended follow-up of the Harvard Six Cities study’, *American Journal of Respiratory and Critical Care Medicine*, 173(6), pp. 667–672. doi: 10.1164/rccm.200503-443OC

Lanzinger, S. *et al.* (2014) ‘The impact of decreases in air temperature and increases in ozone on markers of endothelial function in individuals having type-2 diabetes’, *Environmental Research*, 134, pp. 331–338. doi: 10.1016/j.envres.2014.08.003

Lelieveld, J. *et al.* (2019) ‘Cardiovascular disease burden from ambient air pollution in Europe reassessed using novel hazard ratio functions’, *European Heart Journal*, 40(20), pp. 1590–1596. doi: 10.1093/eurheartj/ehz135

Li, Z. *et al.* (2019) ‘Longitudinal Effect of Ambient Air Pollution and Pollen Exposure on Asthma Control: The Patient-Reported Outcomes Measurement Information System (PROMIS) Pediatric Asthma Study’, *Academic Pediatrics*, 19(6), pp. 615–623. doi: 10.1016/j.acap.2019.03.010

Lipfert, F.W. *et al.* (2006) ‘PM2.5 constituents and related air quality variables as predictors of survival in a cohort of U.S. military veterans’, *Inhalation Toxicology*, 18(9), pp. 645–657. doi: 10.1080/08958370600742946

Lipsett, M.J. *et al.* (2011) ‘Long-term exposure to air pollution and cardiorespiratory disease in the California teachers study cohort’, *American Journal of Respiratory and Critical Care Medicine*, 184(7), pp. 828–835. doi: 10.1164/rccm.201012-2082OC

Little, R.J.A. and Rubin, D.B. (2020) *Statistical analysis with missing data*. 3rd edn. (Wiley series in probability and statistics). Hoboken, NJ: Wiley. Available at: https://onlinelibrary.wiley.com/doi/book/10.1002/9781119482260.

Liu, C. *et al.* (2018) ‘Ambient carbon monoxide and cardiovascular mortality: a nationwide time-series analysis in 272 cities in China’, *The Lancet Planetary Health*, 2(1), e12-e18. doi: 10.1016/S2542-5196(17)30181-X

Lu, K., Kang, J. and Wang, G. (2022) ‘The Impact of Air Quality on Cardiovascular Disease in Shanghai’, *Journal of Healthcare Engineering*, 2022, p. 4421686. doi: 10.1155/2022/4421686

Lüdecke, D. *et al.* (2021) ‘performance: An R Package for Assessment, Comparison and Testing of Statistical Models’, *Journal of Open Source Software*, 6(60), p. 3139. doi: 10.21105/joss.03139

Mahmood, S.S. *et al.* (2014) ‘The Framingham Heart Study and the epidemiology of cardiovascular disease: a historical perspective’, *The Lancet*, 383(9921), pp. 999–1008. doi: 10.1016/S0140-6736(13)61752-3

Mann, J.K. *et al.* (2002) ‘Air pollution and hospital admissions for ischemic heart disease in persons with congestive heart failure or arrhythmia’, *Environmental Health Perspectives*, 110(12), pp. 1247–1252. doi: 10.1289/ehp.021101247

Meng, X. *et al.* (2021) 'Short term associations of ambient nitrogen dioxide with daily total, cardiovascular, and respiratory mortality: multilocation analysis in 398 cities', *BMJ (Clinical Research Ed.)*, 372, n534. doi: 10.1136/bmj.n534

Metz, C.E. (1978) 'Basic principles of ROC analysis', *Seminars in Nuclear Medicine*, 8(4), pp. 283–298. doi: 10.1016/S0001-2998(78)80014-2

Miller, K.A. *et al.* (2007) 'Long-term exposure to air pollution and incidence of cardio-vascular events in women', *The New England Journal of Medicine*, 356(5), pp. 447–458. doi: 10.1056/NEJMoa054409

Miller, M.R. and Newby, D.E. (2020) 'Air pollution and cardiovascular disease: car sick', *Cardiovascular Research*, 116(2), pp. 279–294. doi: 10.1093/cvr/cvz228

Mills, N.L. *et al.* (2009) 'Adverse cardiovascular effects of air pollution', *Nature Clinical Practice. Cardiovascular Medicine*, 6(1), pp. 36–44. doi: 10.1038/ncpcardio1399

Milojevic, A. *et al.* (2014) 'Short-term effects of air pollution on a range of cardiovascular events in England and Wales: case-crossover analysis of the MINAP database, hospital admissions and mortality', *Heart*, 100(14), pp. 1093–1098. doi: 10.1136/heartjnl-2013-304963

Mirowsky, J.E. *et al.* (2015) 'Repeated measures of inflammation, blood pressure, and heart rate variability associated with traffic exposures in healthy adults', *Environmental Health : a Global Access Science Source*, 14, p. 66. doi: 10.1186/s12940-015-0049-0

Motesaddi Zarandi, S. *et al.* (2022) 'Effects of ambient air pollutants on hospital admissions and deaths for cardiovascular diseases: a time series analysis in Tehran', *Environmental Science and Pollution Research*, 29(12), pp. 17997–18009. doi: 10.1007/s11356-021-17051-y

Münzel, T. *et al.* (2017) 'Environmental stressors and cardio-metabolic disease: part II-mechanistic insights', *European Heart Journal*, 38(8), pp. 557–564. doi: 10.1093/eurheartj/ehw294

Münzel, T. *et al.* (2018) 'Effects of gaseous and solid constituents of air pollution on endothelial function', *European Heart Journal*, 39(38), pp. 3543–3550. doi: 10.1093/eurheartj/ehy481

Musil, C.M. *et al.* (2002) 'A comparison of imputation techniques for handling missing data', *Western Journal of Nursing Research*, 24(7), pp. 815–829. doi: 10.1177/019394502762477004

Newby, D.E. *et al.* (2015) 'Expert position paper on air pollution and cardiovascular disease', *European Heart Journal*, 36(2), 83-93b. doi: 10.1093/eurheartj/ehu458

Nguyen, A.-M., Malig, B.J. and Basu, R. (2021) 'The association between ozone and fine particles and mental health-related emergency department visits in California, 2005-2013', *PLoS ONE*, 16(4), e0249675. doi: 10.1371/journal.pone.0249675

Nick, T.G. and Campbell, K.M. (2007) 'Logistic regression', in Ambrosius, W.T. (ed.) *Topics in biostatistics.* (Methods in molecular biology, 404). Totowa, N.J.: Humana Press, pp. 273–301.

Nwanganga, F.C. and Chapple, M. (2020) *Practical machine learning in R*. London: ISTE, Ltd.; Wiley. Available at: https://onlinelibrary.wiley.com/doi/book/10.1002/9781119591542.

O'Brien, E. *et al.* (2023) 'Short-Term Association between Sulfur Dioxide and Mortality: A Multicountry Analysis in 399 Cities', *Environmental Health Perspectives*, 131(3), p. 37002. doi: 10.1289/EHP11112

O'Connor, G.T. *et al.* (1992) 'Multivariate prediction of in-hospital mortality associated with coronary artery bypass graft surgery. Northern New England Cardiovascular Disease Study Group', *Circulation*, 85(6), pp. 2110–2118. doi: 10.1161/01.cir.85.6.2110

OpenStreetMap contributors (2019) *Postleihzahl shapefiles*. Available at: https://www.suche-postleitzahl.org/downloads (Accessed: 2019).

Ostro, B. *et al.* (2010) 'Long-term exposure to constituents of fine particulate air pollution and mortality: results from the California Teachers Study', *Environmental Health Perspectives*, 118(3), pp. 363–369. doi: 10.1289/ehp.0901181

Padula, A.M. *et al.* (2014) 'Exposure to airborne polycyclic aromatic hydrocarbons during pregnancy and risk of preterm birth', *Environmental Research*, 135, pp. 221–226. doi: 10.1016/j.envres.2014.09.014

Pebesma, E. (2018) 'Simple Features for R: Standardized Support for Spatial Vector Data', *The R Journal*, 10(1), p. 439. doi: 10.32614/RJ-2018-009

Pebesma, E., Mailund, T. and Hiebert, J. (2016) 'Measurement Units in R', *The R Journal*, 8(2), p. 486. doi: 10.32614/RJ-2016-061

Pepinsky, T.B. (2018) 'A Note on Listwise Deletion versus Multiple Imputation', *Political Analysis*, 26(4), pp. 480–488. doi: 10.1017/pan.2018.18

Peters, A. *et al.* (2001) 'Particulate air pollution is associated with an acute phase response in men; results from the MONICA-Augsburg Study', *European Heart Journal*, 22(14), pp. 1198–1204. doi: 10.1053/euhj.2000.2483

Peters, A. *et al.* (2004) 'Exposure to traffic and the onset of myocardial infarction', *The New England Journal of Medicine*, 351(17), pp. 1721–1730. doi: 10.1056/NEJMoa040203

Plaia, A. and BONDI, A. (2006) 'Single imputation method of missing values in environmental pollution data sets', *Atmospheric Environment*, 40(38), pp. 7316–7330. doi: 10.1016/j.atmosenv.2006.06.040

Pope, C.A. *et al.* (1995) 'Particulate air pollution as a predictor of mortality in a prospective study of U.S. adults', *American Journal of Respiratory and Critical Care Medicine*, 151(3 Pt 1), pp. 669–674. doi: 10.1164/ajrccm/151.3_Pt_1.669

Pope, C.A. *et al.* (2002) 'Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution', *JAMA*, 287(9), pp. 1132–1141. doi: 10.1001/jama.287.9.1132

Pope, C.A. *et al.* (2004) 'Cardiovascular mortality and long-term exposure to particulate air pollution: epidemiological evidence of general pathophysiological pathways of disease', *Circulation*, 109(1), pp. 71–77. doi: 10.1161/01.CIR.0000108927.80044.7F

Pope, C.A. *et al.* (2006) 'Ischemic heart disease events triggered by short-term exposure to fine particulate air pollution', *Circulation*, 114(23), pp. 2443–2448. doi: 10.1161/CIRCULATIONAHA.106.636977

Pope, C.A. *et al.* (2015) 'Relationships between fine particulate air pollution, cardiometabolic disorders, and cardiovascular mortality', *Circulation Research*, 116(1), pp. 108–115. doi: 10.1161/CIRCRESAHA.116.305060

Pope, C.A. *et al.* (2015) 'Short-Term Exposure to Fine Particulate Matter Air Pollution Is Preferentially Associated With the Risk of ST-Segment Elevation Acute Coronary Events', *Journal of the American Heart Association*, 4(12) (10pp). doi: 10.1161/JAHA.115.002506

Puett, R.C. *et al.* (2008) 'Chronic particulate exposure, mortality, and coronary heart disease in the nurses' health study', *American Journal of Epidemiology*, 168(10), pp. 1161–1168. doi: 10.1093/aje/kwn232

R Core Team (2022) *R: A language and environment for statistical computing. R Foundation for Statistical Computing*. Available at: https://www.r-project.org/.

Raaschou-Nielsen, O. *et al.* (2012) 'Traffic air pollution and mortality from cardiovascular disease and all causes: a Danish cohort study', *Environmental Health*, 11(1), p. 60. doi: 10.1186/1476-069X-11-60

Rage, E. *et al.* (2009) 'Air pollution and asthma severity in adults', *Occupational and Environmental Medicine*, 66(3), pp. 182–188. doi: 10.1136/oem.2007.038349

Robin, X. *et al.* (2011) 'pROC: an open-source package for R and S+ to analyze and compare ROC curves', *BMC Bioinformatics*, 12, p. 77. doi: 10.1186/1471-2105-12-77

Rodríguez, D., Cobo-Cuenca, A.I. and Quiles, R. (2022) 'Effects of air pollution on daily hospital admissions for cardiovascular diseases in Castilla-La Mancha, Spain: a region with moderate air quality', *Air Quality, Atmosphere & Health*, 15(4), pp. 591–604. doi: 10.1007/s11869-021-01144-1

Rodwell, L. *et al.* (2014) 'Comparison of methods for imputing limited-range variables: a simulation study', *BMC Medical Research Methodology*, 14(1), p. 57. doi: 10.1186/1471-2288-14-57

Ross, R.K., Breskin, A. and Westreich, D. (2020) 'When Is a Complete-Case Approach to Missing Data Valid? The Importance of Effect-Measure Modification', *American Journal of Epidemiology*, 189(12), pp. 1583–1589. doi: 10.1093/aje/kwaa124

Rossi, R.J. (2022) *Applied biostatistics for the health sciences*. Hoboken, NJ: John Wiley & Sons.

Rubin, D.B. (1976) 'Inference and Missing Data', *Biometrika*, 63(3), p. 581. doi: 10.2307/2335739

Rückerl, R. *et al.* (2006) 'Air pollution and markers of inflammation and coagulation in patients with coronary heart disease', *American Journal of Respiratory and Critical Care Medicine*, 173(4), pp. 432–441. doi: 10.1164/rccm.200507-1123OC

Samet, J.M. *et al.* (2000) 'Fine particulate air pollution and mortality in 20 U.S. cities, 1987-1994', *The New England Journal of Medicine*, 343(24), pp. 1742–1749. doi: 10.1056/NEJM200012143432401

Samoli, E. *et al.* (2001) 'Investigating regional differences in short-term effects of air pollution on daily mortality in the APHEA project: a sensitivity analysis for controlling long-term trends and seasonality', *Environmental Health Perspectives*, 109(4), pp. 349–353. doi: 10.1289/ehp.01109349

Samoli, E. *et al.* (2007) 'Short-term effects of carbon monoxide on mortality: an analysis within the APHEA project', *Environmental Health Perspectives*, 115(11), pp. 1578–1583. doi: 10.1289/ehp.10375

Sarnat, J.A. *et al.* (2007) 'Panel discussion review: session 1--exposure assessment and related errors in air pollution epidemiologic studies', *Journal of Exposure Science & Environmental Epidemiology*, 17 Suppl 2(2), S75-82. doi: 10.1038/sj.jes.7500621

Schildcrout, J.S. *et al.* (2006) 'Ambient air pollution and asthma exacerbations in children: an eight-city analysis', *American Journal of Epidemiology*, 164(6), pp. 505–517. doi: 10.1093/aje/kwj225

Schneider, A. *et al.* (2008) 'Endothelial dysfunction: associations with exposure to ambient fine particles in diabetic individuals', *Environmental Health Perspectives*, 116(12), pp. 1666–1674. doi: 10.1289/ehp.11666

Schneider, A. *et al.* (2010) 'Association of cardiac and vascular changes with ambient PM2.5 in diabetic individuals', *Particle and Fibre Toxicology*, 7, p. 14. doi: 10.1186/1743-8977-7-14

Schneider, A. *et al.* (2011) 'Ambient PM2.5 exposure up-regulates the expression of costimulatory receptors on circulating monocytes in diabetic individuals', *Environmental Health Perspectives*, 119(6), pp. 778–783. doi: 10.1289/ehp.1002543

Schwartz, J. *et al.* (2008) 'The effect of dose and timing of dose on the association between airborne particles and survival', *Environmental Health Perspectives*, 116(1), pp. 64–69. doi: 10.1289/ehp.9955

SCORE2 working group and ESC Cardiovascular risk collaboration (2021) 'SCORE2 risk prediction algorithms: new models to estimate 10-year risk of cardiovascular disease in Europe', *European Heart Journal*, 42(25), pp. 2439–2454. doi: 10.1093/eurheartj/ehab309

Spencer-Hwang, R. *et al.* (2011) 'Ambient air pollutants and risk of fatal coronary heart disease among kidney transplant recipients', *American Journal of Kidney Diseases : the Official Journal of the National Kidney Foundation*, 58(4), pp. 608–616. doi: 10.1053/j.ajkd.2011.05.017

Stafoggia, M. *et al.* (2014) 'Long-term exposure to ambient air pollution and incidence of cerebrovascular events: results from 11 European cohorts within the ESCAPE project', *Environmental Health Perspectives*, 122(9), pp. 919–925. doi: 10.1289/ehp.1307301

Stieb, D.M., Judek, S. and Burnett, R.T. (2002) 'Meta-analysis of time-series studies of air pollution and mortality: effects of gases and particles and the influence of cause of death, age, and season', *Journal of the Air & Waste Management Association*, 52(4), pp. 470–484. doi: 10.1080/10473289.2002.10470794

Stockfelt, L. *et al.* (2015) 'Long term effects of residential NO(x) exposure on total and cause-specific mortality and incidence of myocardial infarction in a Swedish cohort', *Environmental Research*, 142, pp. 197–206. doi: 10.1016/j.envres.2015.06.045

Thongphunchung, K. *et al.* (2022) 'Outpatient Department Visits and Mortality with Various Causes Attributable to Ambient Air Pollution in the Eastern Economic Corridor of Thailand', *International Journal of Environmental Research and Public Health*, 19(13), p. 7683. doi: 10.3390/ijerph19137683

Tierney, N. and Cook, D. (2023) 'Expanding Tidy Data Principles to Facilitate Missing Data Exploration, Visualization and Assessment of Imputations', *Journal of Statistical Software*, 105(7). doi: 10.18637/jss.v105.i07 van Buuren, S. and Groothuis-Oudshoorn, K. (2011) ‘mice : Multivariate Imputation by Chained Equations in R’, *Journal of Statistical Software*, 45(3), pp. 1–67. doi: 10.18637/jss.v045.i03

van Buuren, S.a.T.Q.o.L.L.T.N. (2018) *Flexible Imputation of Missing Data, Second Edition*. 2nd edn. (Chapman & Hall/CRC Interdisciplinary Statistics). Boca Raton: CRC Press Taylor and Francis Group.

Visseren, F.L.J. *et al.* (2021) ‘2021 ESC Guidelines on cardiovascular disease prevention in clinical practice’, *European Heart Journal*, 42(34), pp. 3227–3337. doi: 10.1093/eurheartj/ehab484

Wang, E., Cook, D. and Hyndman, R.J. (2020) ‘A New Tidy Data Structure to Support Exploration and Modeling of Temporal Data’, *Journal of Computational and Graphical Statistics*, 29(3), pp. 466–478. doi: 10.1080/10618600.2019.1695624

Wei, T. and Simko, V. (2021) *R package 'corrplot': Visualization of a Correlation Matrix (Version 0.92)*. Available at: https://github.com/taiyun/corrplot (Accessed: 9 May 2023).

Wickham, H. (2016) *ggplot 2: Elegant graphics for data analysis*. (Use R!). New York: Springer. Available at: https://ggplot2.tidyverse.org/ (Accessed: 1 January 2023).

Wickham, H. *et al.* (2019) ‘Welcome to the Tidyverse’, *Journal of Open Source Software*, 4(43), p. 1686. doi: 10.21105/joss.01686

Wickham, H. (2022) *_stringr: Simple, Consistent Wrappers for Common String Operations_: R package version 1.4.0*. Available at: https://cran.r-project.org/package=stringr.

Wolf, K. *et al.* (2015) ‘Associations between short-term exposure to particulate matter and ultrafine particles and myocardial infarction in Augsburg, Germany’, *International Journal of Hygiene and Environmental Health*, 218(6), pp. 535–542. doi: 10.1016/j.ijheh.2015.05.002

Wong, T.W. *et al.* (2002) ‘Associations between daily mortalities from respiratory and cardiovascular diseases and air pollution in Hong Kong, China’, *Occupational and Environmental Medicine*, 59(1), pp. 30–35. doi: 10.1136/oem.59.1.30

World Health Organization (2022a) *Ambient (outdoor) air pollution,* 23 July. Available at: https://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health (Accessed: 23 July 2023).

World Health Organization (2022b) ‘Noncommunicable diseases’. *Global Health Observatory,* 9 April. Available at: https://www.who.int/data/gho/data/themes/noncommunicable-diseases (Accessed: 10 April 2022).

