Grant Agreement: 814966

CARES

CITY AIR REMOTE EMISSION SENSING

D4.4 - Report on *openCARES* user functions for structured RES data analyses

WP 4
Task 4.2
Deliverable 4.4

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 814966.

April 2023

Document history and validation

This page is used to follow the deliverable production from its first version until it is approved by the Coordinator. Please give details in the table below about successive releases.

When	Who	Comments
Oct 2020	Naomi Farren, David Carslaw	First draft
March 2023	Naomi Farren, David Carslaw	Final version

Authors: Naomi Farren (U York), David Carslaw (U York), Yoann Bernard (ICCT), Uwe Tietge (ICCT), James Tate (U Leeds), Christopher Rushton (U Leeds), Jens Borken (IIASA)

Contact: Naomi Farren
Department of Chemistry, University of York

Email: naomi.farren@york.ac.uk

Deliverable No. D.4.4: Consortium Internal

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 814966. The sole responsibility for the content of this document lies with the authors. It does not necessarily reflect the opinion of the funding authorities. The funding authorities are not responsible for any use that may be made of the information contained herein.

CARES website: https://cares-project.eu/

Summary

This document summarises the final output from Task 4.2 and Deliverable 4.4 concerning the development of an R package called *openCARES*. The *openCARES* package consists of a suite of tools designed specifically for the analysis of vehicle emission remote sensing data. This means that end users, such as city administrations, can access a consistent, openly available, and comprehensive set of analysis techniques to fully exploit the benefits of real-world emission measurements. A version control system hosted on GitHub has been created to help with robust package development. An important achievement has been to automate the production of a summary analysis report; this is output as an html file and enables users to view a highly detailed analysis of their remote sensing data without needing to perform their own analyses. The *openCARES* package includes highly detailed vehicle emission remote sensing measurements from the City of Milan collected as part of the CARES project. An outline of the data required to apply these methods to other cities is defined to ensure that it is straightforward to use the *openCARES* tools widely.

Attainment of the objectives and explanation of deviations

Description of work related to deliverable as given in DoW

This deliverable consists of a report that outlines the key features of the *openCARES* package and includes a link to the publicly available repository. The *openCARES* package is intended to be an open-access software package designed to be used by non-expert data analysers to answer typical questions on the emission performance of measured vehicles. The summary analysis report can be found in the repository, which is compiled using remote sensing data collected during the CARES project.

Time deviation from original DoW

This deliverable was delayed by 12 months to allow time to focus on the coordination of postponed WP1 activities as a result of the COVID-19 restrictions.

Content deviation from original DoW

An important addition to this deliverable is an example of an automated report produced using vehicle emission remote sensing data collected as part of the CARES project. The report is produced in Quarto and rendered as a web page.

Table of Contents

Introduction

Vehicle emissions remote sensing data has the potential to provide a vast wealth of information on vehicle emissions for use in research, policy and many other applications. However, the collection and use of such data has historically been ad-hoc and inconsistent, in part because emission measurement campaigns tend also to be ad hoc and spread across different countries. As a result, a widely accepted method for the analysis of remote sensing data does not yet exist and there are numerous ways in which the data are processed and presented. There are however many common types of analysis that can usefully be applied to remote emission sensing (RES) measurements.

A barrier to the analysis of RES data is that the data sets can be large, cumbersome and complex. These characteristics mean that it can be difficult for non-expert data analysers to work with such data, let alone fully exploit the valuable information within these data sets. As the number of RES campaigns continues to increase worldwide, there is a growing need to develop effective tools to analyse the data from them. The CARES project makes an important contribution in this respect by aiming to lower the barriers of cost and complexity of making and using RES measurements to promote the increased use of RES.

An important component of the CARES project is the further development of hardware to make remote emission measurements. In addition to the progression of 'traditional' remote sensing, this includes the advancement of plume chase methods and the development of point sampling adjacent to roads. All these approaches have their own challenges in providing robust data using systems that are designed to be practical in use. Equally important is the development of analysis techniques to work with the comprehensive data these approaches provide. The development of a package of tools to support analysis is therefore an essential component of the CARES project.

As interest in RES continues to grow, there is an increasing need to provide analysis that helps inform decision makers as to how best to manage emissions from road vehicles. The ICCT has produced several reports that effectively meet these aims, such as the recent report on Krakow (Bernard et al., 2020), see here. These studies provide an excellent template of how best to present and analyse data from RES and will be built on as the work package develops.

Objectives of this report

This report outlines the consistent set of data analysis tools that have been developed for vehicle emission remote sensing data. A major part of the initial work was to set up the infrastructure and template for making analysis tools available. The main way in which analysis tools have been made available is through the development of an R package using R statistical software. This analysis environment is ideally suited to working with vehicle emissions remote sensing data in providing a platform for effectively disseminating data analysis tools.

