

Air Quality index

By Camilla Colanero, Diletta Pasin & Andrea Pizzolato

INTRODUCTION

DATASET

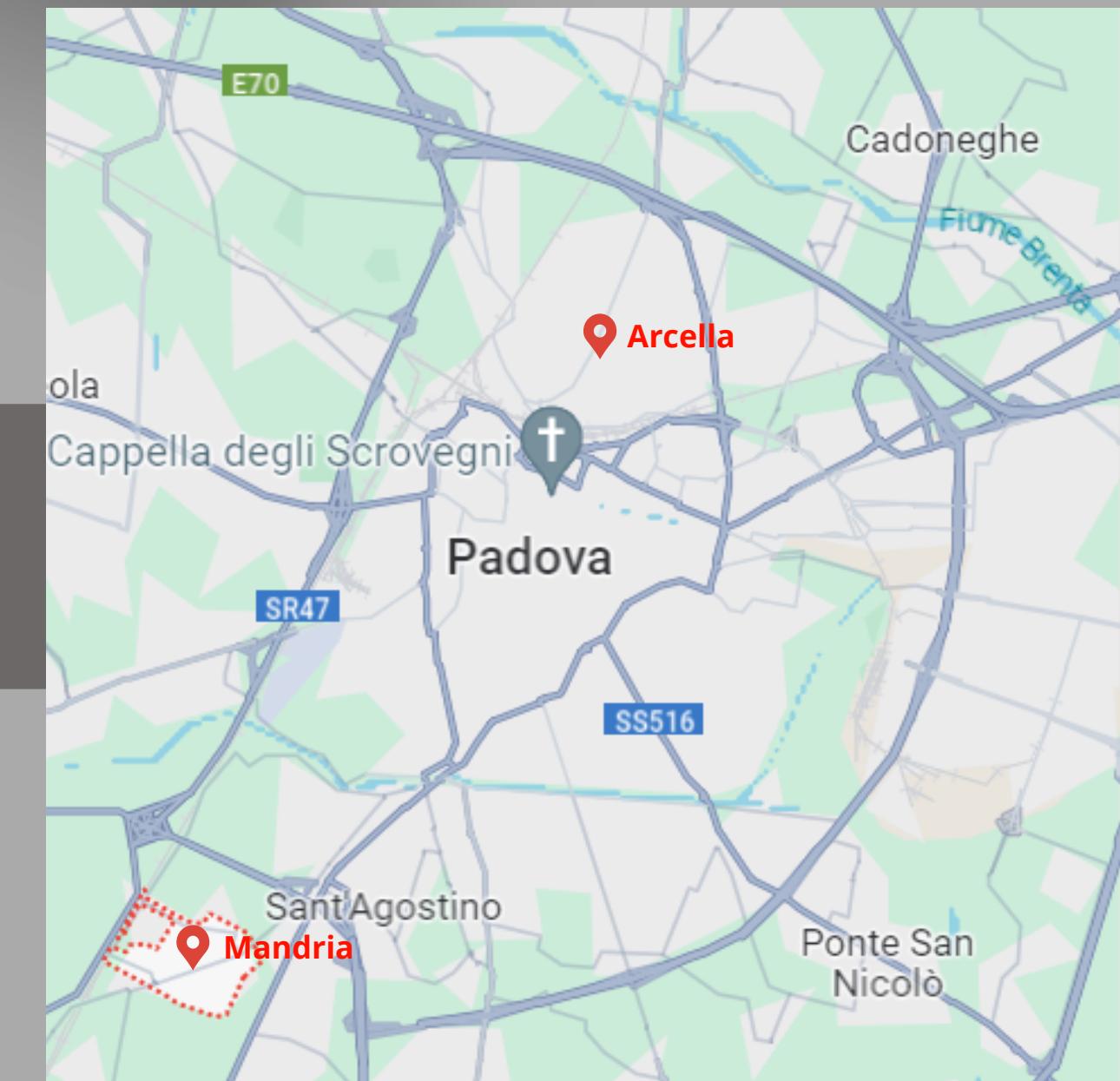


Agenzia Regionale per la Prevenzione
e Protezione Ambientale del Veneto

Arpav **hourly data** about the two stations of Arcella and Mandria

From hourly data to **daily data**:
max of each day

From 1st January 2012 to
30th June 2023



POLLUTANTS

CO

from inexhaust combustion of fossil fuels , can be created domestically , by industrial heating and transports

NO₂

from artificial heating of houses and industries, and combustion of fossil fuels

SO₂

from petrol lavoration cycle, combustion of biomasses and naval transport

O₃

related to seasonal cycles of temperatures and other polluting processes, it is consumed in various atmospherical reactions

PM₁₀

solid residuals from combustion and other reactions, whose diameter is below 10 micrometers

PM_{2.5}

solid residuals from combustion and other reactions, whose diameter is below 2.5 micrometers

AQI

Definition:

The Air Quality index is an indicator developed by government agencies to communicate to the public **how polluted the air** currently is.