Yang, L., Li, C. and Tang, X. (2020) ‘The Impact of PM2.5 on the Host Defense of Respiratory System’, *Frontiers in Cell and Developmental Biology*, 8, p. 91. doi: 10.3389/fcell.2020.00091

Yoon, S.-J. *et al.* (2020) ‘Ambient carbon monoxide exposure and elevated risk of mortality in the glioblastoma patients: A double-cohort retrospective observational study’, *Cancer Medicine*, 9(23), pp. 9018–9026. doi: 10.1002/cam4.3572

Yorifuji, T., Suzuki, E. and Kashima, S. (2014) ‘Cardiovascular emergency hospital visits and hourly changes in air pollution’, *Stroke*, 45(5), pp. 1264–1268. doi: 10.1161/STROKEAHA.114.005227

Zeger, S.L. *et al.* (2008) ‘Mortality in the Medicare population and chronic exposure to fine particulate air pollution in urban centers (2000-2005)’, *Environmental Health Perspectives*, 116(12), pp. 1614–1619. doi: 10.1289/ehp.11449

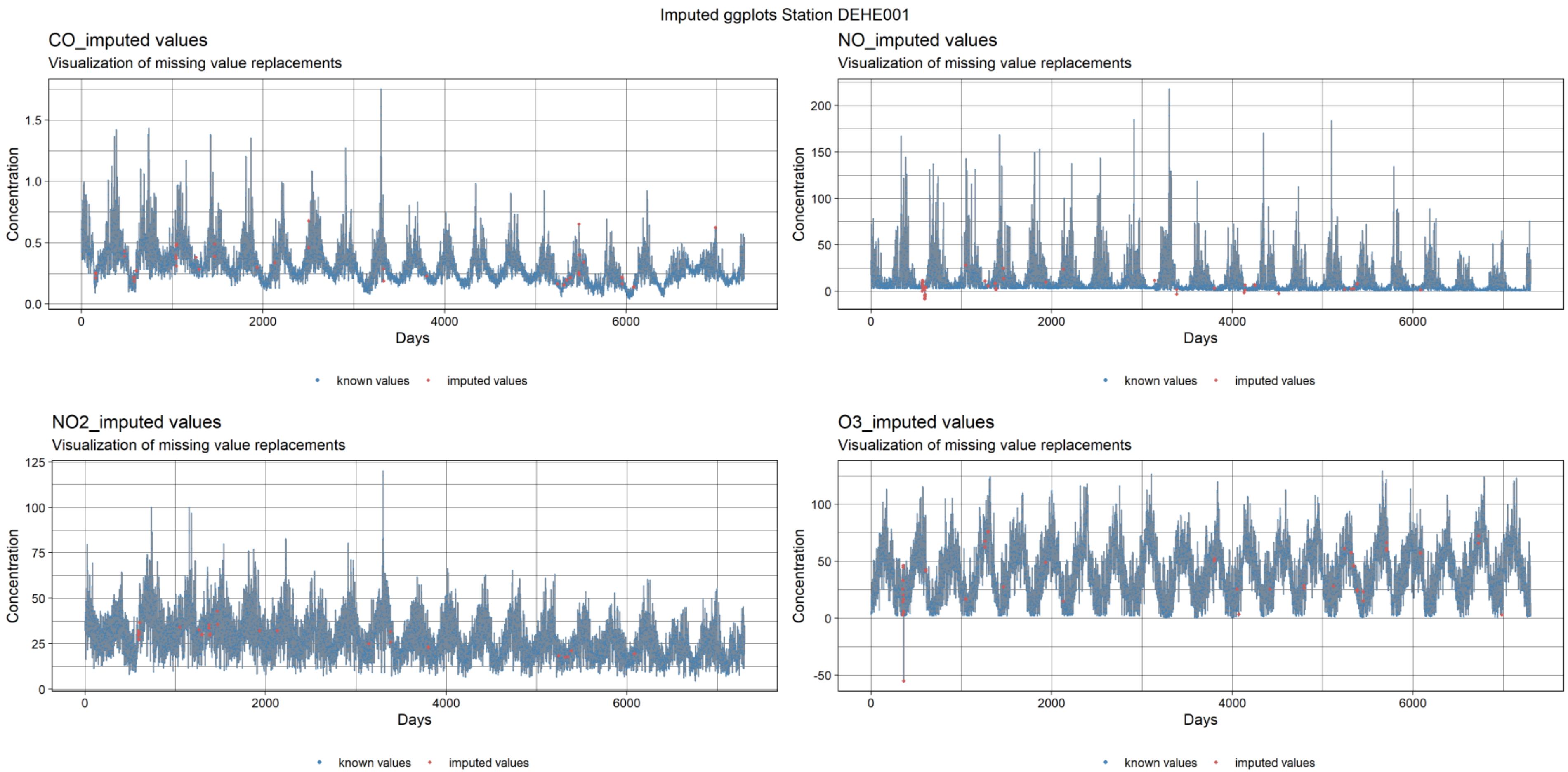
Zeileis, A. and Hothorn, T. (2002) ‘Diagnostic Checking in Regression Relationships’, *R News*, 2(3), pp. 7–10. Available at: https://cran.r-project.org/doc/Rnews/.

Zou, K.H., O'Malley, A.J. and Mauri, L. (2007) ‘Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models’, *Circulation*, 115(5), pp. 654–657. doi: 10.1161/CIRCULATIONAHA.105.594929

Zweig, M.H. and Campbell, G. (1993) ‘Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine’, *Clinical Chemistry*, 39(4), pp. 561–577.

Appendix

Imputed air pollution variables

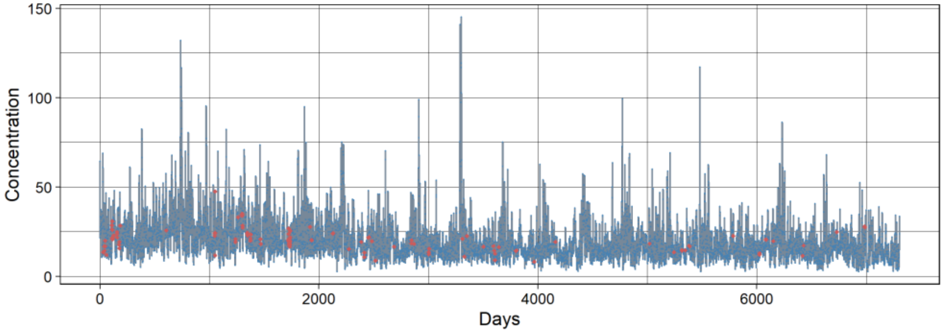


Imputed ggplots Station DEHE001
CO_imputed values
Visualization of missing value replacements
Concentration
Days
known values
imputed values
NO_imputed values
Visualization of missing value replacements
Concentration
Days
known values
imputed values
NO2_imputed values
Visualization of missing value replacements
Concentration
Days
known values
imputed values
O3_imputed values
Visualization of missing value replacements
Concentration
Days
known values
imputed values

Imputed ggplots Station DEHE001

PM10_imputed values

Visualization of missing value replacements



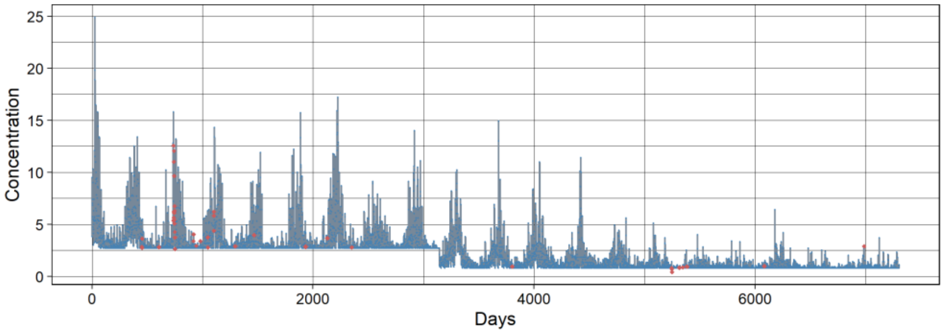
• known values • imputed values

______PM25_no imputed values

No imputation forPM25at station DEHE001
because more than 10% of the values missing

SO2_imputed values

Visualization of missing value replacements



• known values • imputed values

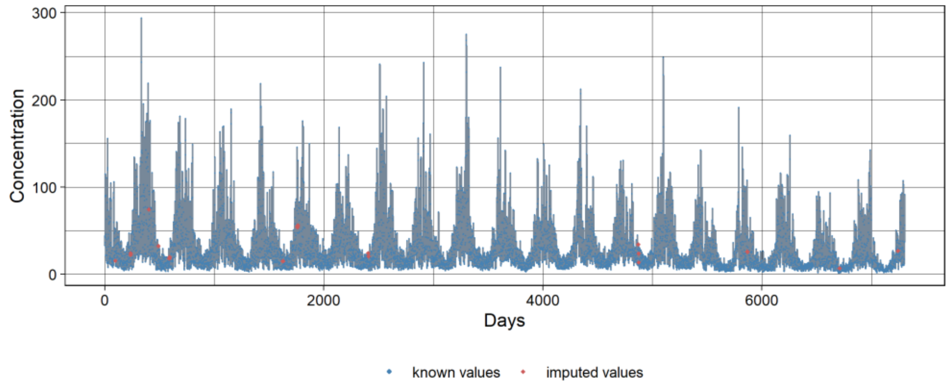
Imputed ggplots Station DEHE005

_____CO_no imputed values

No imputation forCOat station DEHE005
because more than 10% of the values missing

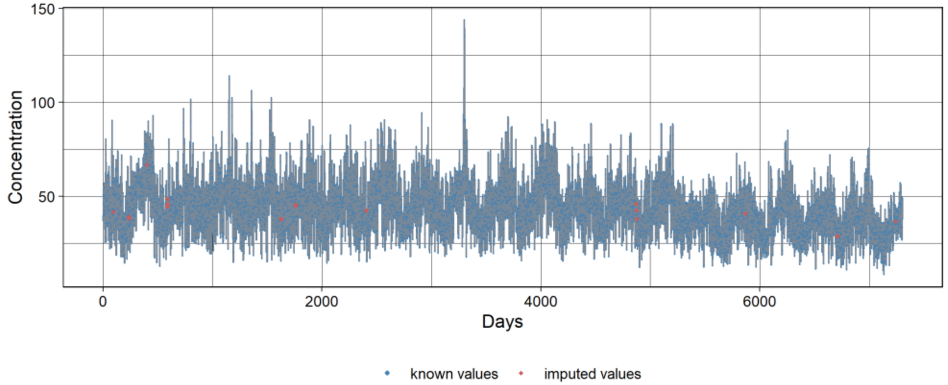
NO_imputed values

Visualization of missing value replacements



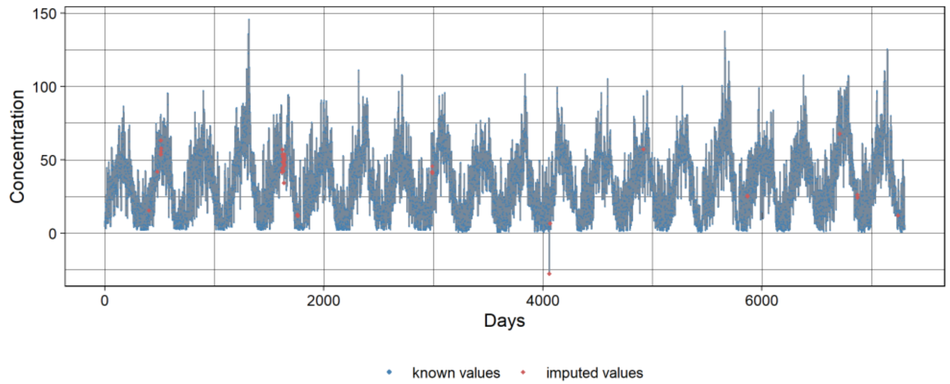
NO2_imputed values

Visualization of missing value replacements



O3_imputed values

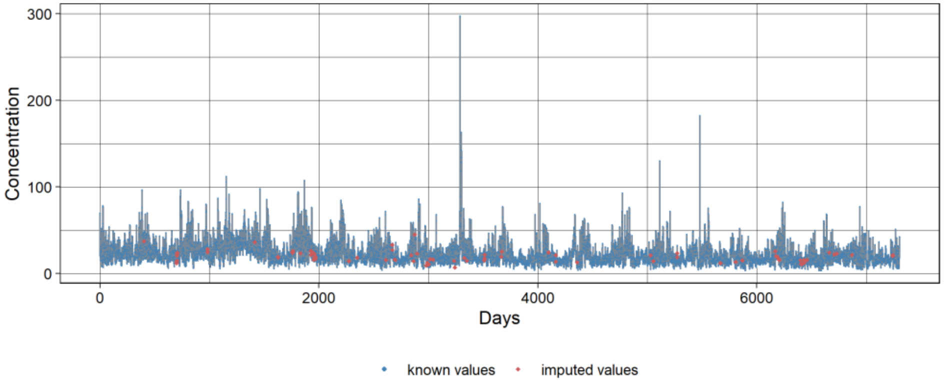
Visualization of missing value replacements



Imputed ggplots Station DEHE005

PM10_imputed values

Visualization of missing value replacements

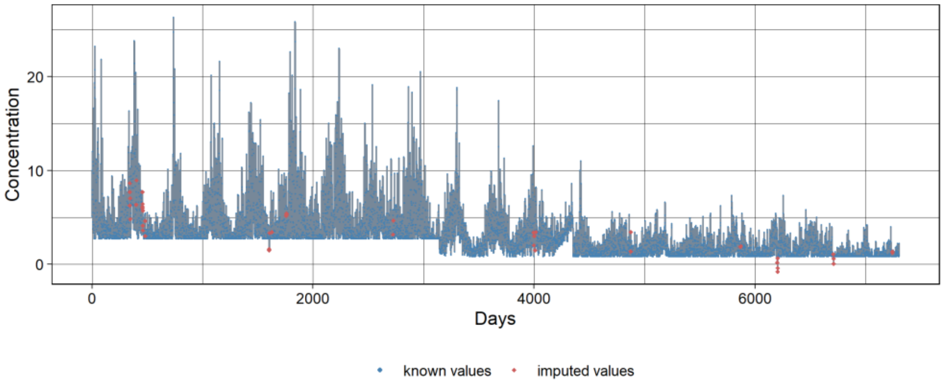


______PM25_no imputed values

No imputation forPM25at station DEHE005
because more than 10% of the values missing

SO2_imputed values

Visualization of missing value replacements



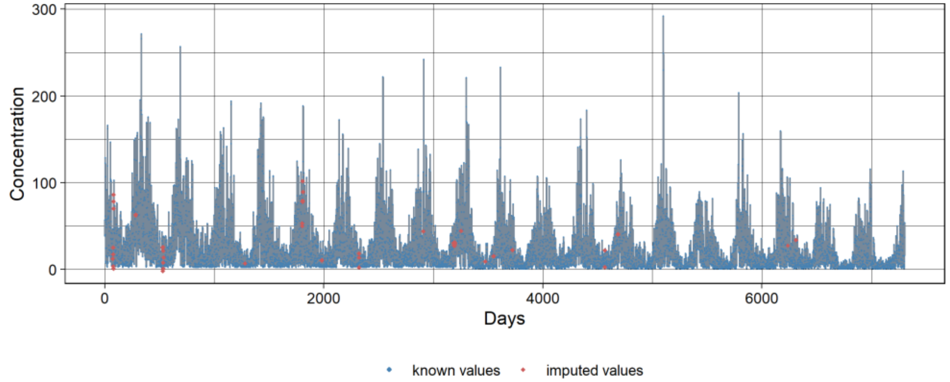
Imputed ggplots Station DEHE008

_____CO_no imputed values

No imputation forCOat station DEHE008
because more than 10% of the values missing

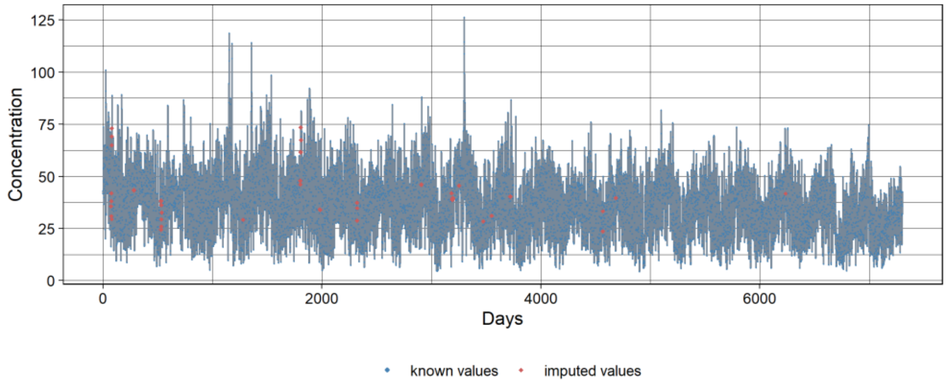
NO_imputed values

Visualization of missing value replacements



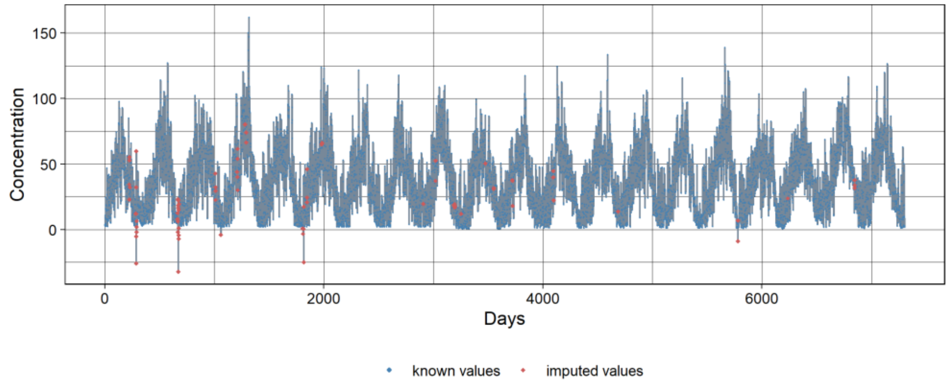
NO2_imputed values

Visualization of missing value replacements



O3_imputed values

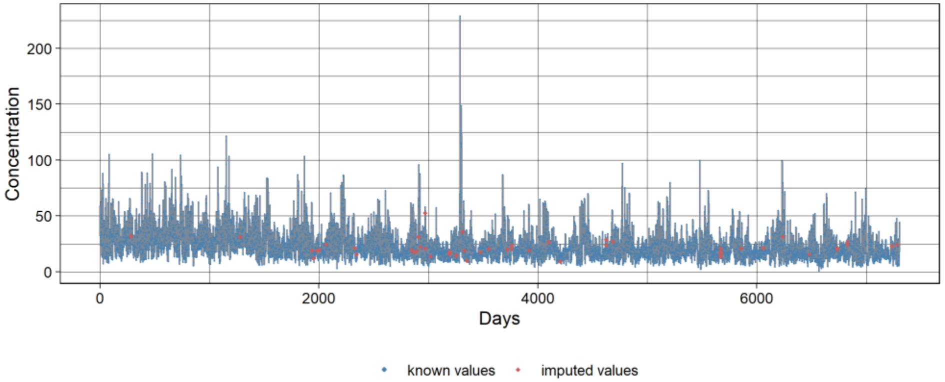
Visualization of missing value replacements



Imputed ggplots Station DEHE008

PM10_imputed values

Visualization of missing value replacements



_____PM25_no imputed values

No imputation forPM25at station DEHE008
because more than 10% of the values missing

_____SO2_no imputed values

No imputation forSO2at station DEHE008
because more than 10% of the values missing

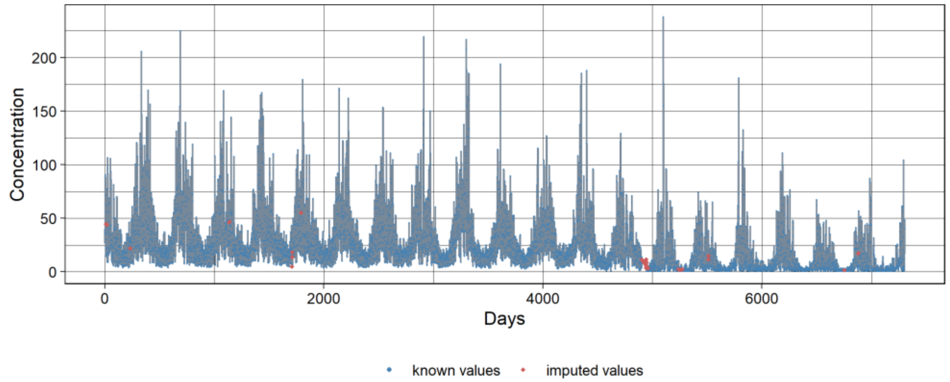
Imputed ggplots Station DEHE011

_____CO_no imputed values

No imputation forCOat station DEHE011
because more than 10% of the values missing

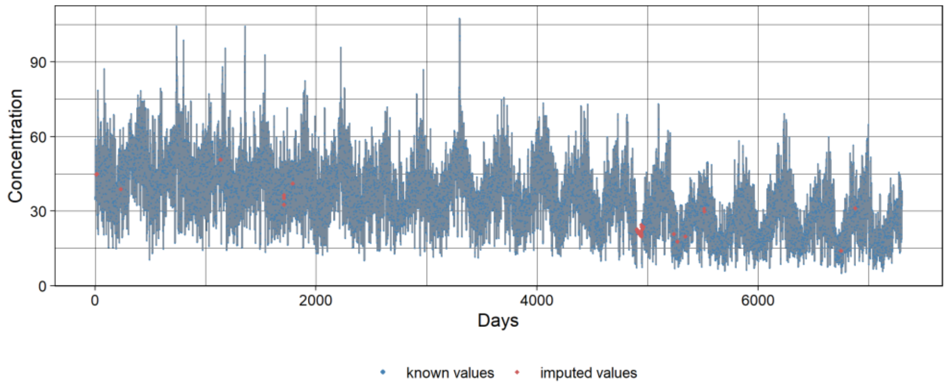
NO_imputed values

Visualization of missing value replacements



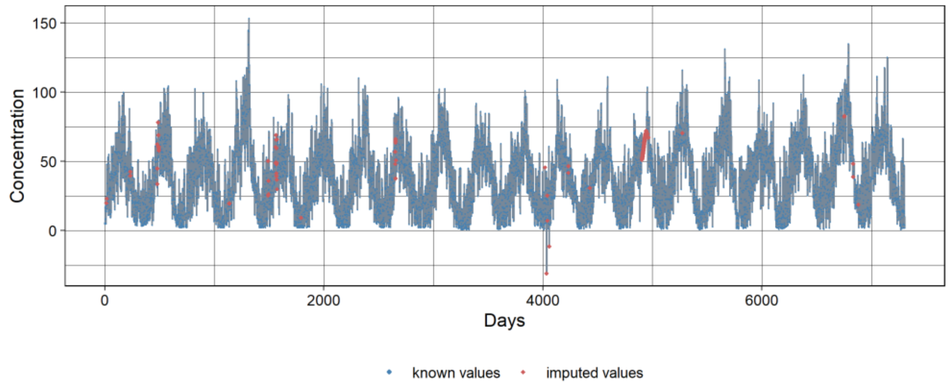
NO2_imputed values

Visualization of missing value replacements



O3_imputed values

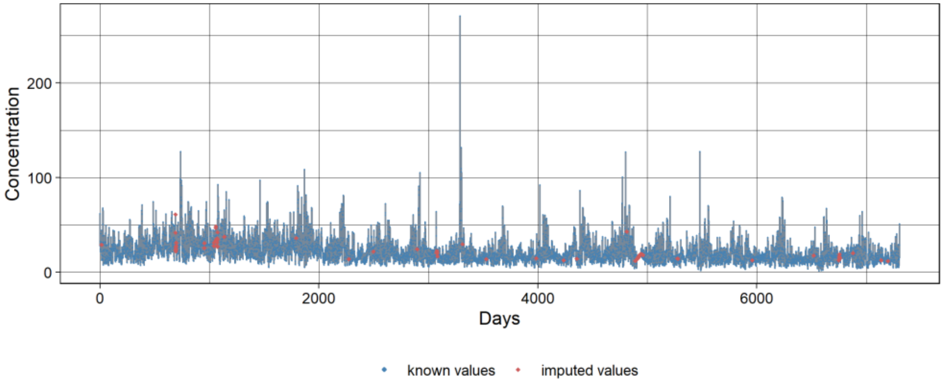
Visualization of missing value replacements