The main objectives of this report are to:

- Introduce the tools used for the analysis of RES data and their benefits.
- Describe how the tools have been developed and provide details on their access and usage.
- Provide examples of the types of analysis that can be undertaken.

Development of an R package (*openCARES)*

Brief introduction to R and its benefits

An initial consideration for the development of tools to analyse RES data is the approach to use. An early decision was made to use R statistical software. While R has its origins in statistics, it has over the past 20 years of development become a rich environment for carrying out data analysis. There are several benefits of using R as part of WP4, which include:

- R is a robust, open-source programming language developed for the purposes of data analysis.
- It inherently includes a wealth of well-written code for analysing data and carrying out a vast range of statistical analyses.
- It has a highly developed packaging system that comprehensively extends the core R capabilities. Indeed, there are now >16,000 packages on the Comprehensive R Archive Network, CRAN (although the actual number of packages will be considerably higher than this).
- It provides excellent graphical capabilities for visualising complex data sets.
- Packages can easily be brought under version control to track all changes made to package code and provide a way for a team to collaborate on code development.
- Recent developments in R packages and associated software have made analysis automation much easier and have extended the core capabilities to include outputs that are interactive and which the user can influence.

In short, the development of a specific R package for analysing RES data is close to ideal and is something that will be fully exploited as part of the CARES project.

Version control

An important aspect of data analysis software is that it is reliable and can be scrutinised. Through using R as a software platform it allows the code to be fully open source. Additionally, the code itself is hosted on GitHub, which allows the code to be brought under a version control system as well as making it easier for collaboration and contributions to the code base.

The *openCARES* package is hosted on a publicly available repository at https://github.com/davidcarslaw/openCARES. Figure 1 shows a screenshot of the *openCARES* development site on GitHub.

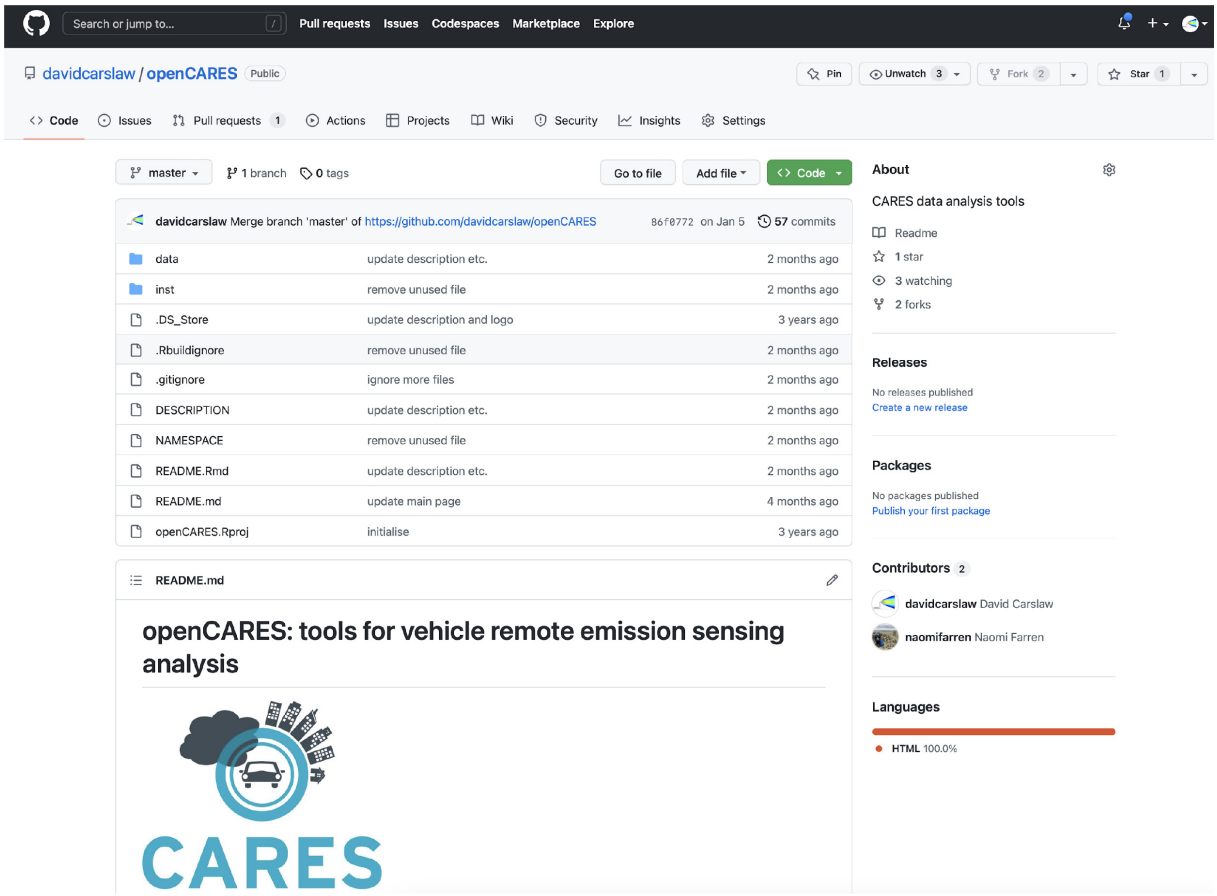


Figure 1. The openCARES development site on GitHub.

Example data

The *openCARES* package currently includes RES data from a measurement campaign conducted in the City of Milan as part of the CARES project. The dataset consists of approximately 35,000 individual vehicle emission measurements. The benefit of incorporating example data in the package itself is that users can work directly with a data set and reproduce different types of analysis as intended. A core set of variables that all field campaign data would be expected to provide has been identified and defined, as outlined below:

Measurement site information:

- Site name
- Latitude and longitude
- Altitude
- Measurement timestamp

Vehicle technical information:

- Euro standard
- Primary fuel type
- Secondary fuel type for bi-fuel and hybrid vehicles
- Vehicle class
- Vehicle manufacturer
- Vehicle category
- Vehicle registration date
- Mileage

Vehicle dynamics information:

- Vehicle speed
- Vehicle Specific Power (VSP)

Environmental variables:

- Ambient temperature
- Relative humidity
- Road gradient

Vehicle emissions data:

- Fuel-specific emissions (g/kg fuel) of CO, NO, NO2, total hydrocarbon (HC), CH4.
- Validity status of emission measurement

Types of analysis

The main types of analysis that would be addressed are summarised in the Description of the Action in the CARES Grant Agreement (Task 4.2), an excerpt of which is below:

> *We address the following functionalities / applications as determined above: Fleet average emission rate; average emission rate by vehicle type (light-duty, heavy-duty truck, bus, ...) and propulsion type (gasoline, diesel, gas, ... and hybrid combinations); by emission standard (Euro class) and by model year. Screening for high-emitters in the data. Emission factors as a function of engine load (VSP); as a function of ambient temperature; as a function of humidity; as a function of vehicle age. Data will be clustered by vehicle manufacturer, brand and model family. All these factors have been found to affect emission rates. The robustness of the resulting emission rate essentially depends on the amount of data (sample size) and the quality of measurements. The functions will first be developed based on the large RES data sampled already in past years (CONOX) and analysed by members of the consortium. The software will include the statistical approaches necessary to robustly detect high-emitters and more generally draw conclusions from the data that are defensible.*

These types of analysis form what might be considered 'core information' related to RES data analysis. They are very common types of analysis that can quickly reveal the nature and characteristics of emissions measurements. Table 1 summarises some of the routine ways in which RES data is analysed. Often, combinations of different analyses are required, which is an aspect that *openCARES* also addresses. Table 1 shows the common types of analysis undertaken on RES data.

Table 1 Common type of analysis undertaken on RES data.

Analysis type	Comments
Euro standard by main vehicle classes	This is a common type of analysis that helps show progress through the main emission standard changes.
Fuel type	Important to show gasoline, diesel, hybrid etc. influence on emissions.
Year of manufacture or vehicle age	Useful to show how emissions change over time and within Euro standards. Alternatively, analysis by age helps to show degradation effects.
Vehicle manufacturer effects	Vehicle emissions can show strong variations due to vehicle manufacturer, engine family and even vehicle model.
Variation with VSP	Vehicle specific power, VSP, is an important variable used for the analysis of RS data; useful to consider the effect of engine load, to allow 'corrections' between different sites and as the basis of g km^{-1} estimates.
Ambient conditions	Emissions of different pollutants can have important variations with ambient factors such as temperature.

Automated analysis report production

During the initial development of the *openCARES* package it became clear that there would be significant advantages to automating the production of a summary analysis report, which could be applied to new data sets, e.g. those collected as part of city demonstrations. The production of an automated data analysis report has several compelling advantages. First, it removes the need for users to develop their own analyses. Second, it is an efficient way of providing a highly detailed analysis of the most common and useful ways of analysing RES data. Furthermore, the development of an automated analysis approach based on an R package under a version control system means that it can easily be extended based on feedback from users.