Imputed ggplots Station DEHE011

PM10_imputed values

Visualization of missing value replacements

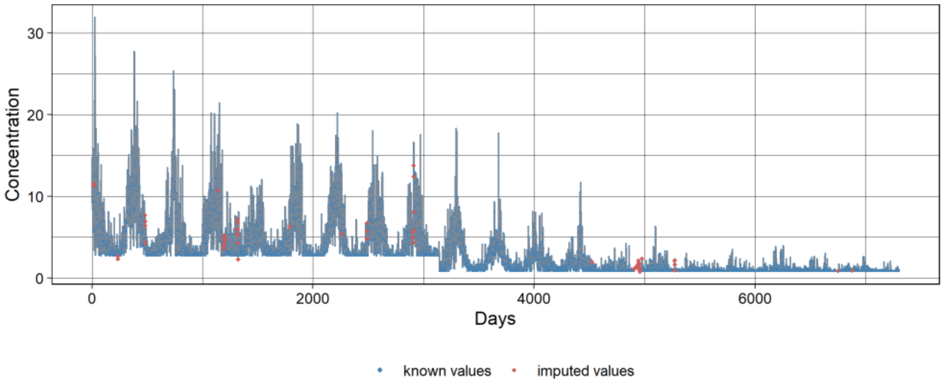


_____PM25_no imputed values

No imputation forPM25at station DEHE011
because more than 10% of the values missing

SO2_imputed values

Visualization of missing value replacements



Imputed ggplots Station DEHE013

______CO_no imputed values

No imputation for this airpollutant at station DEHE013
because more than 10% of the values missing

______NO_no imputed values

No imputation for this airpollutant at station DEHE013
because more than 10% of the values missing

______NO2_no imputed values

No imputation for this airpollutant at station DEHE013
because more than 10% of the values missing

______O3_no imputed values

No imputation for this airpollutant at station DEHE013
because more than 10% of the values missing

Imputed ggplots Station DEHE013

______PM10_no imputed values

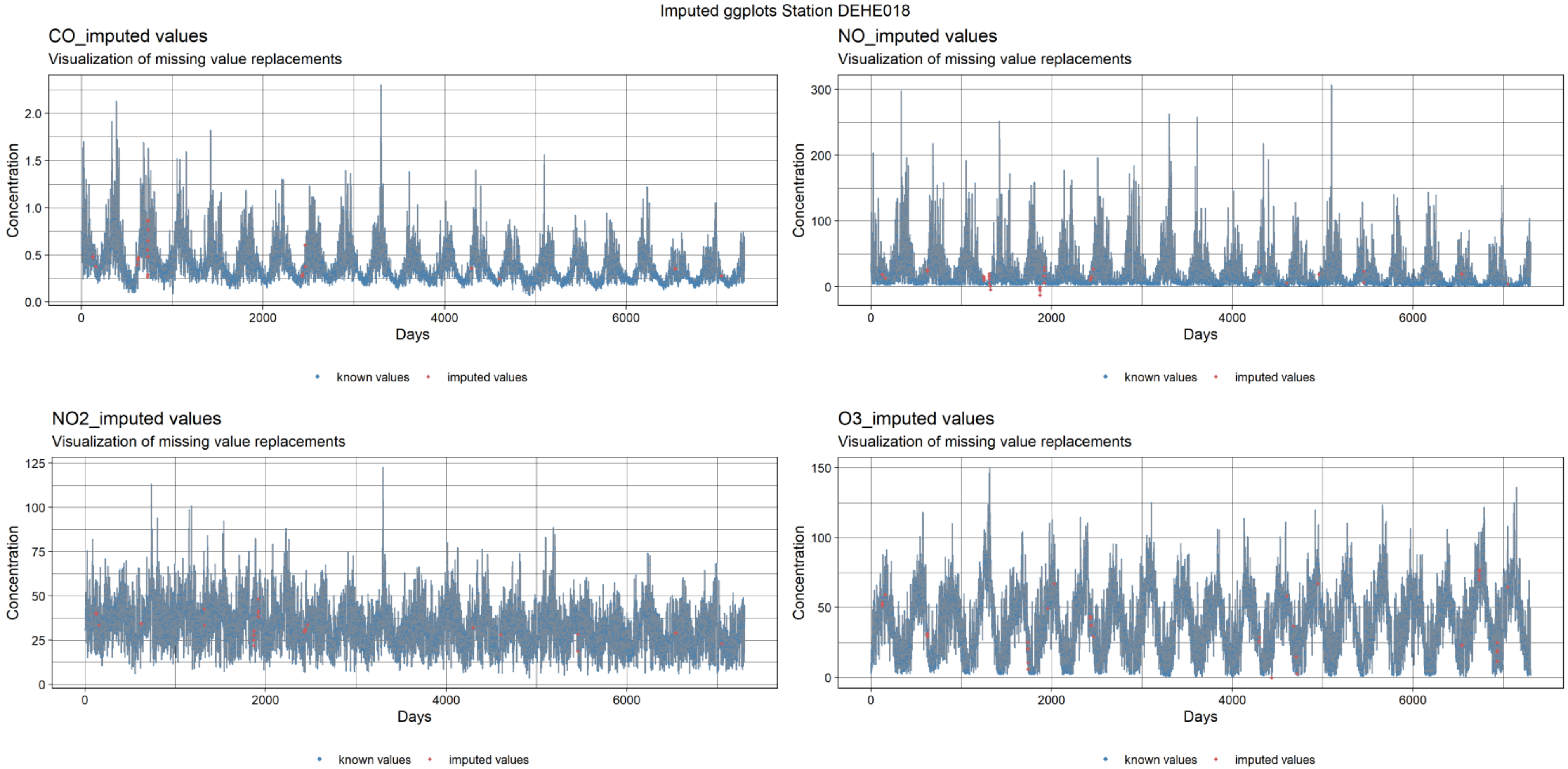
No imputation for this airpollutant at station DEHE013
because more than 10% of the values missing

______PM25_no imputed values

No imputation for this airpollutant at station DEHE013
because more than 10% of the values missing

______SO2_no imputed values

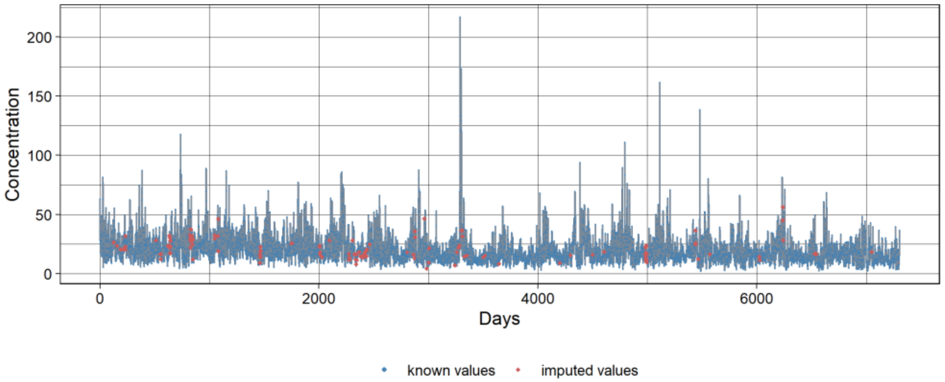
No imputation for this airpollutant at station DEHE013
because more than 10% of the values missing


Imputed ggplots Station DEHE018
CO_imputed values
Visualization of missing value replacements
Concentration
2.0
1.5
1.0
0.5
0.0
0
2000
4000
6000
Days
known values
imputed values
NO_imputed values
Visualization of missing value replacements
Concentration
300
200
100
0
0
2000
4000
6000
Days
known values
imputed values
NO2_imputed values
Visualization of missing value replacements
Concentration
125
100
75
50
25
0
0
2000
4000
6000
Days
known values
imputed values
O3_imputed values
Visualization of missing value replacements
Concentration
150
100
50
0
0
2000
4000
6000
Days
known values
imputed values

Imputed ggplots Station DEHE018

PM10_imputed values

Visualization of missing value replacements

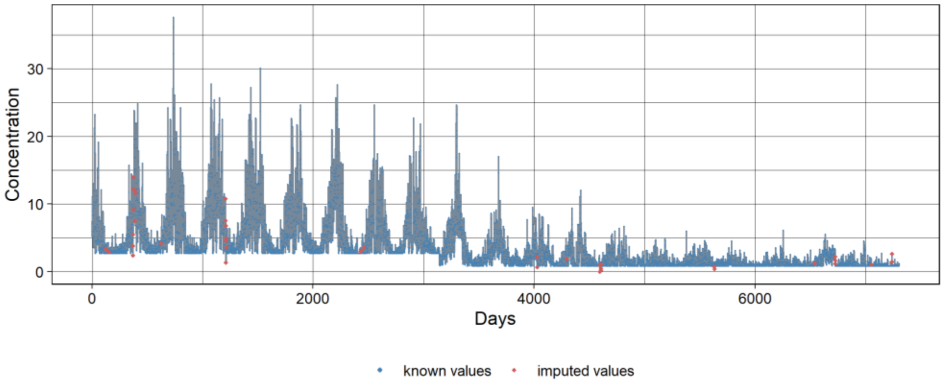


______PM25_no imputed values

No imputation forPM25at station DEHE018
because more than 10% of the values missing

SO2_imputed values

Visualization of missing value replacements



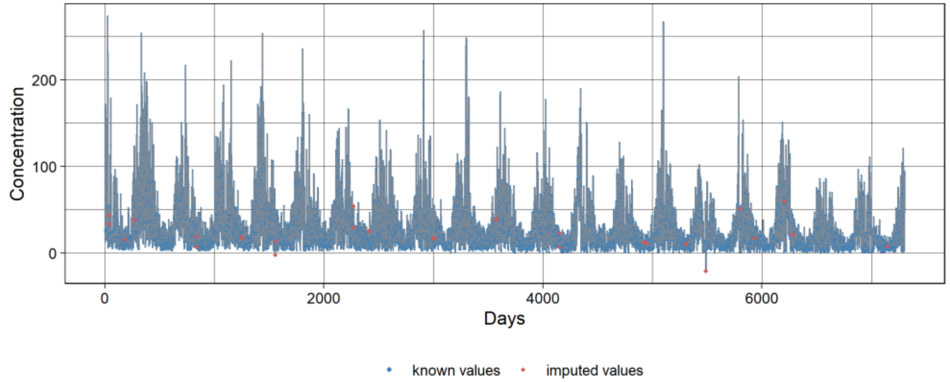
Imputed ggplots Station DEHE020

______CO_no imputed values

No imputation forCOat station DEHE020
because more than 10% of the values missing

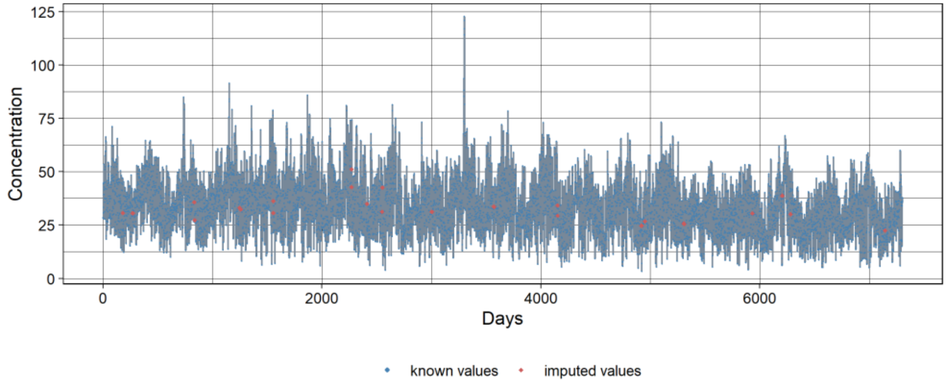
NO_imputed values

Visualization of missing value replacements



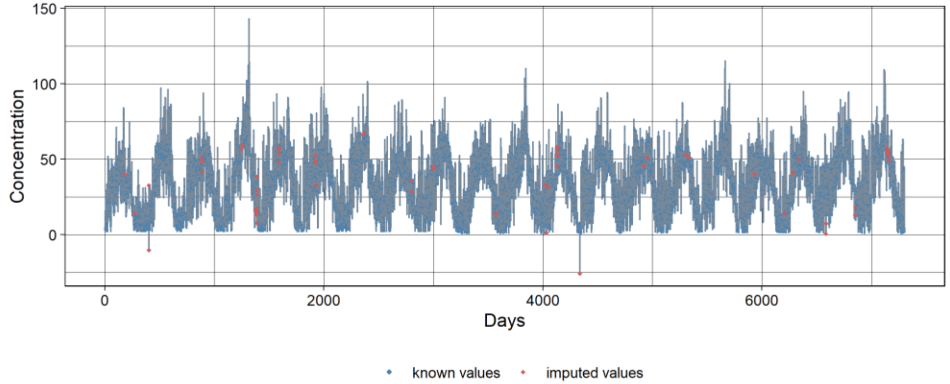
NO2_imputed values

Visualization of missing value replacements



O3_imputed values

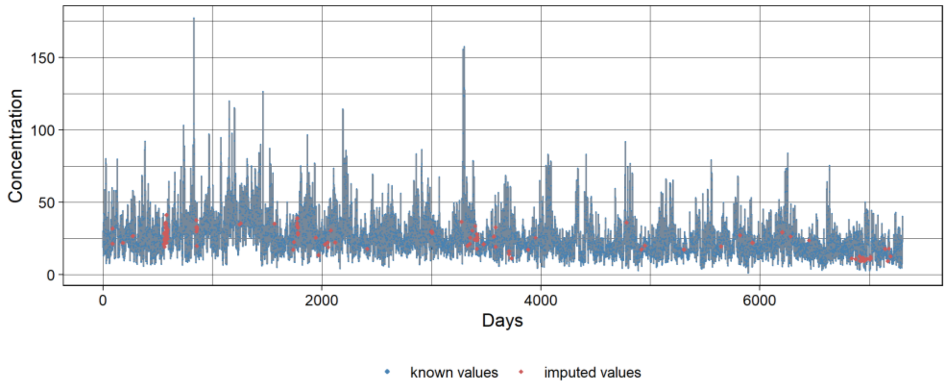
Visualization of missing value replacements



Imputed ggplots Station DEHE020

PM10_imputed values

Visualization of missing value replacements

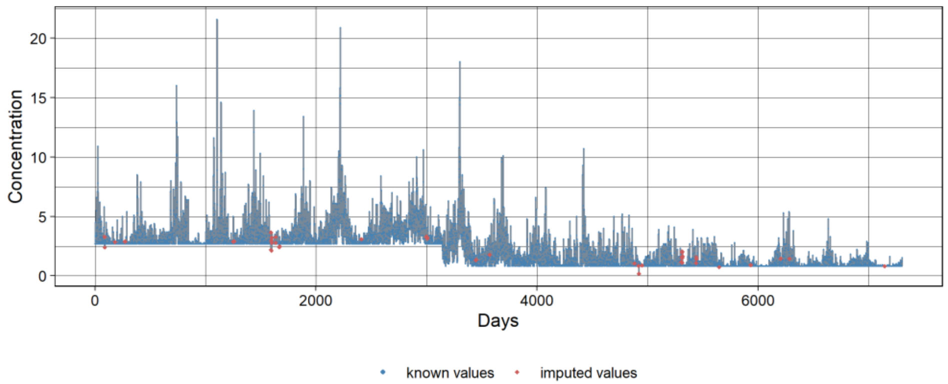


_____PM25_no imputed values

No imputation forPM25at station DEHE020
because more than 10% of the values missing

SO2_imputed values

Visualization of missing value replacements



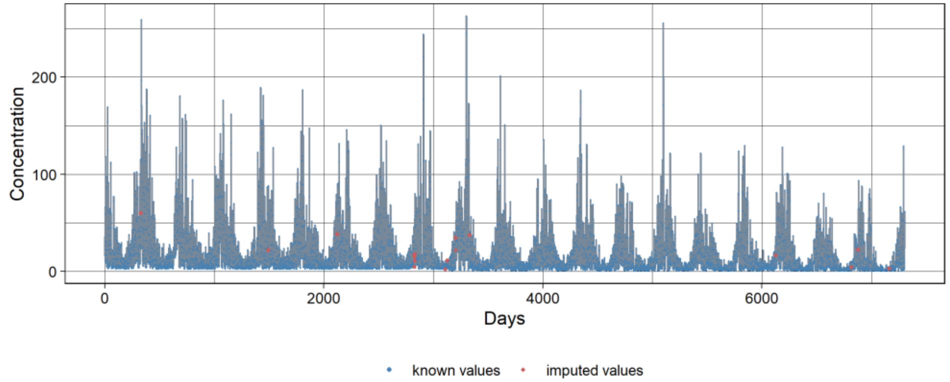
Imputed ggplots Station DEHE022

______CO_no imputed values

No imputation forCOat station DEHE022
because more than 10% of the values missing

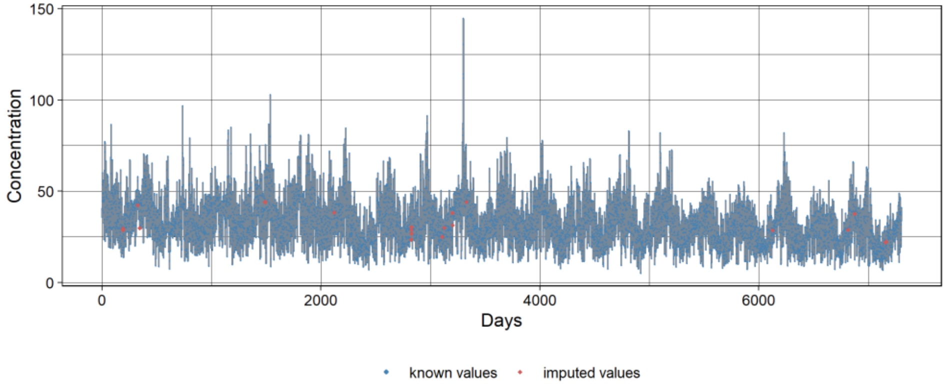
NO_imputed values

Visualization of missing value replacements



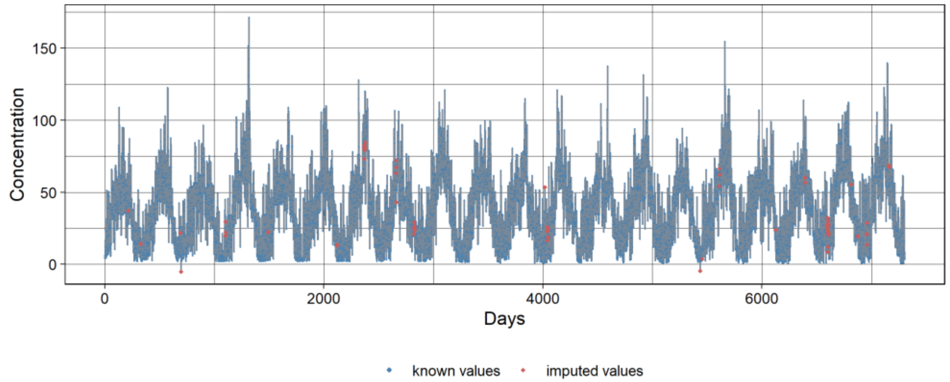
NO2_imputed values

Visualization of missing value replacements



O3_imputed values

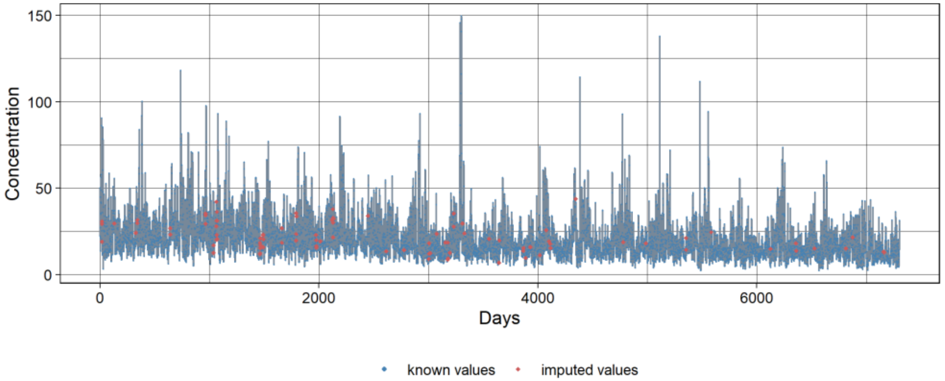
Visualization of missing value replacements



Imputed ggplots Station DEHE022

PM10_imputed values

Visualization of missing value replacements

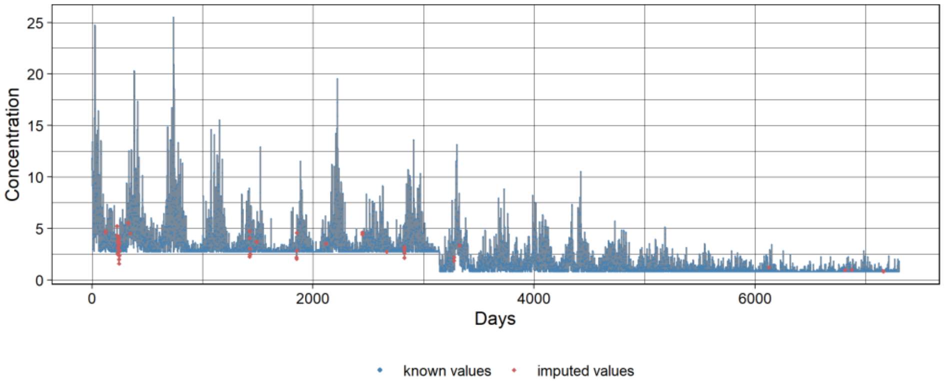


______PM25_no imputed values

No imputation forPM25at station DEHE022
because more than 10% of the values missing

SO2_imputed values

Visualization of missing value replacements



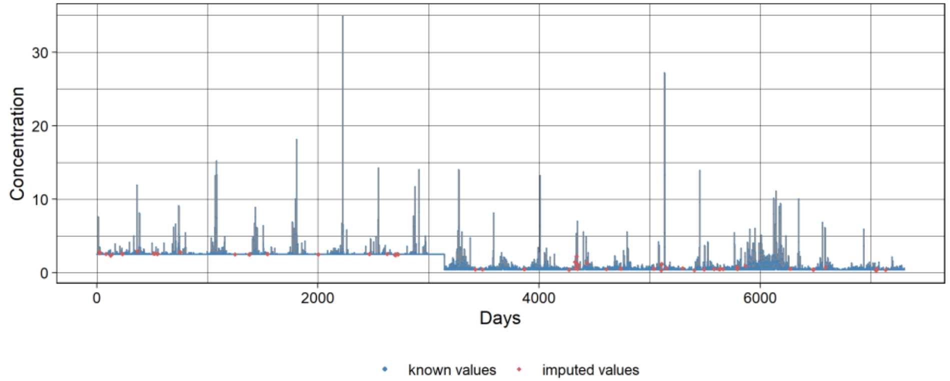
Imputed ggplots Station DEHE024

_____CO_no imputed values

No imputation forCOat station DEHE024
because more than 10% of the values missing

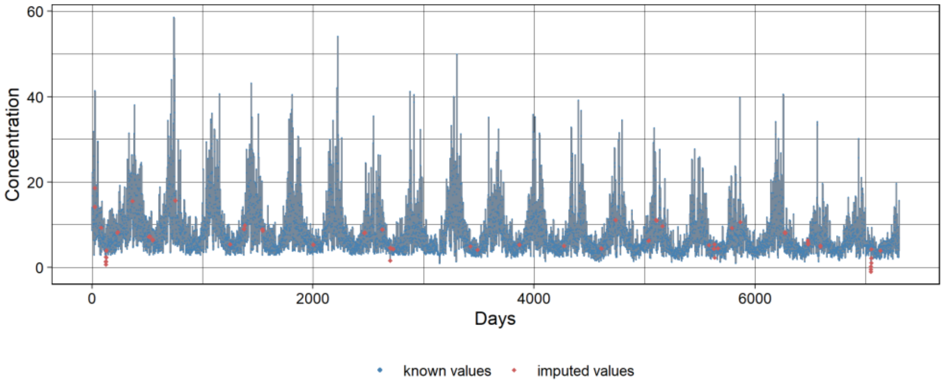
NO_imputed values

Visualization of missing value replacements



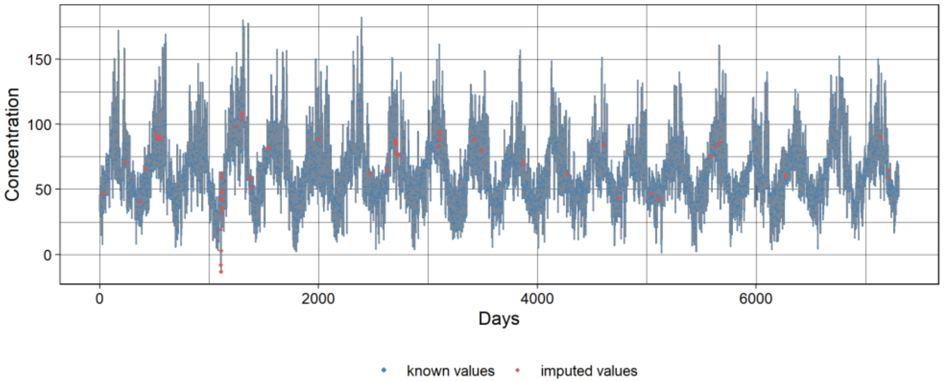
NO2_imputed values

Visualization of missing value replacements



O3_imputed values

Visualization of missing value replacements



Imputed ggplots Station DEHE024

______PM10_no imputed values

No imputation forPM10at station DEHE024
because more than 10% of the values missing

______PM25_no imputed values

No imputation forPM25at station DEHE024
because more than 10% of the values missing

______SO2_no imputed values

No imputation forSO2at station DEHE024
because more than 10% of the values missing

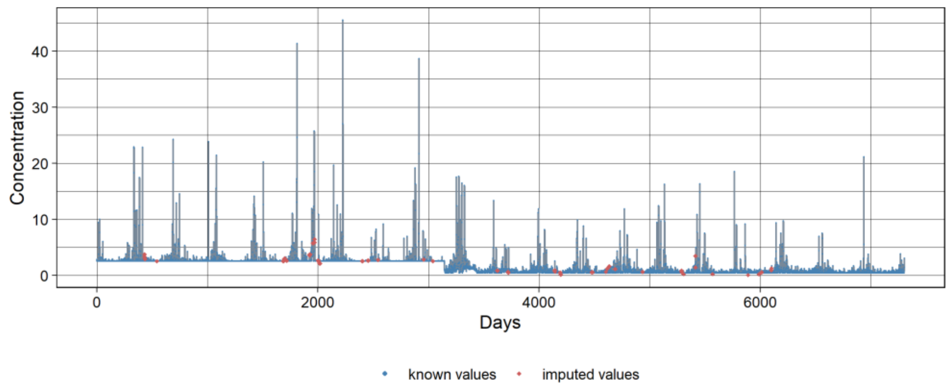
Imputed ggplots Station DEHE026

______CO_no imputed values

No imputation forCOat station DEHE026
because more than 10% of the values missing

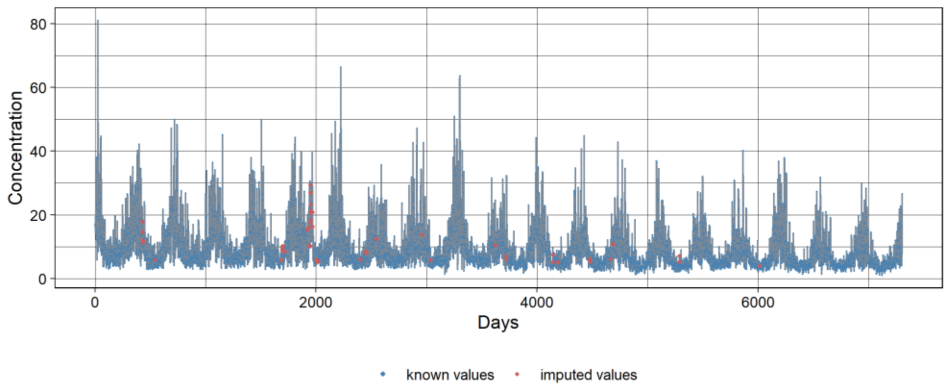
NO_imputed values

Visualization of missing value replacements



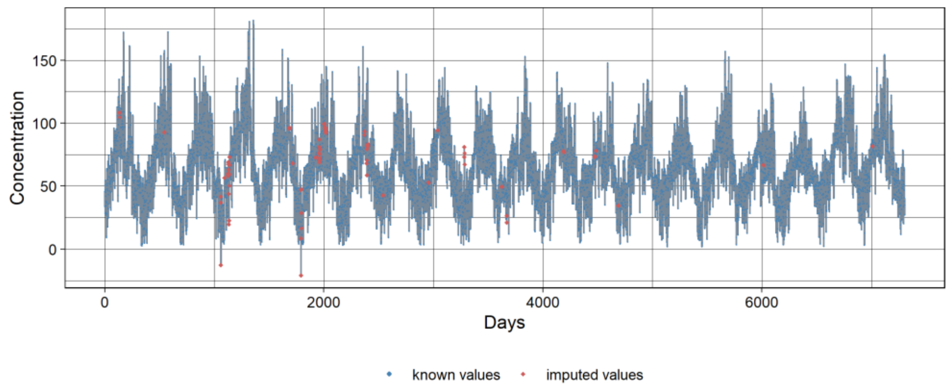
NO2_imputed values

Visualization of missing value replacements



O3_imputed values

Visualization of missing value replacements



Imputed ggplots Station DEHE026

______PM10_no imputed values

No imputation forPM10at station DEHE026
because more than 10% of the values missing

______PM25_no imputed values

No imputation forPM25at station DEHE026
because more than 10% of the values missing

______SO2_no imputed values

No imputation forSO2at station DEHE026
because more than 10% of the values missing

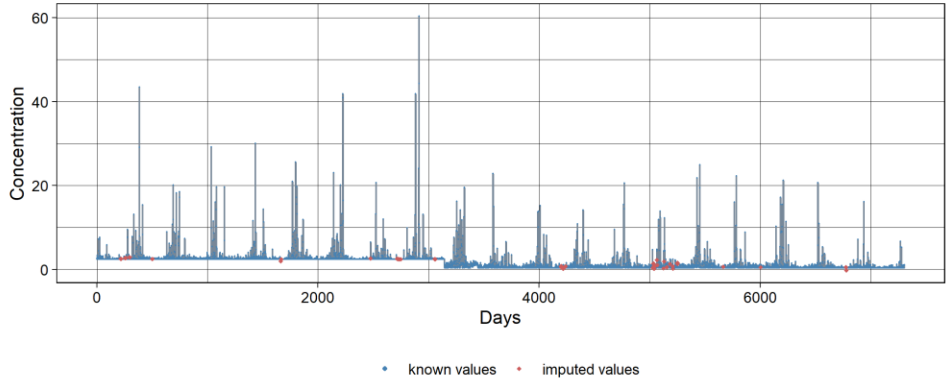
Imputed ggplots Station DEHE028

______CO_no imputed values

No imputation forCOat station DEHE028
because more than 10% of the values missing

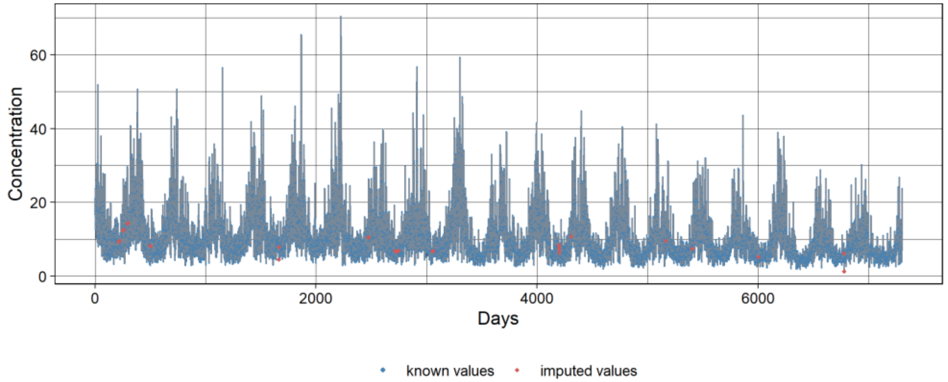
NO_imputed values

Visualization of missing value replacements



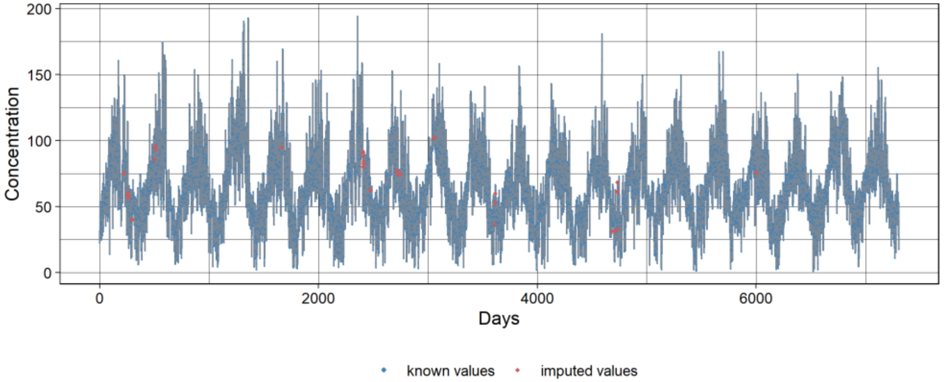
NO2_imputed values

Visualization of missing value replacements



O3_imputed values

Visualization of missing value replacements



Imputed ggplots Station DEHE028

_____PM10_no imputed values

No imputation forPM10at station DEHE028
because more than 10% of the values missing

_____PM25_no imputed values

No imputation forPM25at station DEHE028
because more than 10% of the values missing

_____SO2_no imputed values

No imputation forSO2at station DEHE028
because more than 10% of the values missing

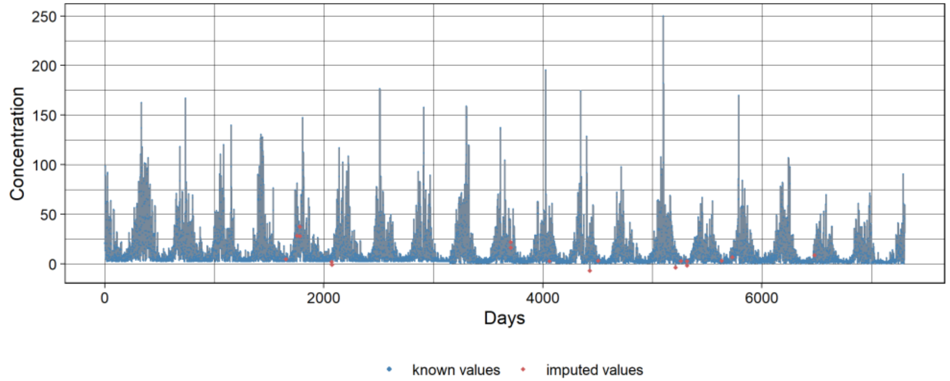
Imputed ggplots Station DEHE030

______CO_no imputed values

No imputation forCOat station DEHE030
because more than 10% of the values missing

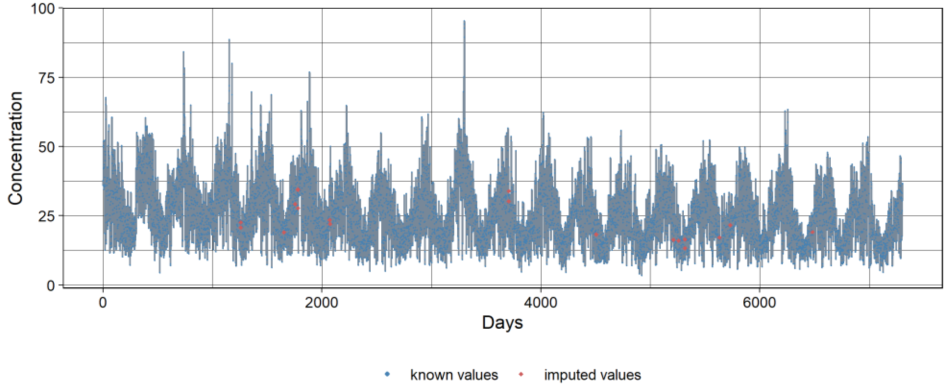
NO_imputed values

Visualization of missing value replacements



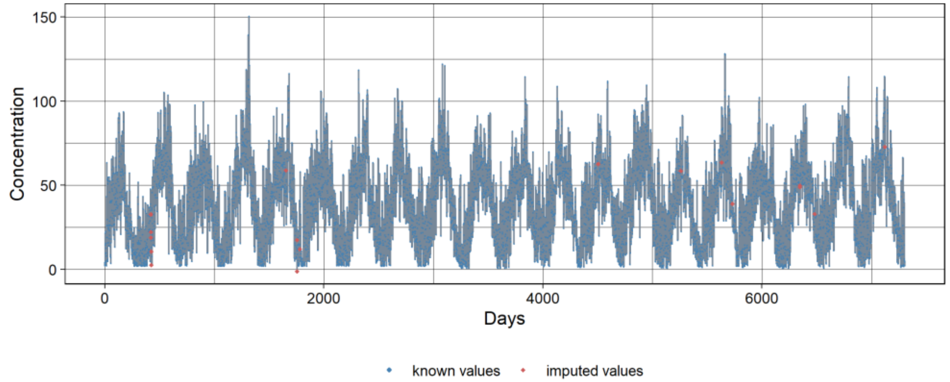
NO2_imputed values

Visualization of missing value replacements



O3_imputed values

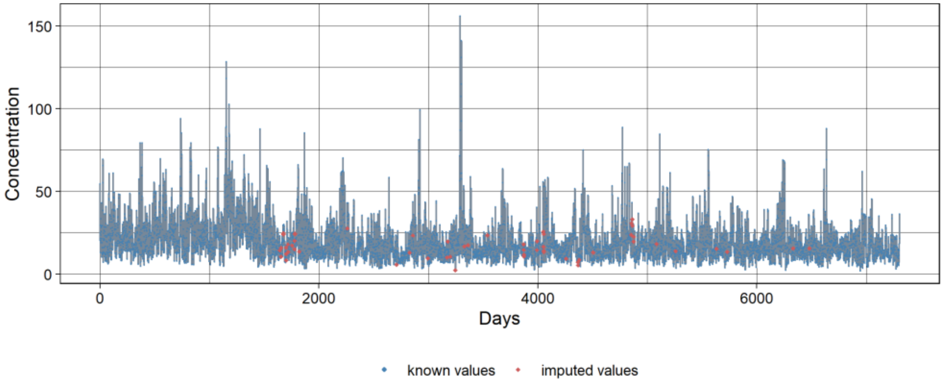
Visualization of missing value replacements



Imputed ggplots Station DEHE030

PM10_imputed values

Visualization of missing value replacements



_____PM25_no imputed values

No imputation forPM25at station DEHE030
because more than 10% of the values missing

_____SO2_no imputed values

No imputation forSO2at station DEHE030
because more than 10% of the values missing

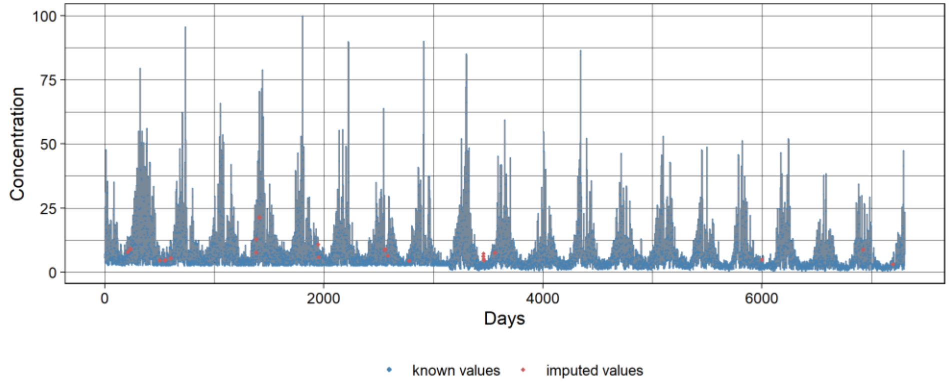
Imputed ggplots Station DEHE032

_____CO_no imputed values

No imputation forCOat station DEHE032
because more than 10% of the values missing