A report has been developed in a format known as Quarto which is an open-source scientific and technical publishing system. This approach combines the elements of traditional report writing, e.g. as in Microsoft Word, with data analysis in R.[1] There are also other significant advantages to adopting this approach for users. For example, because RES data can be highly detailed and consist of considerable amounts of data (easily of the order of 100,000 measurements and 100 variables), the ability to condense and present this information in a compact and informative way is compelling. Additionally, making it easy for users to download the data used to make plots for further analysis is also attractive.

The *openCARES* package includes a Quarto file *Summary-analysis.qmd*, which can be compiled to produce a rich-format html file that can be viewed in a browser. The package also includes a data set collected as part of the CARES project from the city of Milan using the Edar system from HEAT. The data set includes measurements from over 35,000 vehicles, carefully matched with vehicle technical information such as Euro status, fuel type etc. This data set is typical of the type that can be expected from a remote sensing campaign in Europe and elsewhere. The fields (column names) required are explained in the package itself but include those that would typically be expected to be available – and indeed necessary, to conduct useful analysis. Figure 2 shows an example section of the automated report produced in Quarto - Table 1 details the individual data fields required for the analysis.

Figures 3 - 7 show additional example sections of the automated report produced using RES data from the 2021 Milan campaign. The full version of the example report is available in the html file attached to this document. On the left is a table of contents to navigate through the sections of the report.

[1] Note that other languages are also possible, such as Python.

Data field specifications

A description of the individual data fields required for analysis is provided in Table 1. The data must be in the correct format and units prior to running the analysis. End users have the option to download and edit their own version of the source code.

Table 1. Required data fields for the production of the automated summary report.

Name	Description
site_name	Site name
latitude	Latitude (dec °)
longitude	Longitude (dec °)
altitude	Altitude (m)
slope	Road slope (°)
amb_temp	Ambient temperature (°C)
amb_rhum	Relative humidity (%)
date_time	Local date and time (YYYY-MM-DD HH:mm:ss)
co_fm	Fuel-specific CO emissions (g kg^{-1})
no_fm	Fuel-specific NO emissions (g kg^{-1})
no2_fm	Fuel-specific NO_2 emissions (g kg^{-1})

Figure 2. A screen capture of Table 1 in the summary report, which provides a list of the individual data fields required for the analysis.