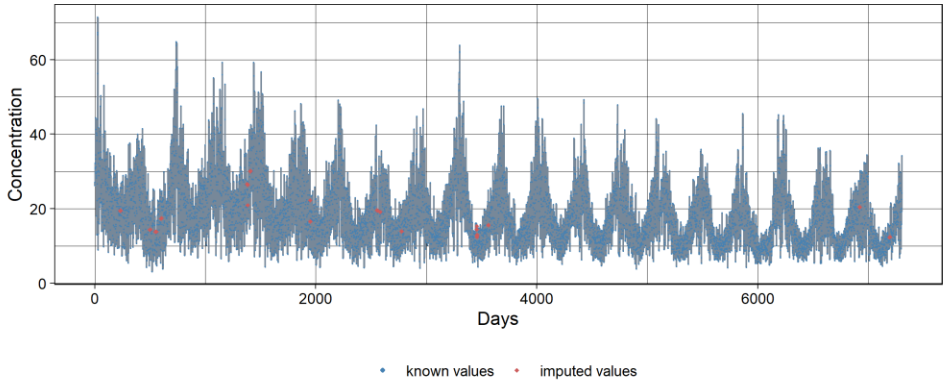
NO_imputed values

Visualization of missing value replacements



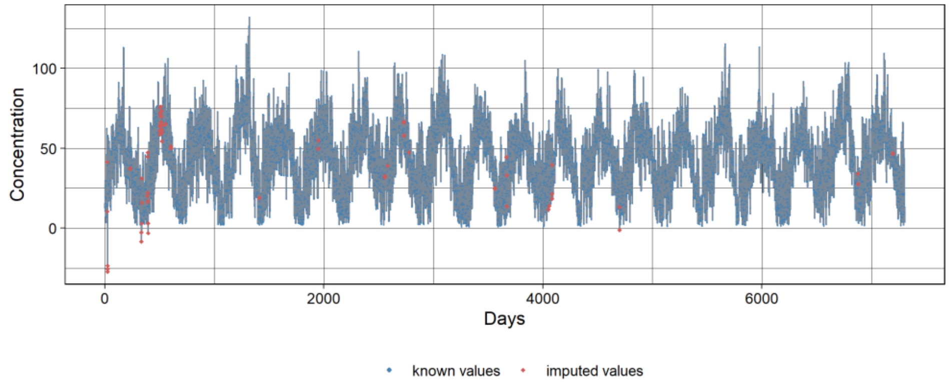
NO2_imputed values

Visualization of missing value replacements



O3_imputed values

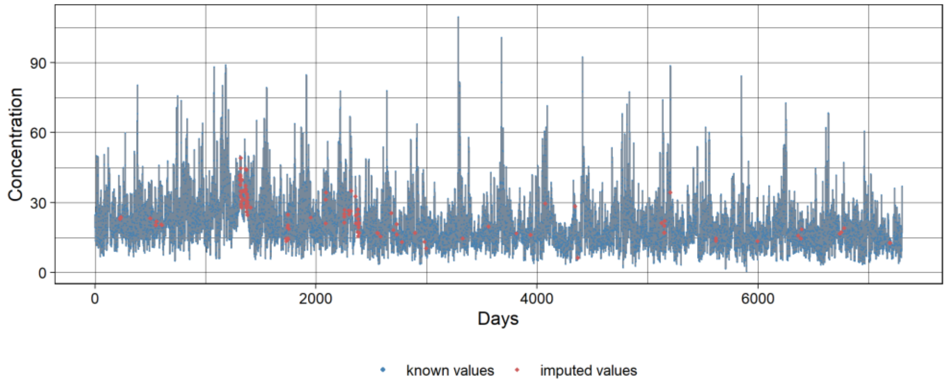
Visualization of missing value replacements



Imputed ggplots Station DEHE032

PM10_imputed values

Visualization of missing value replacements



_____PM25_no imputed values

No imputation forPM25at station DEHE032
because more than 10% of the values missing

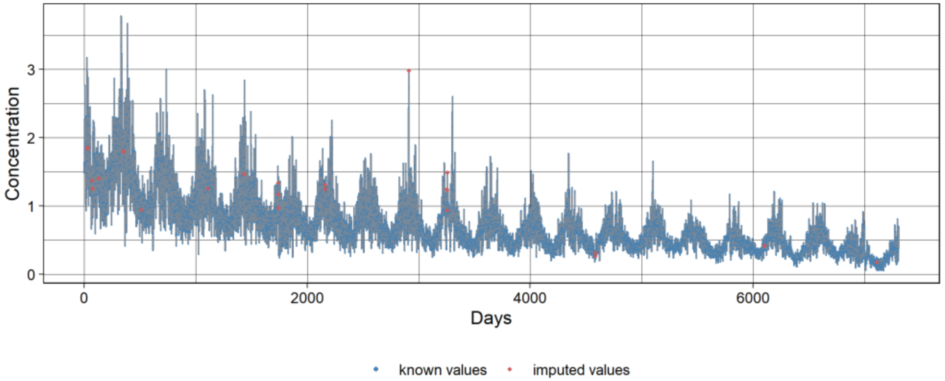
_____SO2_no imputed values

No imputation forSO2at station DEHE032
because more than 10% of the values missing

Imputed ggplots Station DEHE037

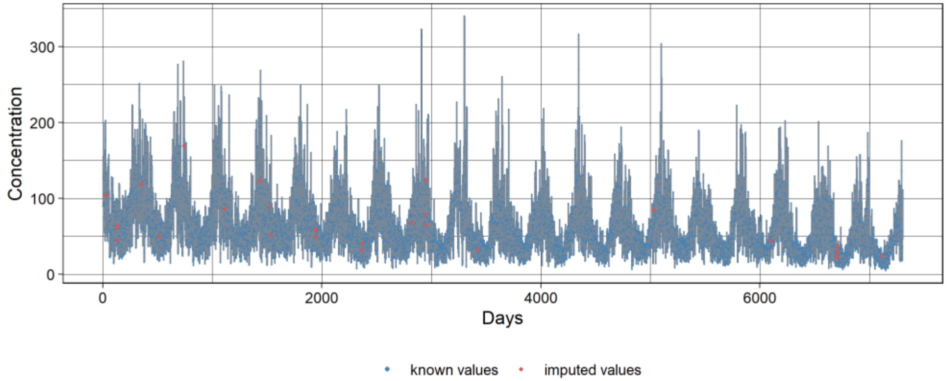
CO_imputed values

Visualization of missing value replacements



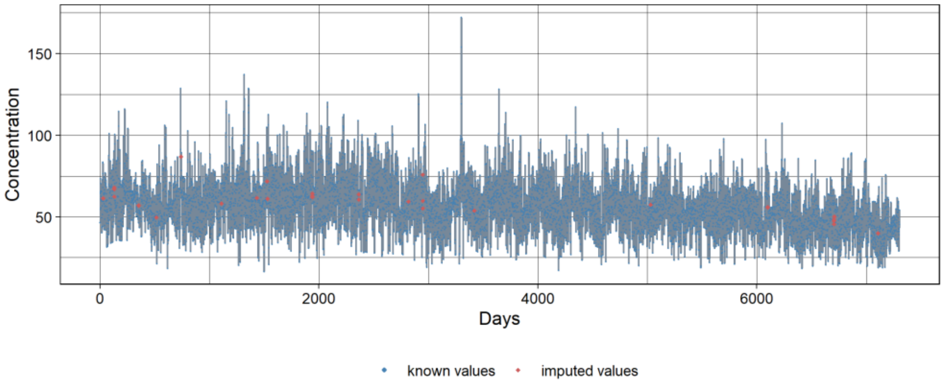
NO_imputed values

Visualization of missing value replacements



NO2_imputed values

Visualization of missing value replacements



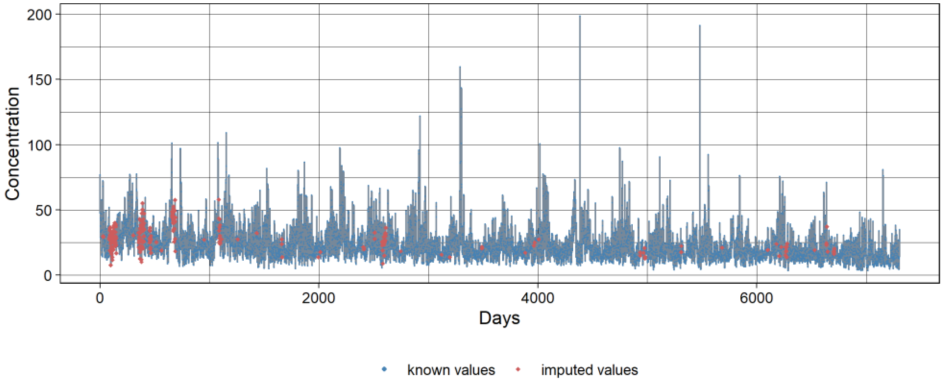
______O3_no imputed values

No imputation forO3at station DEHE037
because more than 10% of the values missing

Imputed ggplots Station DEHE037

PM10_imputed values

Visualization of missing value replacements



______PM25_no imputed values

No imputation forPM25at station DEHE037
because more than 10% of the values missing

______SO2_no imputed values

No imputation forSO2at station DEHE037
because more than 10% of the values missing

Imputed ggplots Station DEHE039

______CO_no imputed values

No imputation for this airpollutant at station DEHE039
because more than 10% of the values missing

______NO_no imputed values

No imputation for this airpollutant at station DEHE039
because more than 10% of the values missing

______NO2_no imputed values

No imputation for this airpollutant at station DEHE039
because more than 10% of the values missing

______O3_no imputed values

No imputation for this airpollutant at station DEHE039
because more than 10% of the values missing

Imputed ggplots Station DEHE039

______PM10_no imputed values

No imputation for this airpollutant at station DEHE039
because more than 10% of the values missing

______PM25_no imputed values

No imputation for this airpollutant at station DEHE039
because more than 10% of the values missing

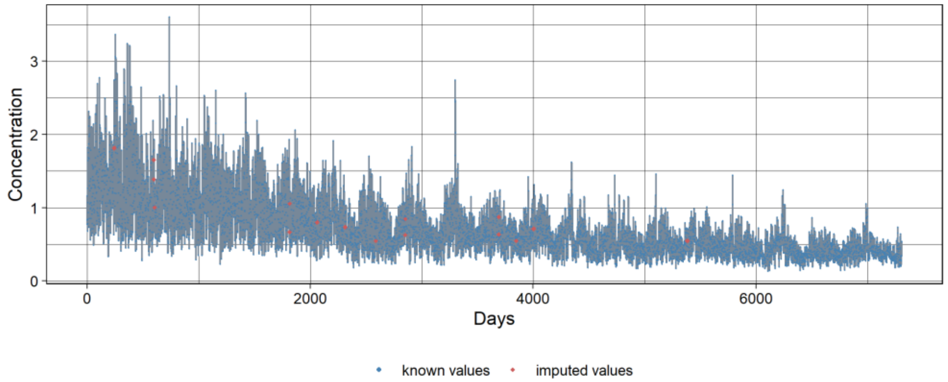
______SO2_no imputed values

No imputation for this airpollutant at station DEHE039
because more than 10% of the values missing

Imputed ggplots Station DEHE040

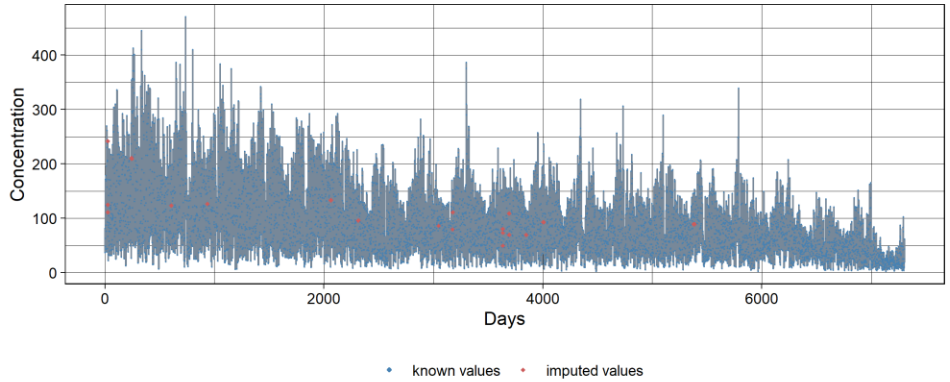
CO_imputed values

Visualization of missing value replacements



NO_imputed values

Visualization of missing value replacements



NO2_imputed values

Visualization of missing value replacements

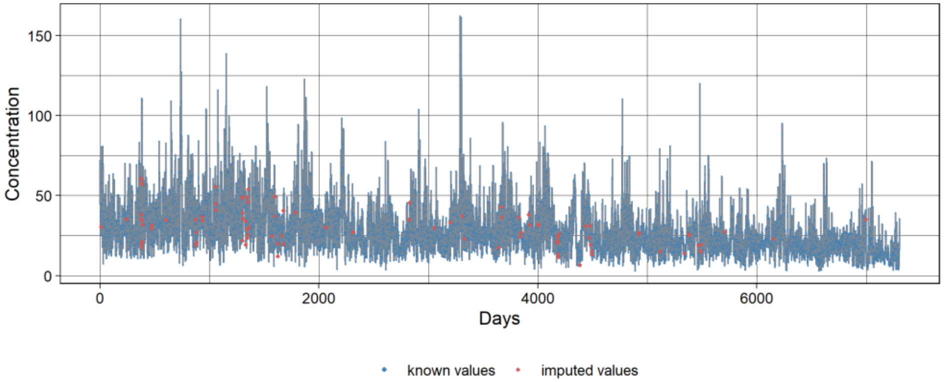
______O3_no imputed values

No imputation forO3at station DEHE040
because more than 10% of the values missing

Imputed ggplots Station DEHE040

PM10_imputed values

Visualization of missing value replacements



_____PM25_no imputed values

No imputation forPM25at station DEHE040
because more than 10% of the values missing

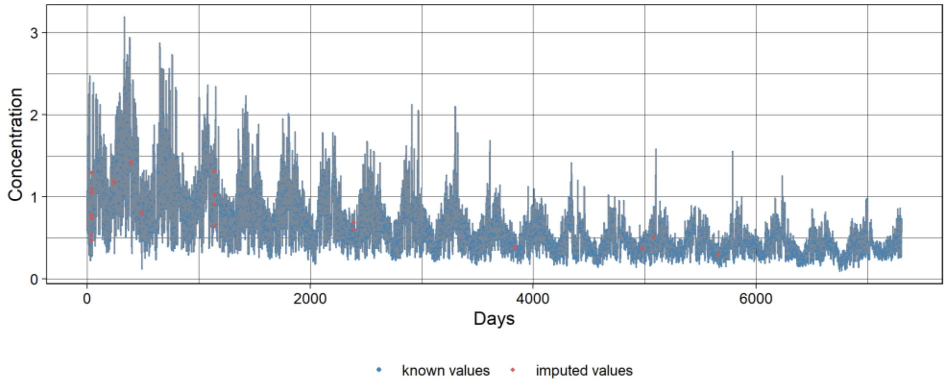
_____SO2_no imputed values

No imputation forSO2at station DEHE040
because more than 10% of the values missing

Imputed ggplots Station DEHE041

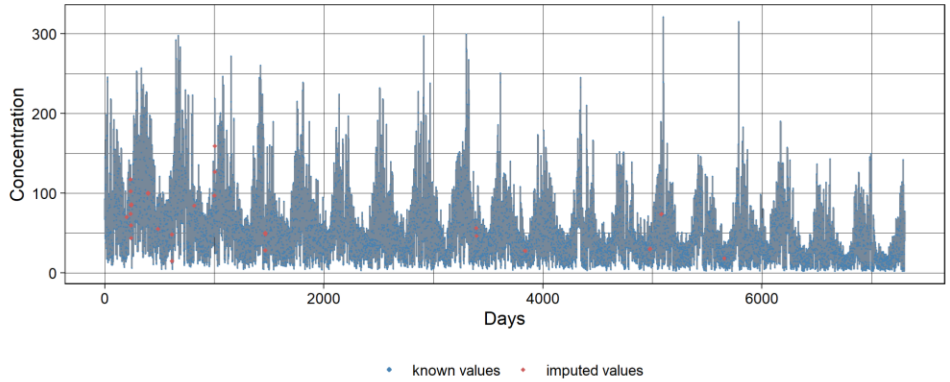
CO_imputed values

Visualization of missing value replacements



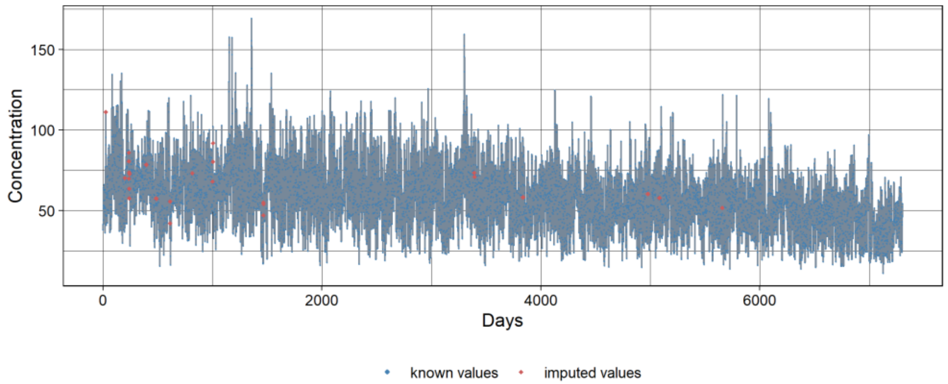
NO_imputed values

Visualization of missing value replacements



NO2_imputed values

Visualization of missing value replacements



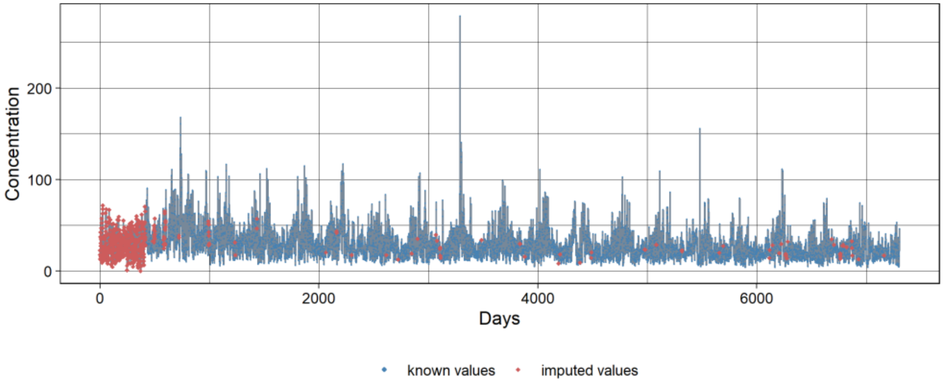
_____O3_no imputed values

No imputation forO3at station DEHE041
because more than 10% of the values missing

Imputed ggplots Station DEHE041

PM10_imputed values

Visualization of missing value replacements



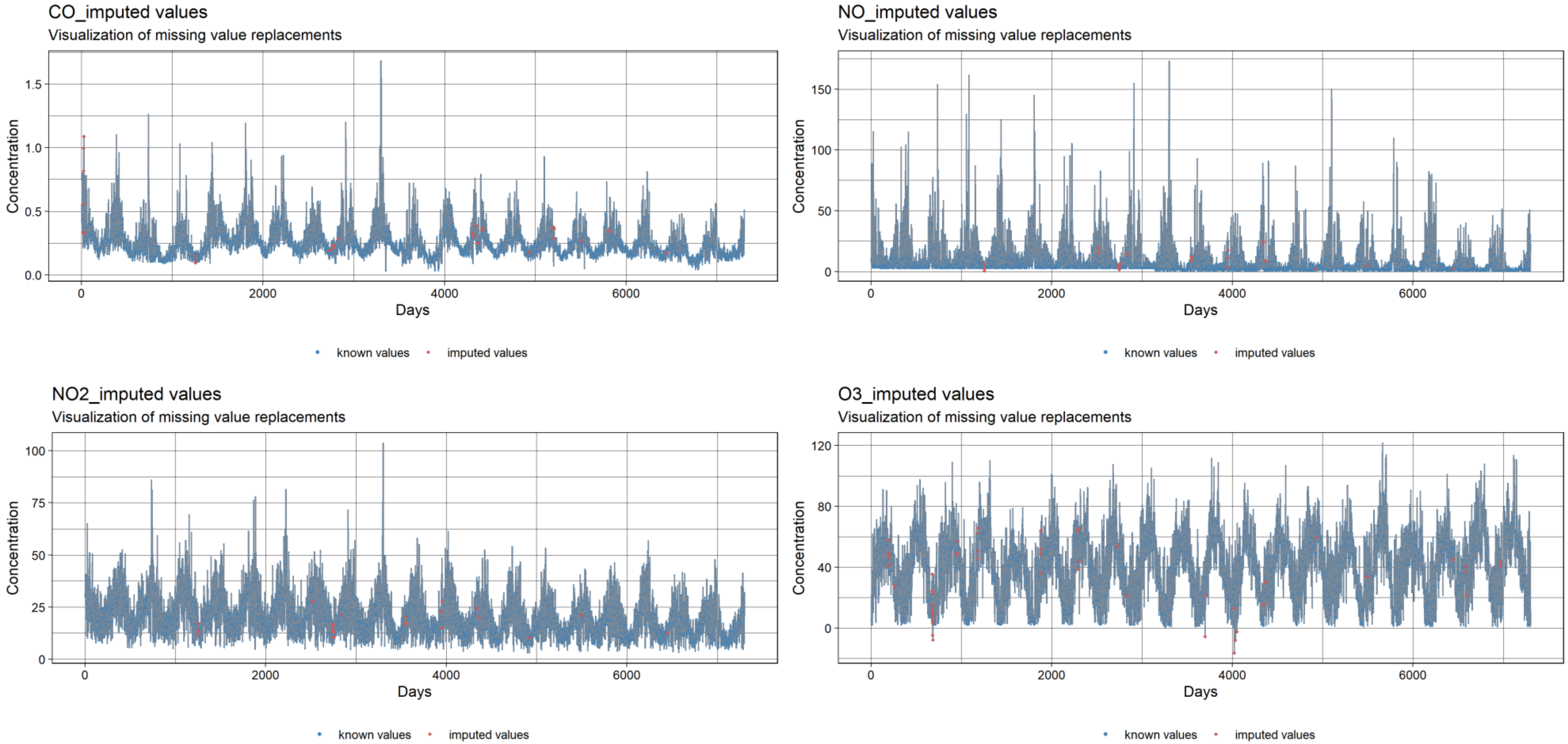
_____PM25_no imputed values

No imputation forPM25at station DEHE041
because more than 10% of the values missing

_____SO2_no imputed values

No imputation forSO2at station DEHE041
because more than 10% of the values missing

Imputed ggplots Station DEHE042



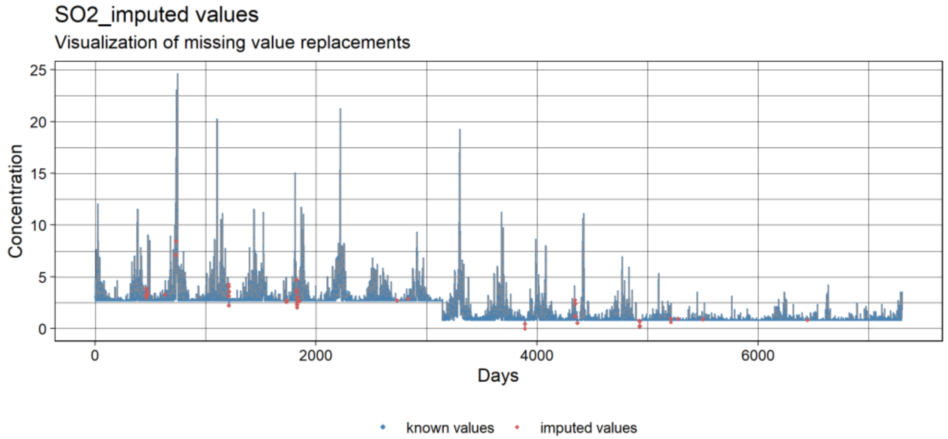
Imputed ggplots Station DEHE042

_____PM10_no imputed values

No imputation forPM10at station DEHE042
because more than 10% of the values missing

_____PM25_no imputed values

No imputation forPM25at station DEHE042
because more than 10% of the values missing



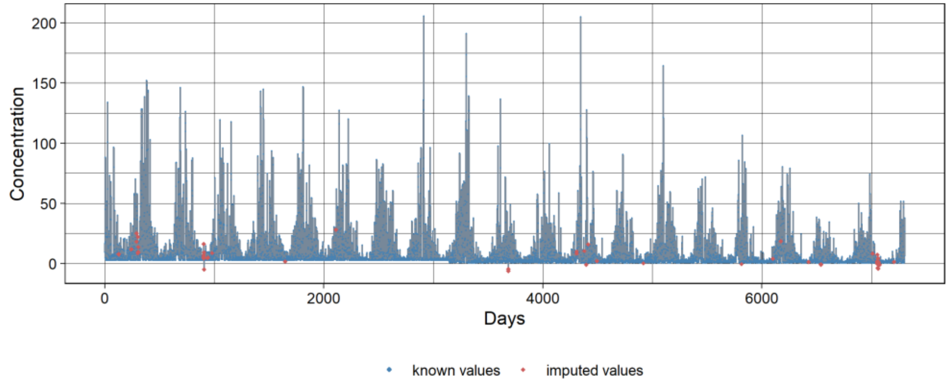
Imputed ggplots Station DEHE043

______CO_no imputed values

No imputation forCOat station DEHE043
because more than 10% of the values missing

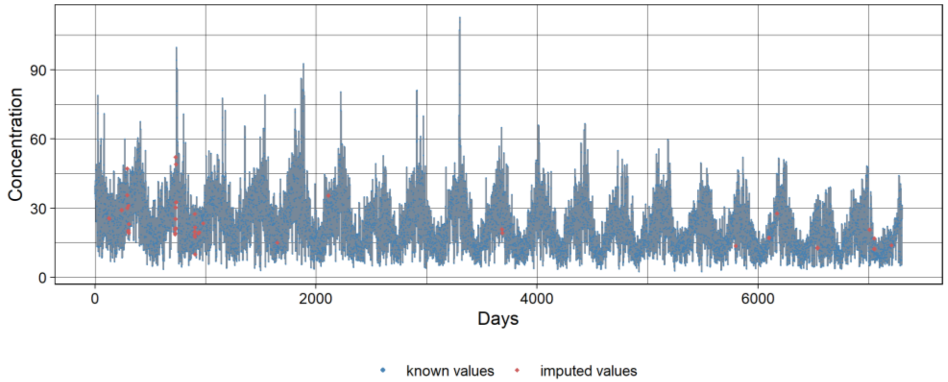
NO_imputed values

Visualization of missing value replacements



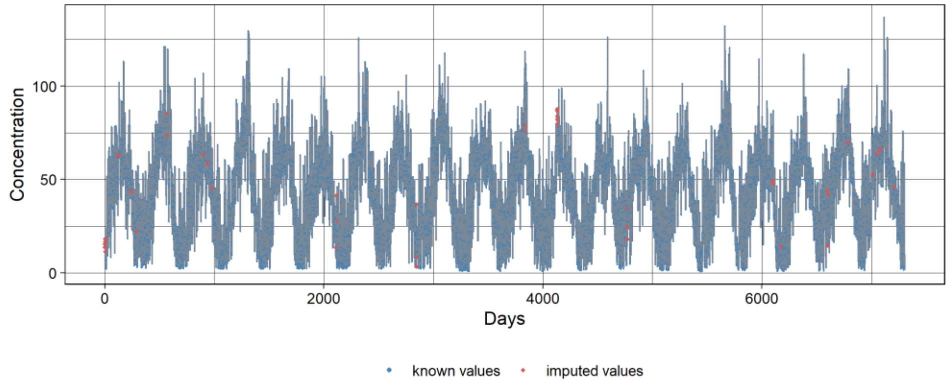
NO2_imputed values

Visualization of missing value replacements



O3_imputed values

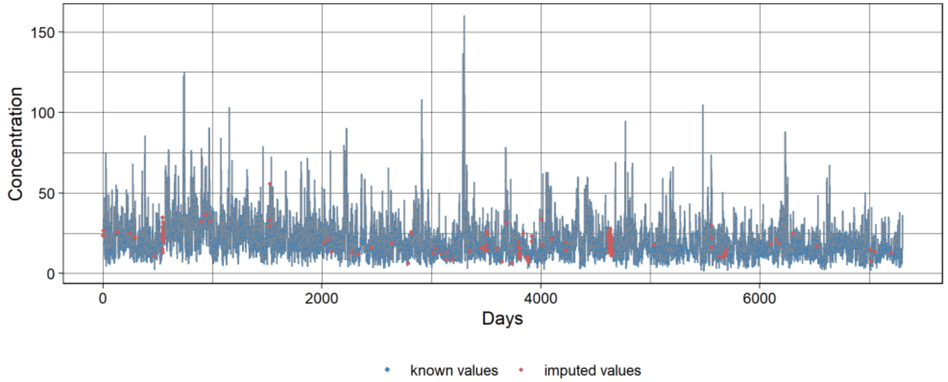
Visualization of missing value replacements



Imputed ggplots Station DEHE043

PM10_imputed values

Visualization of missing value replacements



______PM25_no imputed values

No imputation forPM25at station DEHE043
because more than 10% of the values missing

______SO2_no imputed values

No imputation forSO2at station DEHE043
because more than 10% of the values missing

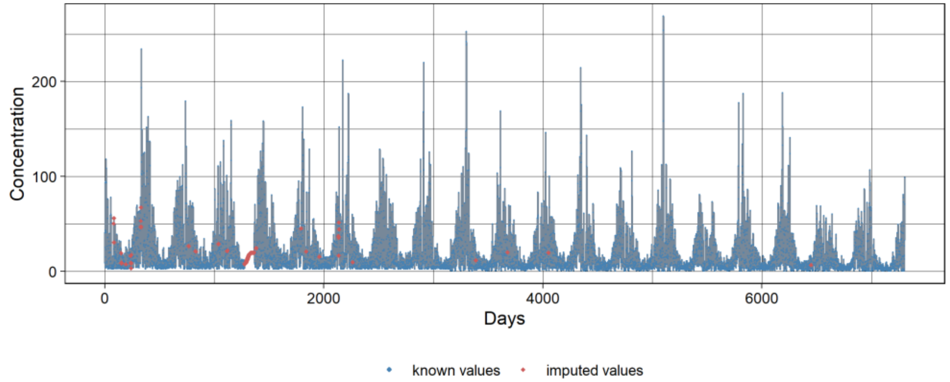
Imputed ggplots Station DEHE044

______CO_no imputed values

No imputation forCOat station DEHE044
because more than 10% of the values missing

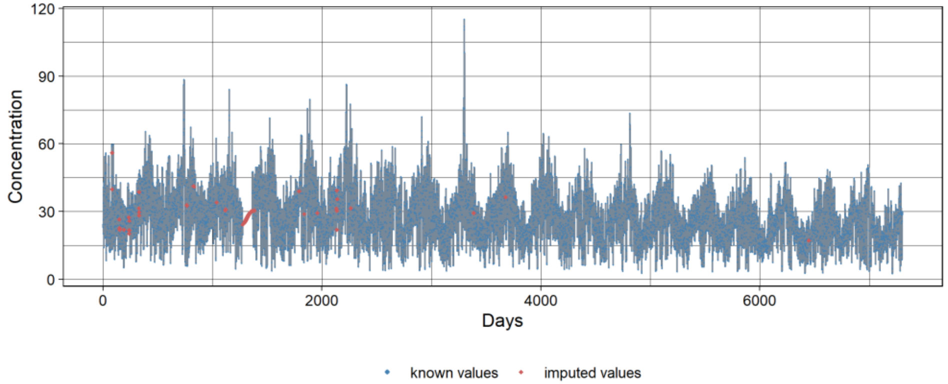
NO_imputed values

Visualization of missing value replacements



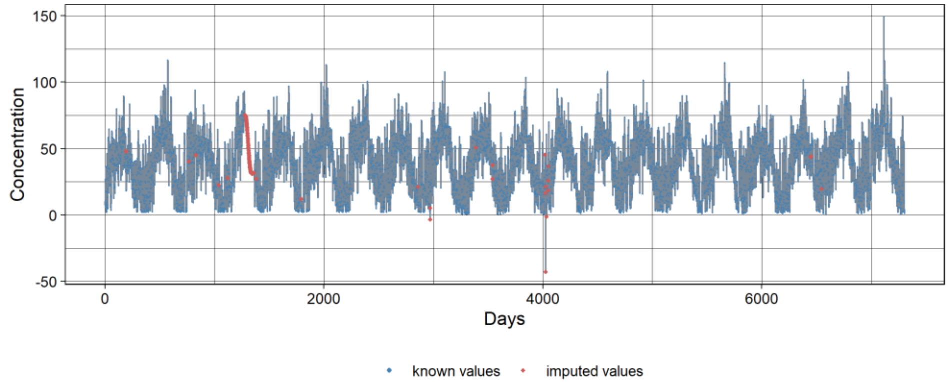
NO2_imputed values

Visualization of missing value replacements



O3_imputed values

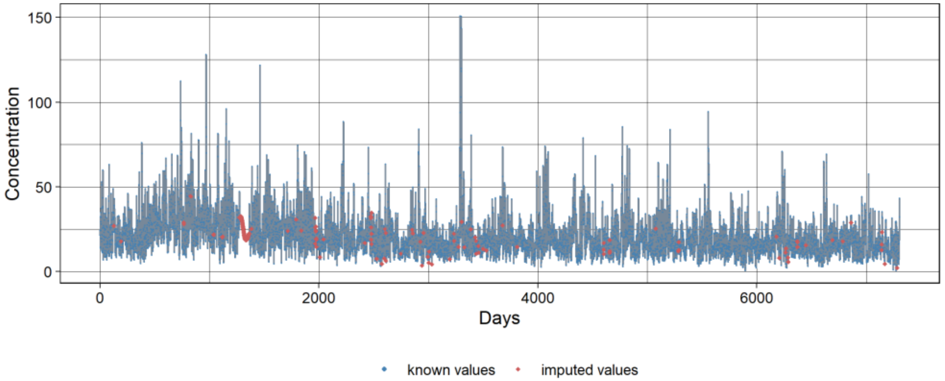
Visualization of missing value replacements



Imputed ggplots Station DEHE044

PM10_imputed values

Visualization of missing value replacements



_____PM25_no imputed values

No imputation forPM25at station DEHE044
because more than 10% of the values missing

_____SO2_no imputed values

No imputation forSO2at station DEHE044
because more than 10% of the values missing

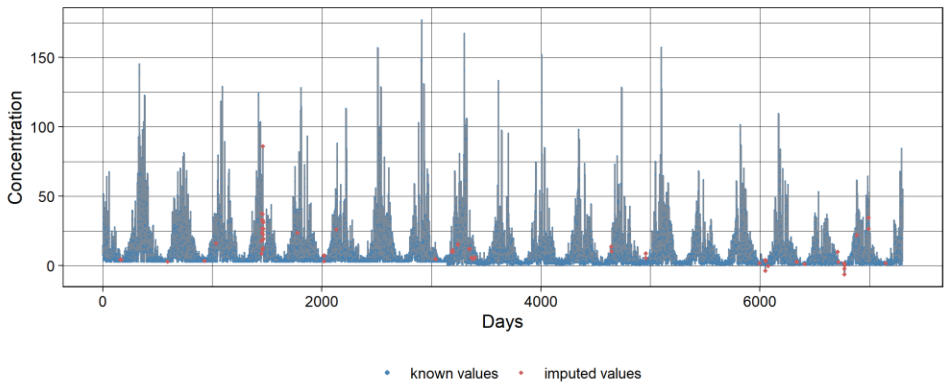
Imputed ggplots Station DEHE045

_____CO_no imputed values

No imputation forCOat station DEHE045
because more than 10% of the values missing

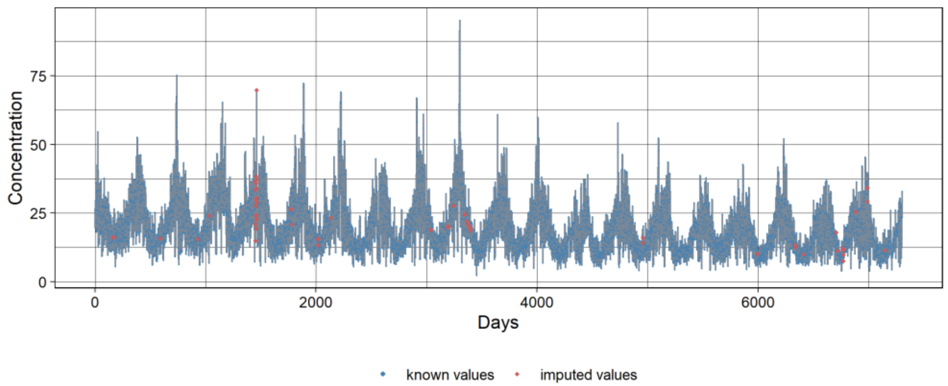
NO_imputed values

Visualization of missing value replacements



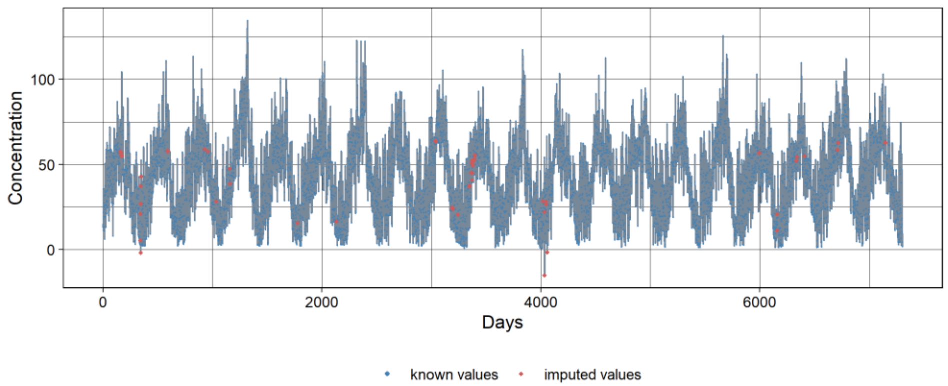
NO2_imputed values

Visualization of missing value replacements



O3_imputed values

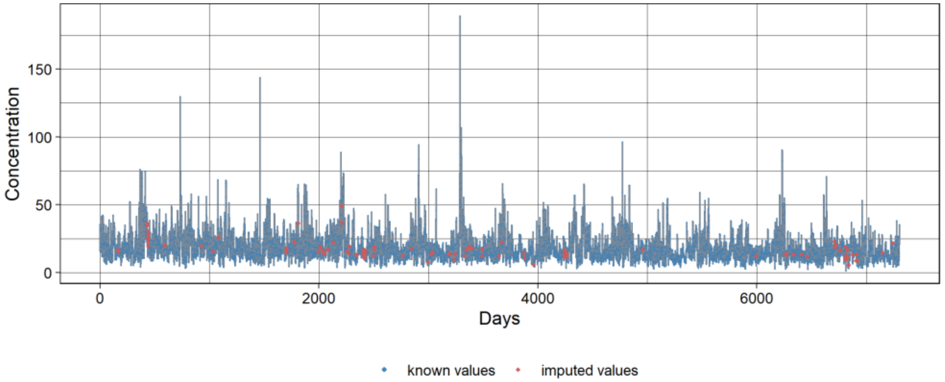
Visualization of missing value replacements



Imputed ggplots Station DEHE045

PM10_imputed values

Visualization of missing value replacements

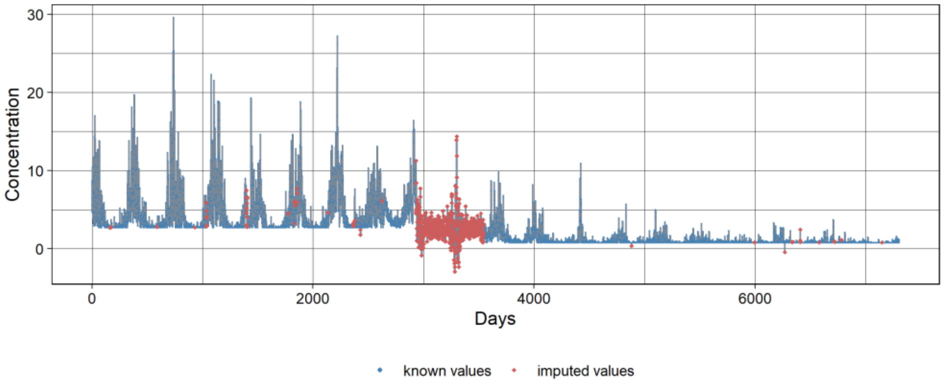


______PM25_no imputed values

No imputation forPM25at station DEHE045
because more than 10% of the values missing

SO2_imputed values

Visualization of missing value replacements



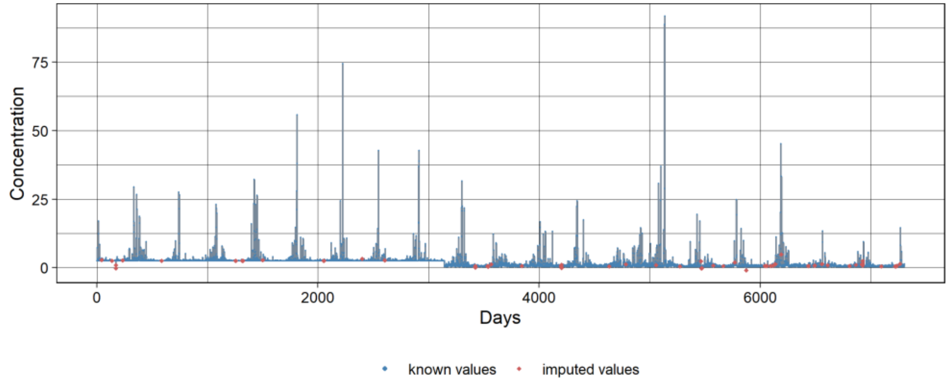
Imputed ggplots Station DEHE046

_____CO_no imputed values

No imputation forCOat station DEHE046
because more than 10% of the values missing