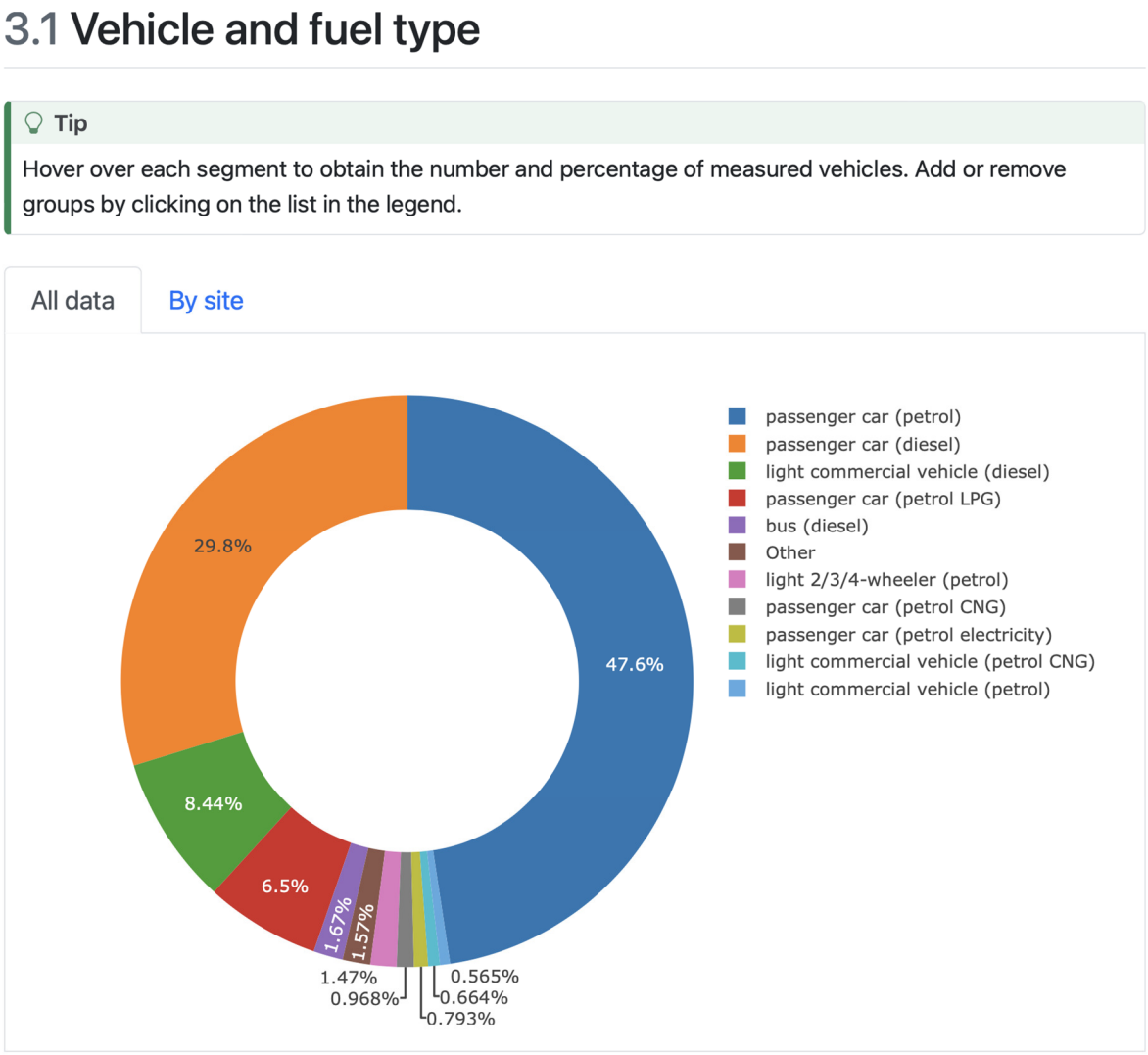


Figure 3 shows the vehicle fleet composition, grouped by vehicle type and fuel type. Vehicles that comprise less than 0.5% of the fleet are assigned as 'Other'. The vehicle and fuel type breakdown for each site can also be viewed by clicking the 'By site' tab.

3.3 Manufacturers

Note

The manufacturers are assigned to groups e.g. "VWG" includes Audi, Bentley, Lamborghini, Porsche, Seat, Skoda and Volkswagen. The size of each rectangle is proportional to the share of each manufacturer / manufacturer group.

All fuel types | Petrol | Diesel

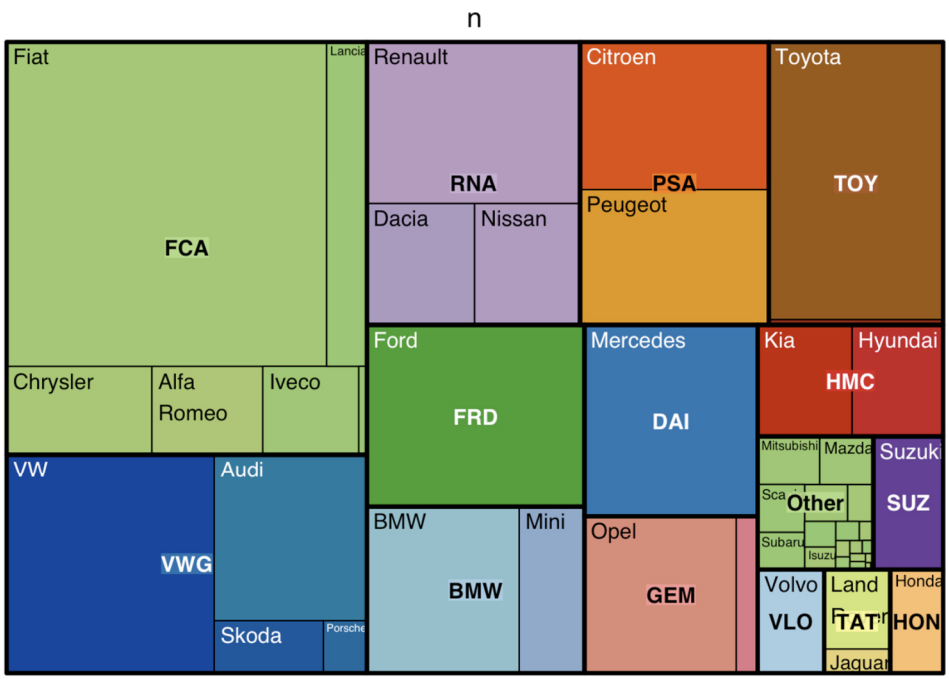


Figure 4 shows the manufacturer composition of the vehicle fleet. The manufacturers are assigned to groups, e.g. 'RNA' includes Renault, Dacia and Nissan, as shown by the purple rectangles. The size of the rectangles is proportional to the share of each manufacturer in the measured fleet. The manufacturer composition grouped by the main fuel types (petrol and diesel) can also be viewed by clicking on the relevant tabs.