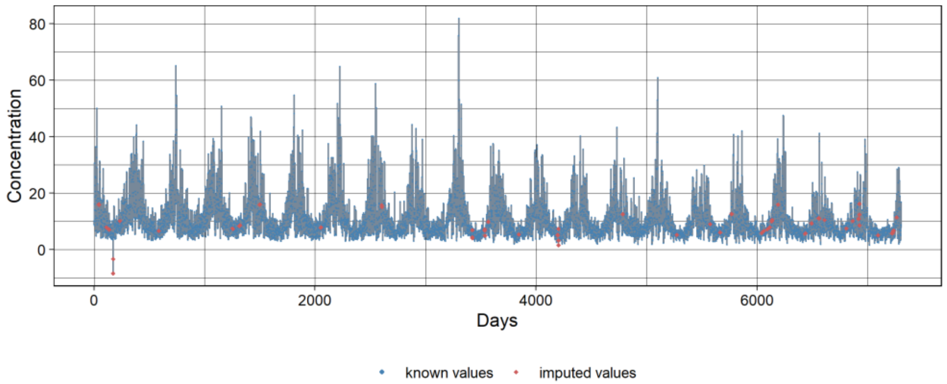
NO_imputed values

Visualization of missing value replacements



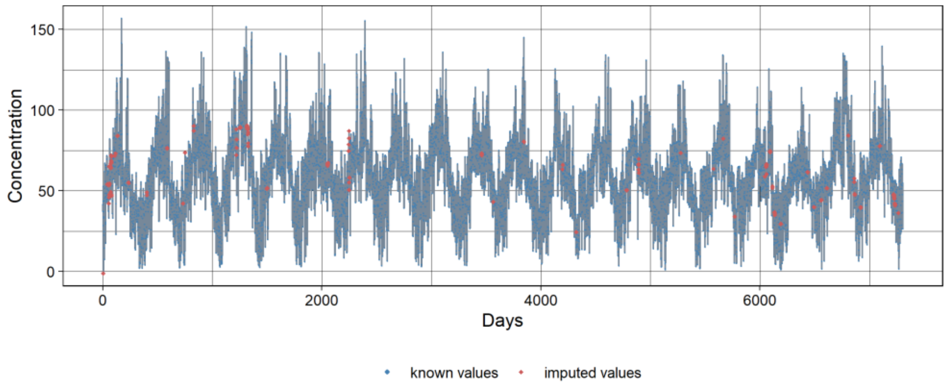
NO2_imputed values

Visualization of missing value replacements



O3_imputed values

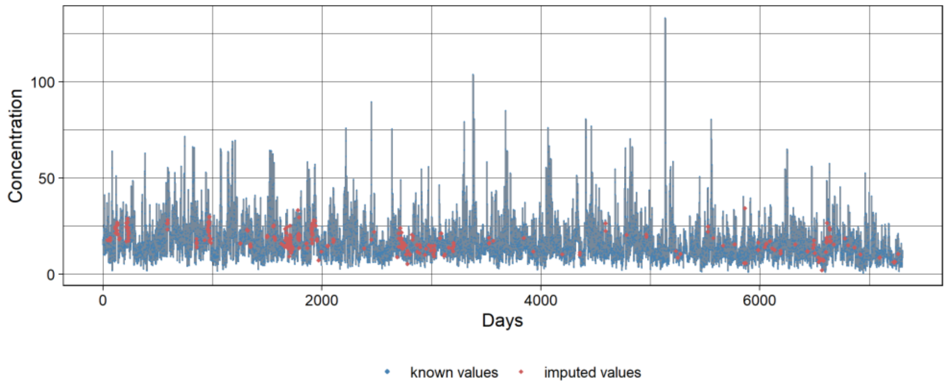
Visualization of missing value replacements



Imputed ggplots Station DEHE046

PM10_imputed values

Visualization of missing value replacements



_____PM25_no imputed values

No imputation forPM25at station DEHE046
because more than 10% of the values missing

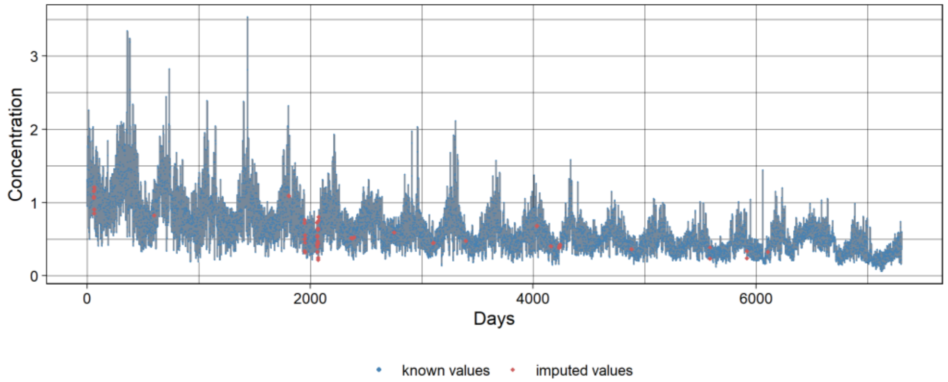
_____SO2_no imputed values

No imputation forSO2at station DEHE046
because more than 10% of the values missing

Imputed ggplots Station DEHE049

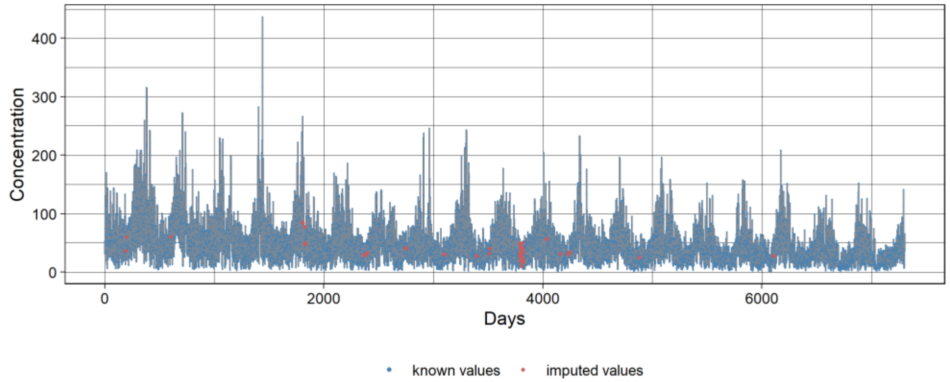
CO_imputed values

Visualization of missing value replacements



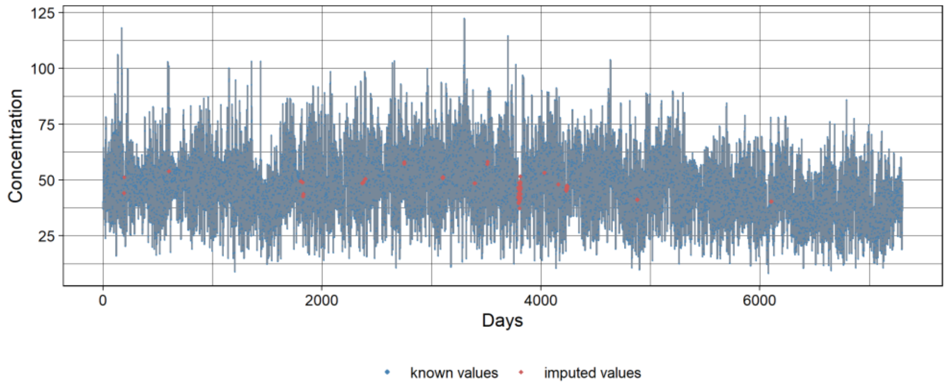
NO_imputed values

Visualization of missing value replacements



NO2_imputed values

Visualization of missing value replacements



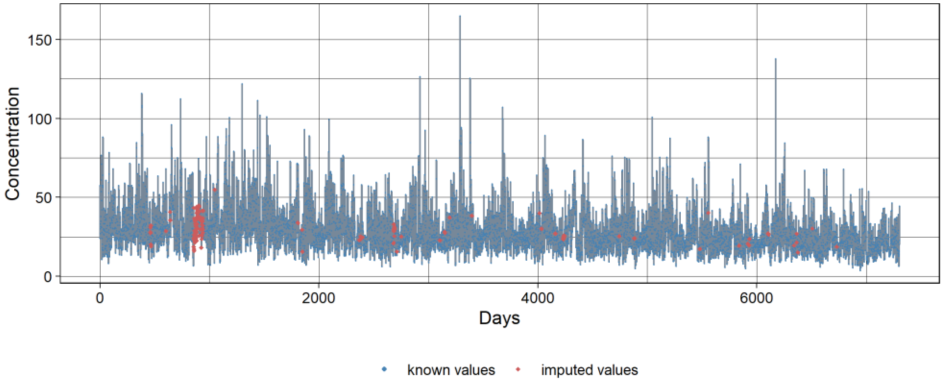
______O3_no imputed values

No imputation forO3at station DEHE049
because more than 10% of the values missing

Imputed ggplots Station DEHE049

PM10_imputed values

Visualization of missing value replacements



_____PM25_no imputed values

No imputation forPM25at station DEHE049
because more than 10% of the values missing

_____SO2_no imputed values

No imputation forSO2at station DEHE049
because more than 10% of the values missing

Imputed ggplots Station DEHE050

______CO_no imputed values

No imputation for this airpollutant at station DEHE050
because more than 10% of the values missing

______NO_no imputed values

No imputation for this airpollutant at station DEHE050
because more than 10% of the values missing

______NO2_no imputed values

No imputation for this airpollutant at station DEHE050
because more than 10% of the values missing

______O3_no imputed values

No imputation for this airpollutant at station DEHE050
because more than 10% of the values missing

Imputed ggplots Station DEHE050

______PM10_no imputed values

No imputation for this airpollutant at station DEHE050
because more than 10% of the values missing

______PM25_no imputed values

No imputation for this airpollutant at station DEHE050
because more than 10% of the values missing

______SO2_no imputed values

No imputation for this airpollutant at station DEHE050
because more than 10% of the values missing

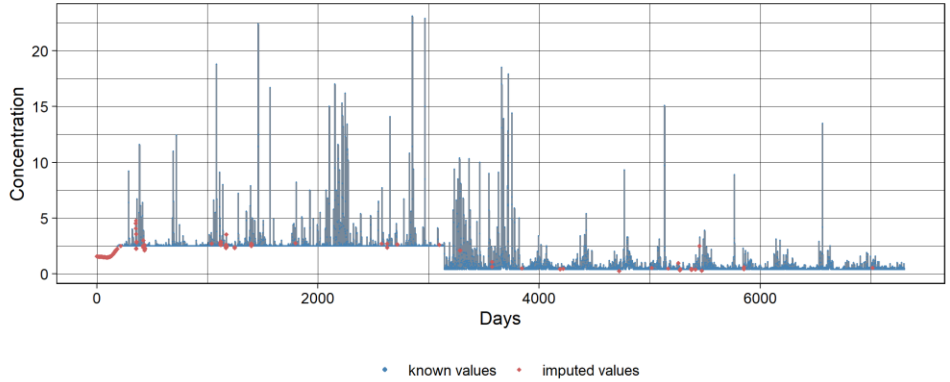
Imputed ggplots Station DEHE051

______CO_no imputed values

No imputation forCOat station DEHE051
because more than 10% of the values missing

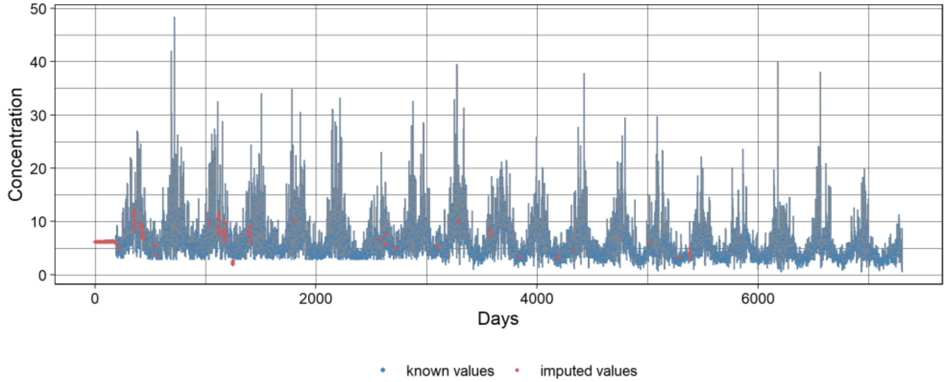
NO_imputed values

Visualization of missing value replacements



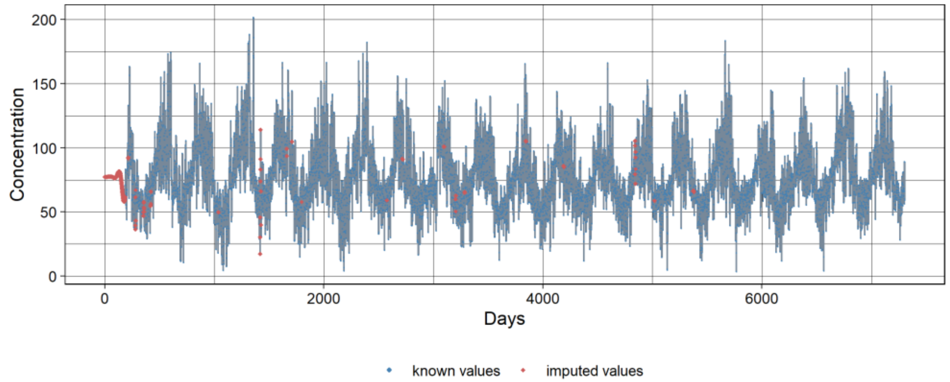
NO2_imputed values

Visualization of missing value replacements



O3_imputed values

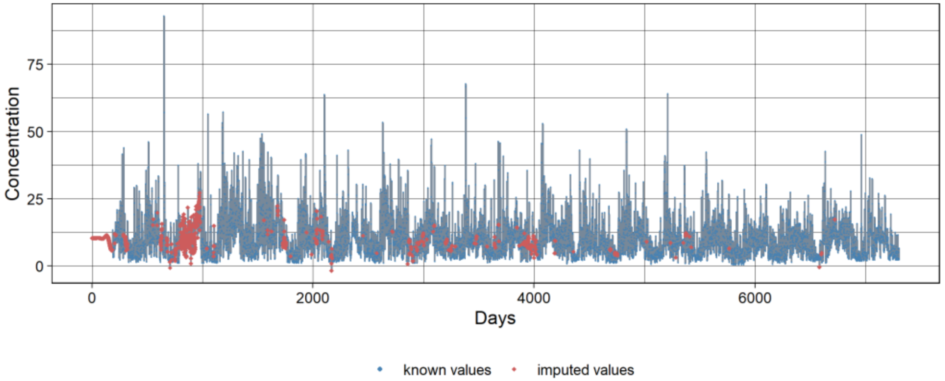
Visualization of missing value replacements



Imputed ggplots Station DEHE051

PM10_imputed values

Visualization of missing value replacements

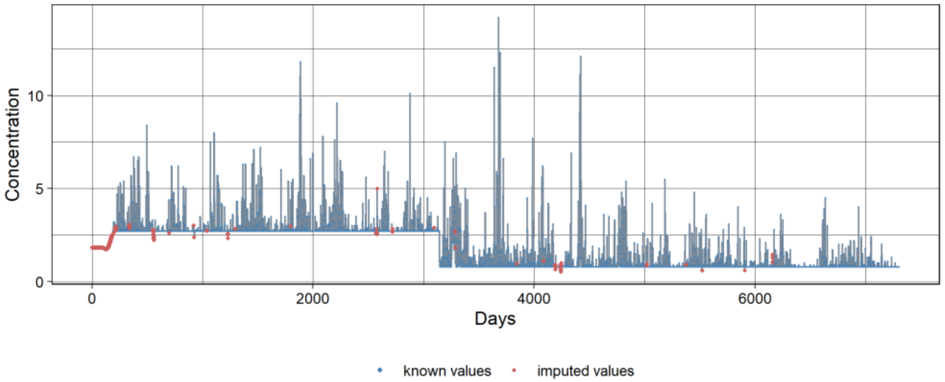


______PM25_no imputed values

No imputation forPM25at station DEHE051
because more than 10% of the values missing

SO2_imputed values

Visualization of missing value replacements



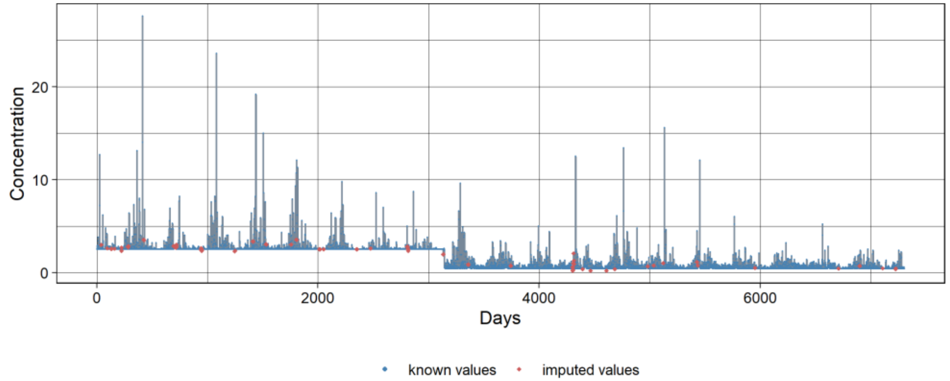
Imputed ggplots Station DEHE052

______CO_no imputed values

No imputation forCOat station DEHE052
because more than 10% of the values missing

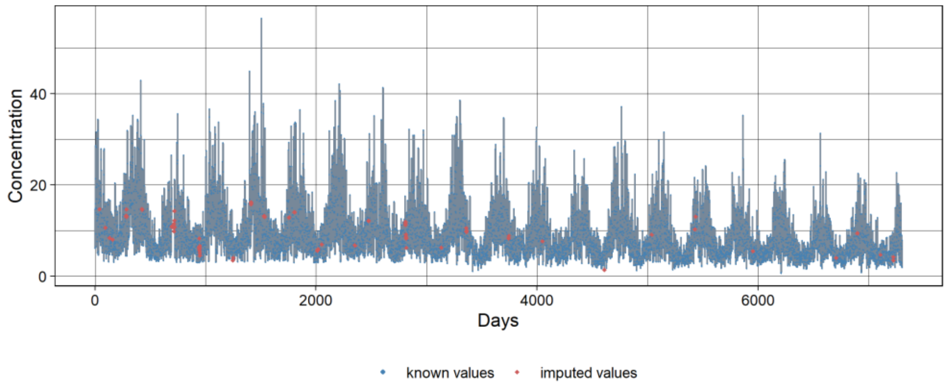
NO_imputed values

Visualization of missing value replacements



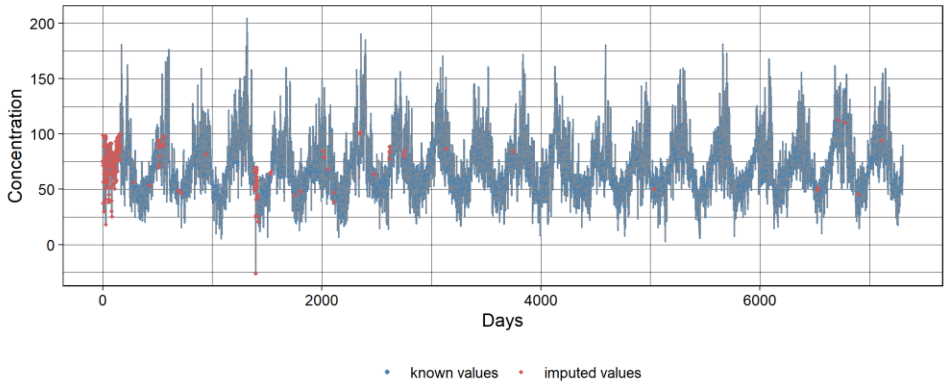
NO2_imputed values

Visualization of missing value replacements



O3_imputed values

Visualization of missing value replacements



Imputed ggplots Station DEHE052

______PM10_no imputed values

No imputation forPM10at station DEHE052
because more than 10% of the values missing

______PM25_no imputed values

No imputation forPM25at station DEHE052
because more than 10% of the values missing

______SO2_no imputed values

No imputation forSO2at station DEHE052
because more than 10% of the values missing

Imputed ggplots Station DEHE059

______CO_no imputed values

No imputation for this airpollutant at station DEHE059
because more than 10% of the values missing

______NO_no imputed values

No imputation for this airpollutant at station DEHE059
because more than 10% of the values missing

______NO2_no imputed values

No imputation for this airpollutant at station DEHE059
because more than 10% of the values missing

______O3_no imputed values

No imputation for this airpollutant at station DEHE059
because more than 10% of the values missing

Imputed ggplots Station DEHE059

______PM10_no imputed values

No imputation for this airpollutant at station DEHE059
because more than 10% of the values missing

______PM25_no imputed values

No imputation for this airpollutant at station DEHE059
because more than 10% of the values missing

______SO2_no imputed values

No imputation for this airpollutant at station DEHE059
because more than 10% of the values missing

Imputed ggplots Station DEHE060

______CO_no imputed values

No imputation for this airpollutant at station DEHE060
because more than 10% of the values missing

______NO_no imputed values

No imputation for this airpollutant at station DEHE060
because more than 10% of the values missing

______NO2_no imputed values

No imputation for this airpollutant at station DEHE060
because more than 10% of the values missing

______O3_no imputed values

No imputation for this airpollutant at station DEHE060
because more than 10% of the values missing

Imputed ggplots Station DEHE060

_____PM10_no imputed values

No imputation for this airpollutant at station DEHE060
because more than 10% of the values missing

_____PM25_no imputed values

No imputation for this airpollutant at station DEHE060
because more than 10% of the values missing

_____SO2_no imputed values

No imputation for this airpollutant at station DEHE060
because more than 10% of the values missing

Imputed ggplots Station DEHE061

______CO_no imputed values

No imputation for this airpollutant at station DEHE061
because more than 10% of the values missing

______NO_no imputed values

No imputation for this airpollutant at station DEHE061
because more than 10% of the values missing

______NO2_no imputed values

No imputation for this airpollutant at station DEHE061
because more than 10% of the values missing

______O3_no imputed values

No imputation for this airpollutant at station DEHE061
because more than 10% of the values missing

Imputed ggplots Station DEHE061

______PM10_no imputed values

No imputation for this airpollutant at station DEHE061
because more than 10% of the values missing

______PM25_no imputed values

No imputation for this airpollutant at station DEHE061
because more than 10% of the values missing

______SO2_no imputed values

No imputation for this airpollutant at station DEHE061
because more than 10% of the values missing

Imputed ggplots Station DEHE062

______CO_no imputed values

No imputation for this airpollutant at station DEHE062
because more than 10% of the values missing

______NO_no imputed values

No imputation for this airpollutant at station DEHE062
because more than 10% of the values missing

______NO2_no imputed values

No imputation for this airpollutant at station DEHE062
because more than 10% of the values missing

______O3_no imputed values

No imputation for this airpollutant at station DEHE062
because more than 10% of the values missing

Imputed ggplots Station DEHE062

______PM10_no imputed values

No imputation for this airpollutant at station DEHE062
because more than 10% of the values missing

______PM25_no imputed values

No imputation for this airpollutant at station DEHE062
because more than 10% of the values missing

______SO2_no imputed values

No imputation for this airpollutant at station DEHE062
because more than 10% of the values missing

Imputed ggplots Station DEHE063

_____CO_no imputed values

No imputation for this airpollutant at station DEHE063
because more than 10% of the values missing

_____NO_no imputed values

No imputation for this airpollutant at station DEHE063
because more than 10% of the values missing

_____NO2_no imputed values

No imputation for this airpollutant at station DEHE063
because more than 10% of the values missing

_____O3_no imputed values

No imputation for this airpollutant at station DEHE063
because more than 10% of the values missing

Imputed ggplots Station DEHE063

______PM10_no imputed values

No imputation for this airpollutant at station DEHE063
because more than 10% of the values missing

______PM25_no imputed values

No imputation for this airpollutant at station DEHE063
because more than 10% of the values missing

______SO2_no imputed values

No imputation for this airpollutant at station DEHE063
because more than 10% of the values missing

Imputed ggplots Station DEHE112

_____CO_no imputed values

No imputation for this airpollutant at station DEHE112
because more than 10% of the values missing

_____NO_no imputed values

No imputation for this airpollutant at station DEHE112
because more than 10% of the values missing

_____NO2_no imputed values

No imputation for this airpollutant at station DEHE112
because more than 10% of the values missing

_____O3_no imputed values

No imputation for this airpollutant at station DEHE112
because more than 10% of the values missing

Imputed ggplots Station DEHE112

______PM10_no imputed values

No imputation for this airpollutant at station DEHE112
because more than 10% of the values missing

______PM25_no imputed values

No imputation for this airpollutant at station DEHE112
because more than 10% of the values missing

______SO2_no imputed values

No imputation for this airpollutant at station DEHE112
because more than 10% of the values missing

Imputed ggplots Station DEHE116

_____CO_no imputed values

No imputation for this airpollutant at station DEHE116
because more than 10% of the values missing

_____NO_no imputed values

No imputation for this airpollutant at station DEHE116
because more than 10% of the values missing

_____NO2_no imputed values

No imputation for this airpollutant at station DEHE116
because more than 10% of the values missing

_____O3_no imputed values

No imputation for this airpollutant at station DEHE116
because more than 10% of the values missing

Imputed ggplots Station DEHE116

______PM10_no imputed values

No imputation for this airpollutant at station DEHE116
because more than 10% of the values missing

______PM25_no imputed values

No imputation for this airpollutant at station DEHE116
because more than 10% of the values missing

______SO2_no imputed values

No imputation for this airpollutant at station DEHE116
because more than 10% of the values missing

Imputed ggplots Station DEHE131

______CO_no imputed values

No imputation for this airpollutant at station DEHE131
because more than 10% of the values missing

______NO_no imputed values

No imputation for this airpollutant at station DEHE131
because more than 10% of the values missing

______NO2_no imputed values

No imputation for this airpollutant at station DEHE131
because more than 10% of the values missing

______O3_no imputed values

No imputation for this airpollutant at station DEHE131
because more than 10% of the values missing

Imputed ggplots Station DEHE131

______PM10_no imputed values

No imputation for this airpollutant at station DEHE131
because more than 10% of the values missing

______PM25_no imputed values

No imputation for this airpollutant at station DEHE131
because more than 10% of the values missing

______SO2_no imputed values

No imputation for this airpollutant at station DEHE131
because more than 10% of the values missing

Imputed ggplots Station DEHE134

______CO_no imputed values

No imputation for this airpollutant at station DEHE134
because more than 10% of the values missing

______NO_no imputed values

No imputation for this airpollutant at station DEHE134
because more than 10% of the values missing

______NO2_no imputed values

No imputation for this airpollutant at station DEHE134
because more than 10% of the values missing

______O3_no imputed values

No imputation for this airpollutant at station DEHE134
because more than 10% of the values missing

Imputed ggplots Station DEHE134

_____PM10_no imputed values

No imputation for this airpollutant at station DEHE134
because more than 10% of the values missing

_____PM25_no imputed values

No imputation for this airpollutant at station DEHE134
because more than 10% of the values missing

_____SO2_no imputed values

No imputation for this airpollutant at station DEHE134
because more than 10% of the values missing

Imputed ggplots Station DEHE135

______CO_no imputed values

No imputation for this airpollutant at station DEHE135
because more than 10% of the values missing

______NO_no imputed values

No imputation for this airpollutant at station DEHE135
because more than 10% of the values missing

______NO2_no imputed values

No imputation for this airpollutant at station DEHE135
because more than 10% of the values missing

______O3_no imputed values

No imputation for this airpollutant at station DEHE135
because more than 10% of the values missing

Imputed ggplots Station DEHE135

______PM10_no imputed values

No imputation for this airpollutant at station DEHE135
because more than 10% of the values missing

______PM25_no imputed values

No imputation for this airpollutant at station DEHE135
because more than 10% of the values missing

______SO2_no imputed values

No imputation for this airpollutant at station DEHE135
because more than 10% of the values missing

Imputed ggplots Station DEHE150

______CO_no imputed values

No imputation for this airpollutant at station DEHE150
because more than 10% of the values missing

______NO_no imputed values

No imputation for this airpollutant at station DEHE150
because more than 10% of the values missing

______NO2_no imputed values

No imputation for this airpollutant at station DEHE150
because more than 10% of the values missing

______O3_no imputed values

No imputation for this airpollutant at station DEHE150
because more than 10% of the values missing

Imputed ggplots Station DEHE150

______PM10_no imputed values

No imputation for this airpollutant at station DEHE150
because more than 10% of the values missing

______PM25_no imputed values

No imputation for this airpollutant at station DEHE150
because more than 10% of the values missing

______SO2_no imputed values

No imputation for this airpollutant at station DEHE150
because more than 10% of the values missing

Imputed ggplots Station DEHE160

______CO_no imputed values

No imputation for this airpollutant at station DEHE160
because more than 10% of the values missing

______NO_no imputed values

No imputation for this airpollutant at station DEHE160
because more than 10% of the values missing

______NO2_no imputed values

No imputation for this airpollutant at station DEHE160
because more than 10% of the values missing

______O3_no imputed values

No imputation for this airpollutant at station DEHE160
because more than 10% of the values missing

Imputed ggplots Station DEHE160

______PM10_no imputed values

No imputation for this airpollutant at station DEHE160
because more than 10% of the values missing

______PM25_no imputed values

No imputation for this airpollutant at station DEHE160
because more than 10% of the values missing

______SO2_no imputed values

No imputation for this airpollutant at station DEHE160
because more than 10% of the values missing

Imputed ggplots Station DEHE161

______CO_no imputed values

No imputation for this airpollutant at station DEHE161
because more than 10% of the values missing

______NO_no imputed values

No imputation for this airpollutant at station DEHE161
because more than 10% of the values missing

______NO2_no imputed values

No imputation for this airpollutant at station DEHE161
because more than 10% of the values missing

______O3_no imputed values

No imputation for this airpollutant at station DEHE161
because more than 10% of the values missing

Imputed ggplots Station DEHE161

______PM10_no imputed values

No imputation for this airpollutant at station DEHE161
because more than 10% of the values missing

______PM25_no imputed values

No imputation for this airpollutant at station DEHE161
because more than 10% of the values missing

______SO2_no imputed values

No imputation for this airpollutant at station DEHE161
because more than 10% of the values missing

Imputed ggplots Station DEHE162

______CO_no imputed values

No imputation for this airpollutant at station DEHE162
because more than 10% of the values missing

______NO_no imputed values

No imputation for this airpollutant at station DEHE162
because more than 10% of the values missing

______NO2_no imputed values

No imputation for this airpollutant at station DEHE162
because more than 10% of the values missing

______O3_no imputed values

No imputation for this airpollutant at station DEHE162
because more than 10% of the values missing

Imputed ggplots Station DEHE162

______PM10_no imputed values

No imputation for this airpollutant at station DEHE162
because more than 10% of the values missing

______PM25_no imputed values

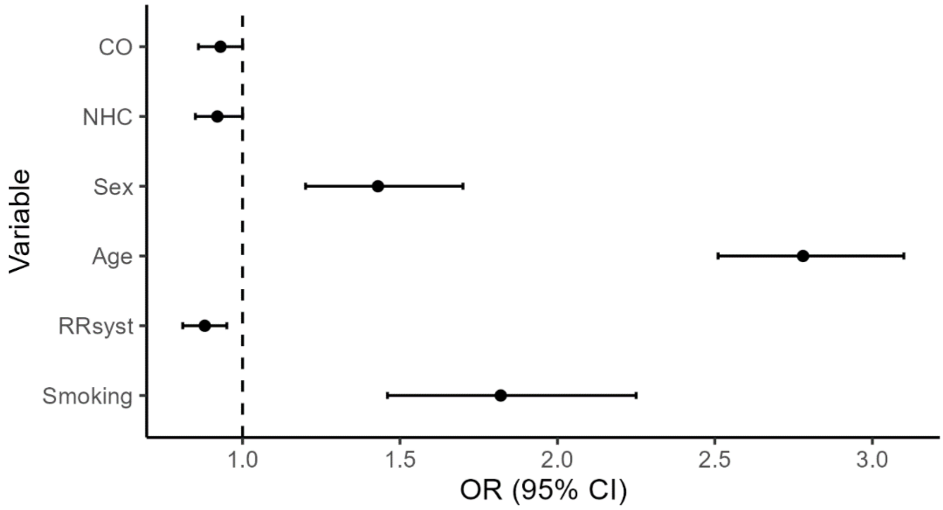
No imputation for this airpollutant at station DEHE162
because more than 10% of the values missing

______SO2_no imputed values

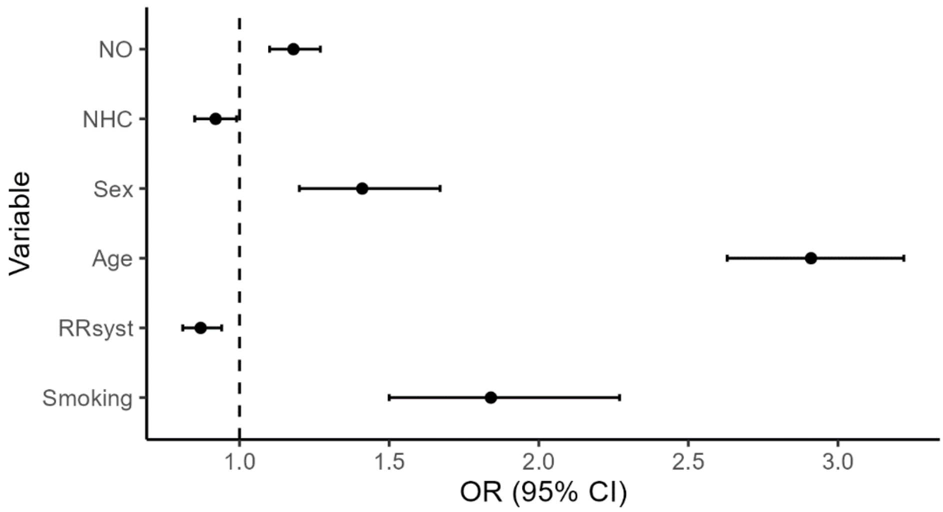
No imputation for this airpollutant at station DEHE162
because more than 10% of the values missing

ORs in multivariate models single air pollutant models except PM_{25} and SO_2

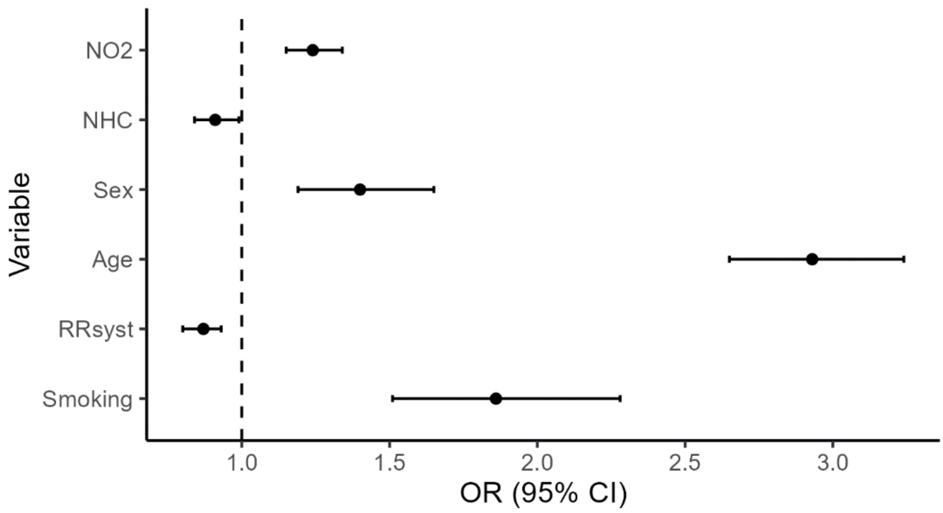
OR model 2 with CO

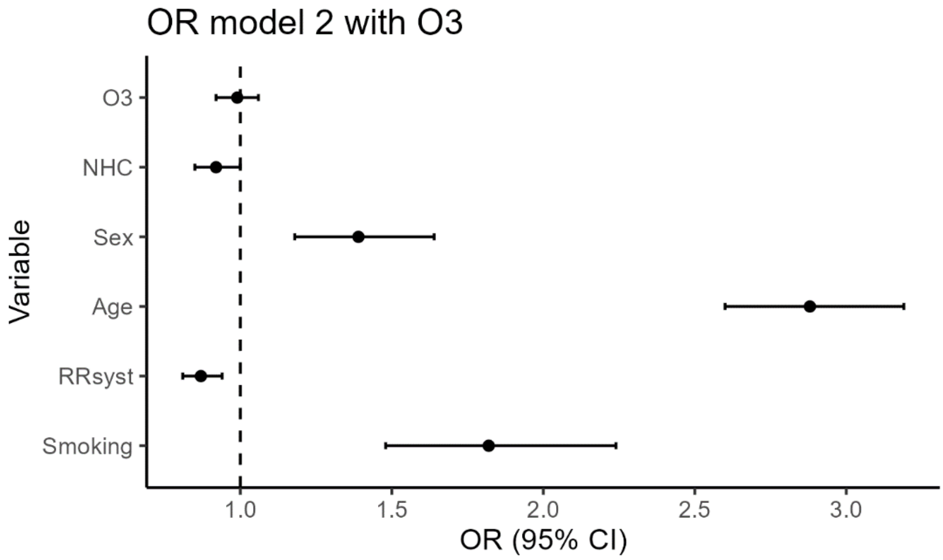


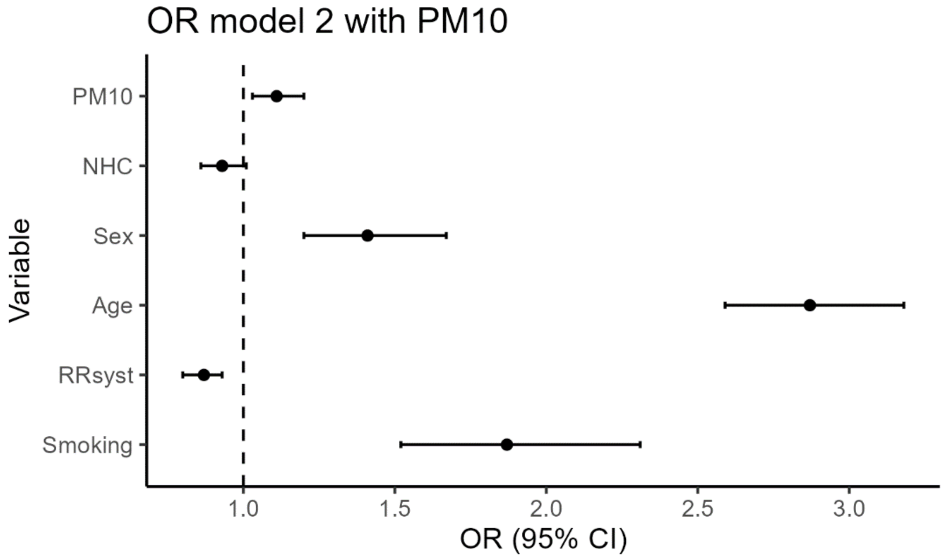
OR model 2 with NO



OR model 2 with NO2




OR model 2 with O3
Variable
O3
NHC
Sex
Age
RRsyst
Smoking
1.0
1.5
2.0
2.5
3.0
OR (95% CI)


OR model 2 with PM10
Variable
PM10
NHC
Sex
Age
RRsyst
Smoking
1.0
1.5
2.0
2.5
3.0
OR (95% CI)

Publications and congress contributions

Congress contribution: 89th annual meeting of the German society of cardiology (DGK) 2023 (Presentation):
Maitra, R.A. *et al.* (2023) ‘Association of air pollution and mortality in individuals with high cardiovascular risk’, *Clinical Research in Cardiology*, 112(7), p. 1005.
doi: 10.1007/s00392-023-02180-w

This contribution was awarded with the 2nd place of the Hans-Blömer Young Investigator Award, German cardiac society.

Congress contribution: Science day 2022 of the Justus-Liebig University, Giessen (Presentation):
Maitra, R.A. *et al.* (2022): ‘Prognostic impact of environmental exposition in prevalent cardiac disease’

Ehrenwörtliche Erklärung

„Hiermit erkläre ich, dass ich die vorliegende Arbeit selbständig und ohne unzulässige Hilfe oder Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Alle Textstellen, die wörtlich oder sinngemäß aus veröffentlichten oder nichtveröffentlichten Schriften entnommen sind, und alle Angaben, die auf mündlichen Auskünften beruhen, sind als solche kenntlich gemacht. Bei den von mir durchgeführten und in der Dissertation erwähnten Untersuchungen habe ich die Grundsätze guter wissenschaftlicher Praxis, wie sie in der „Satzung der Justus-Liebig-Universität Gießen zur Sicherung guter wissenschaftlicher Praxis" niedergelegt sind, eingehalten sowie ethische, datenschutzrechtliche und tierschutzrechtliche Grundsätze befolgt. Ich versichere, dass Dritte von mir weder unmittelbar noch mittelbar geldwerte Leistungen für Arbeiten erhalten haben, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen, und dass die vorgelegte Arbeit weder im Inland noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde zum Zweck einer Promotion oder eines anderen Prüfungsverfahrens vorgelegt wurde. Alles aus anderen Quellen und von anderen Personen übernommene Material, das in der Arbeit verwendet wurde oder auf das direkt Bezug genommen wird, wurde als solches kenntlich gemacht. Insbesondere wurden alle Personen genannt, die direkt und indirekt an der Entstehung der vorliegenden Arbeit beteiligt waren. Mit der Überprüfung meiner Arbeit durch eine Plagiatserkennungssoftware bzw. ein internetbasiertes Softwareprogramm erkläre ich mich einverstanden."

Ort/Datum Unterschrift

Acknowledgements

I am deeply grateful to my supervisor Prof. Dr. Till Keller for introducing me to my doctoral research project and supporting me in its realization. I would like to thank him for his trust and confidence in me and his guidance throughout the project.

I would like to express a special thanks to my co-supervisor PD Dr. med. Beatrice von Jeinsen. Her invaluable and constructive feedback, as well as her creative ideas were a great support to me and have contributed significantly to the success of this project. In particular, she presented the results of this work at the 89th annual meeting of the German society of cardiology (DGK) 2023, while I was prevented from doing so, having to stand my second state examination

I am equally thankful to Dimitri Grün, for his prior analysis enabling the idea of this project, for his technical advice and support in statistical questions. Also, I would like to thank Dr. med. vet. Sandra Voß, Nora Staubach, Claudia Brüderle, Monika Rieschel and all the other staff members of the Franz-Groedel-Institute at the Kerckhoff-Klinik in Bad Nauheim for their support during the realisation of my dissertation project.

I would like to thank Mrs. Bärbel Oehme and Dr. Florian Ditas who have provided me with publicly available air pollution data of the HLNUG Hessisches Landesamt für Naturschutz, Umwelt und Geologie (Hessian State Agency for Nature Conservation, Environment and Geology).

I am thankful for the generous scholarship and valuable research seminars provided to me by the Trainee program of the Justus-Liebig University, Giessen. My great appreciation goes to Prof. Dr. Katrin Susanne Lips who so supportively guided us through this program. I would also like to thank the DZHK Young program (German Centre for Cardiovascular Research) for enabling me to participate in an R Course and the German Academic Scholarship Foundation for supporting the additional semester needed to realise my research project.

I am very grateful to my family, Carola, Robin, Miriam and Lilith for always supporting and inspiring me, being there for me and believing in me and my abilities. Thank you for the endless support, inspiration and discussions – Sarah-Sophie, Merle, Sophia, Vera, Leo, Hanna, Laura and Lenni. And thank you to all my other friends supporting me along the way. Lastly, thank you to Konstantin, for believing in me, his trust, support, the numerous fruitful discussions, encouragement, and love.

Der Lebenslauf wurde aus der elektronischen Version der Arbeit entfernt.

The curriculum vitae was removed from the electronic version of the paper.