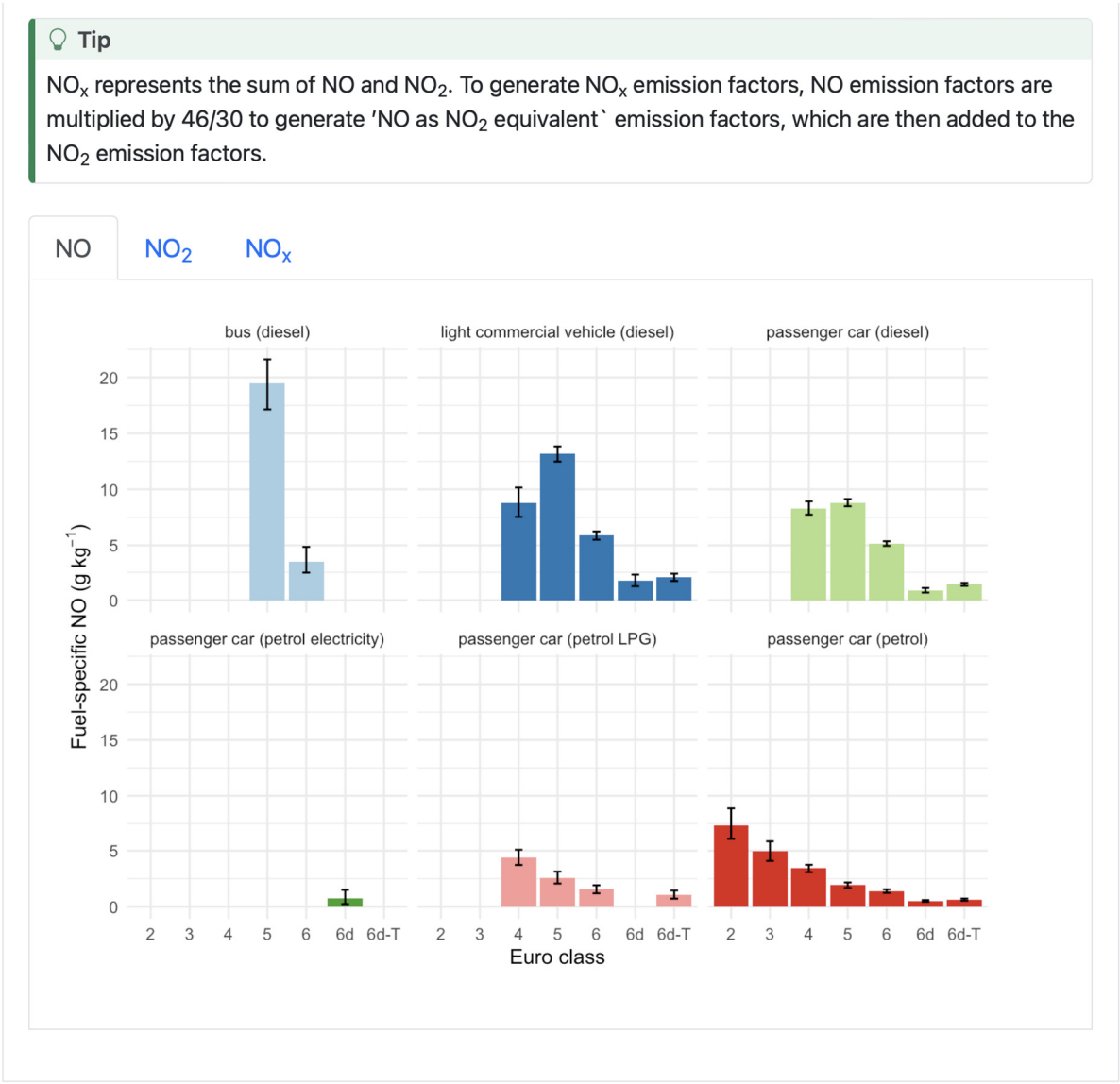


Figure 5. Example output from the summary report to show average fuel-specific emission factors, grouped by vehicle type, fuel type and Euro class.

It can be useful to disaggregate the data by multiple factors to obtain more detailed insight into emissions behaviour. Figure 5 shows average fuel-specific emission factors for NO, grouped by vehicle type, fuel type and Euro class. The data has been considered in this way for multiple pollutants, including NO, NO2, CO and HC. These can be navigated using the relevant tabs in the Quarto report.

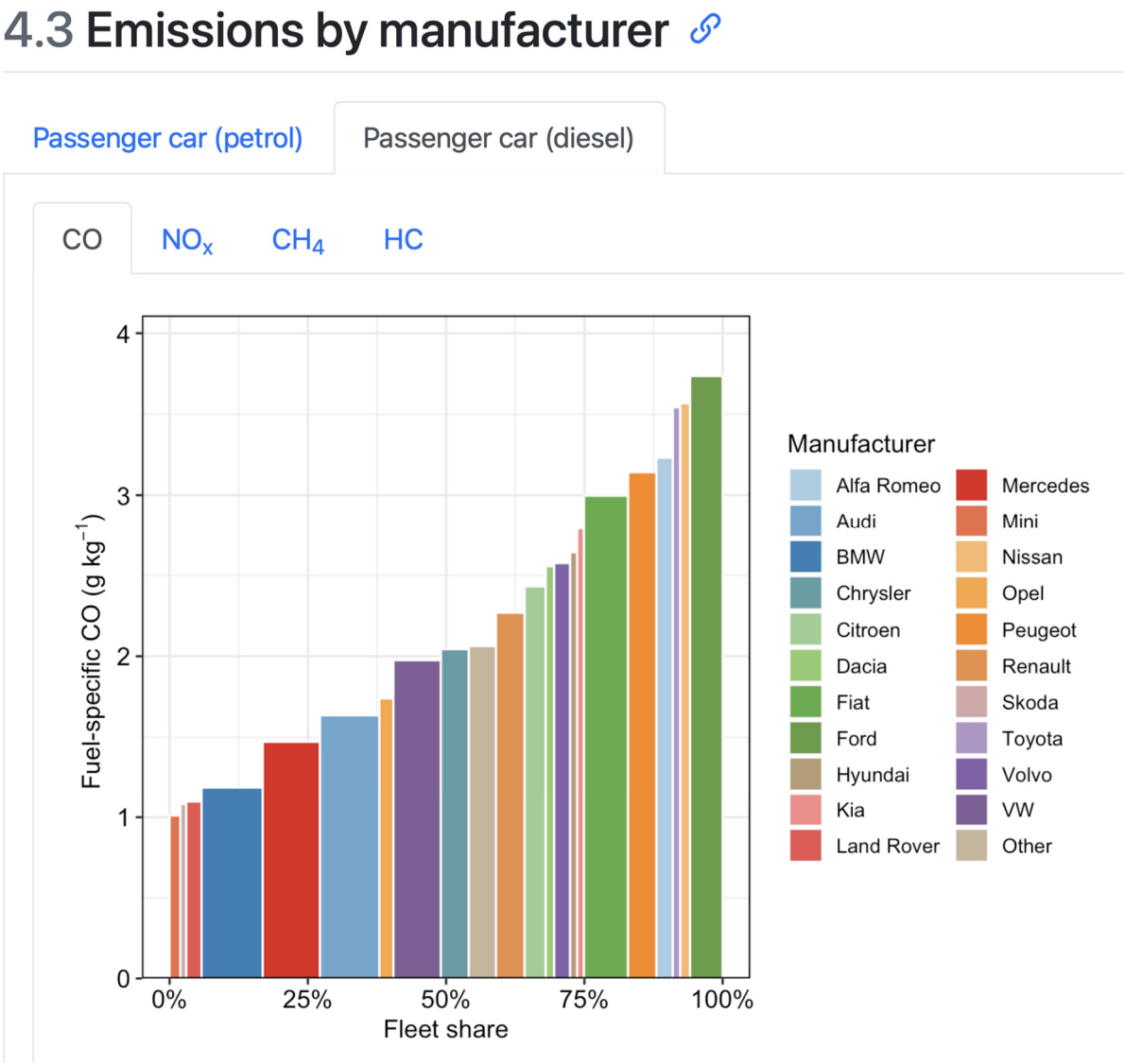
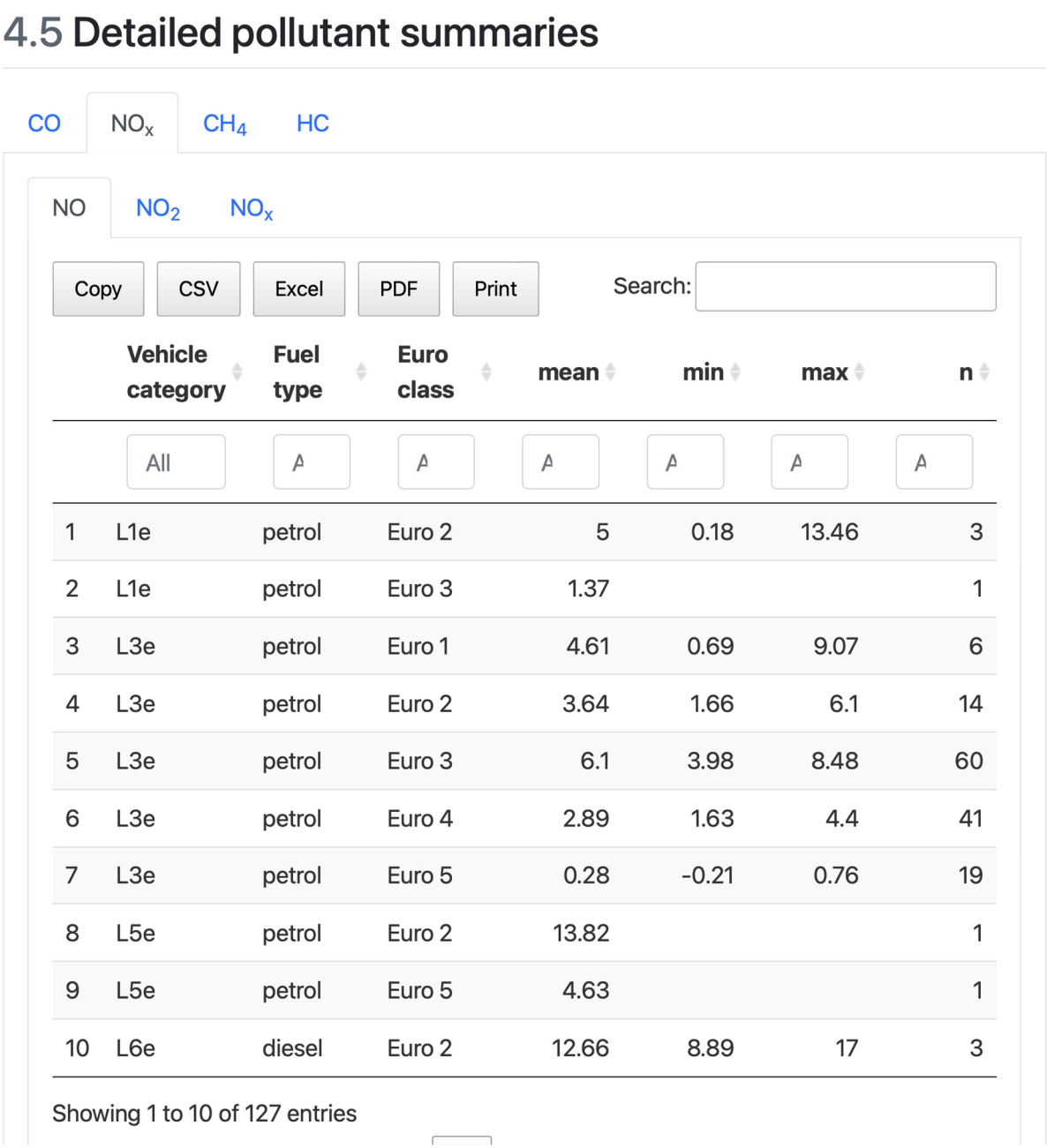


Figure 6 shows average fuel-specific emission factors for CO, grouped by vehicle type, fuel type and manufacturer. The tabs can be used to navigate through different fuel types and pollutants. Importantly the width of the rectangles represents the relevant share of each manufacturer for each group of measured vehicles.



4.5 Detailed pollutant summaries

CO | NO_x | CH_4 | HC

NO | NO_2 | NO_x

Copy CSV Excel PDF Print Search:

	Vehicle category	Fuel type	Euro class	mean	min	max	n
	All						
1	L1e	petrol	Euro 2	5	0.18	13.46	3
2	L1e	petrol	Euro 3	1.37			1
3	L3e	petrol	Euro 1	4.61	0.69	9.07	6
4	L3e	petrol	Euro 2	3.64	1.66	6.1	14
5	L3e	petrol	Euro 3	6.1	3.98	8.48	60
6	L3e	petrol	Euro 4	2.89	1.63	4.4	41
7	L3e	petrol	Euro 5	0.28	-0.21	0.76	19
8	L5e	petrol	Euro 2	13.82			1
9	L5e	petrol	Euro 5	4.63			1
10	L6e	diesel	Euro 2	12.66	8.89	17	3

Showing 1 to 10 of 127 entries

Figure 7. Example output from the summary report to show detailed pollutant summary tables. These can be filtered directly by the user and exported in a range of different formats, e.g. copied to the clipboard, .csv and Excel, for further analysis and presentation.

Future potential

The openCARES R package is publicly accessible and ready for end users to use their own vehicle emission remote sensing data to produce summary analysis reports. End users can also

use the existing code as a starting point for developing their own code for further analysis. The Github version control system means that anyone can suggest new developments for the openCARES package.

References

Bernard, Y.; Tietge, U.; Pniewska, I. (2020). *Remote sensing of motor vehicle emissions in Krakow* (ICCT consulting report). International Council on Clean Transportation. https://theicct.org/publications/remote-sensing-krakow-sept2020.