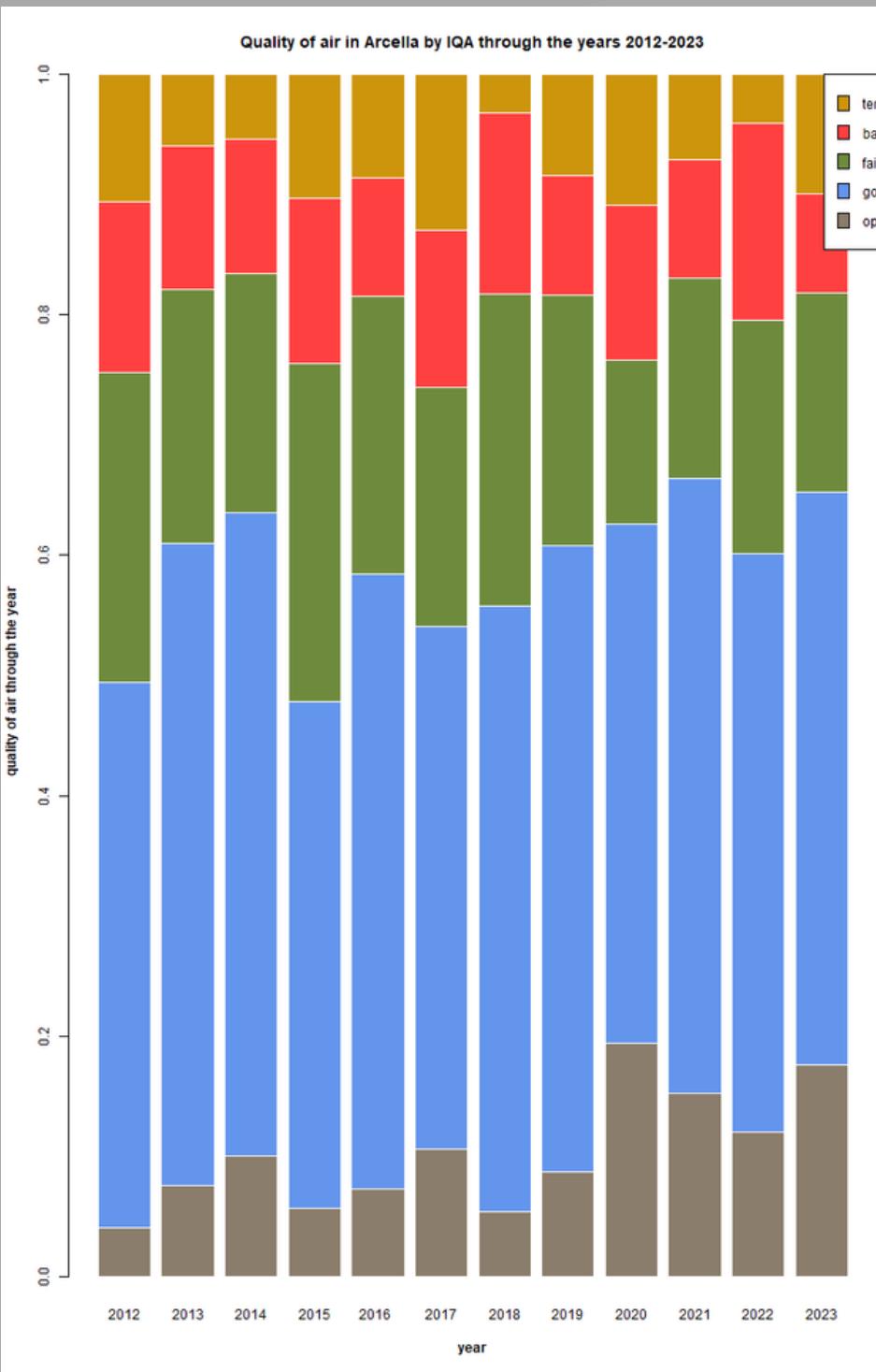
Calculation:

For each pollutant the data is divided by the **national standard** and multiplied by 100 to get the AQI for the pollutant. Then the **maximum** is taken.

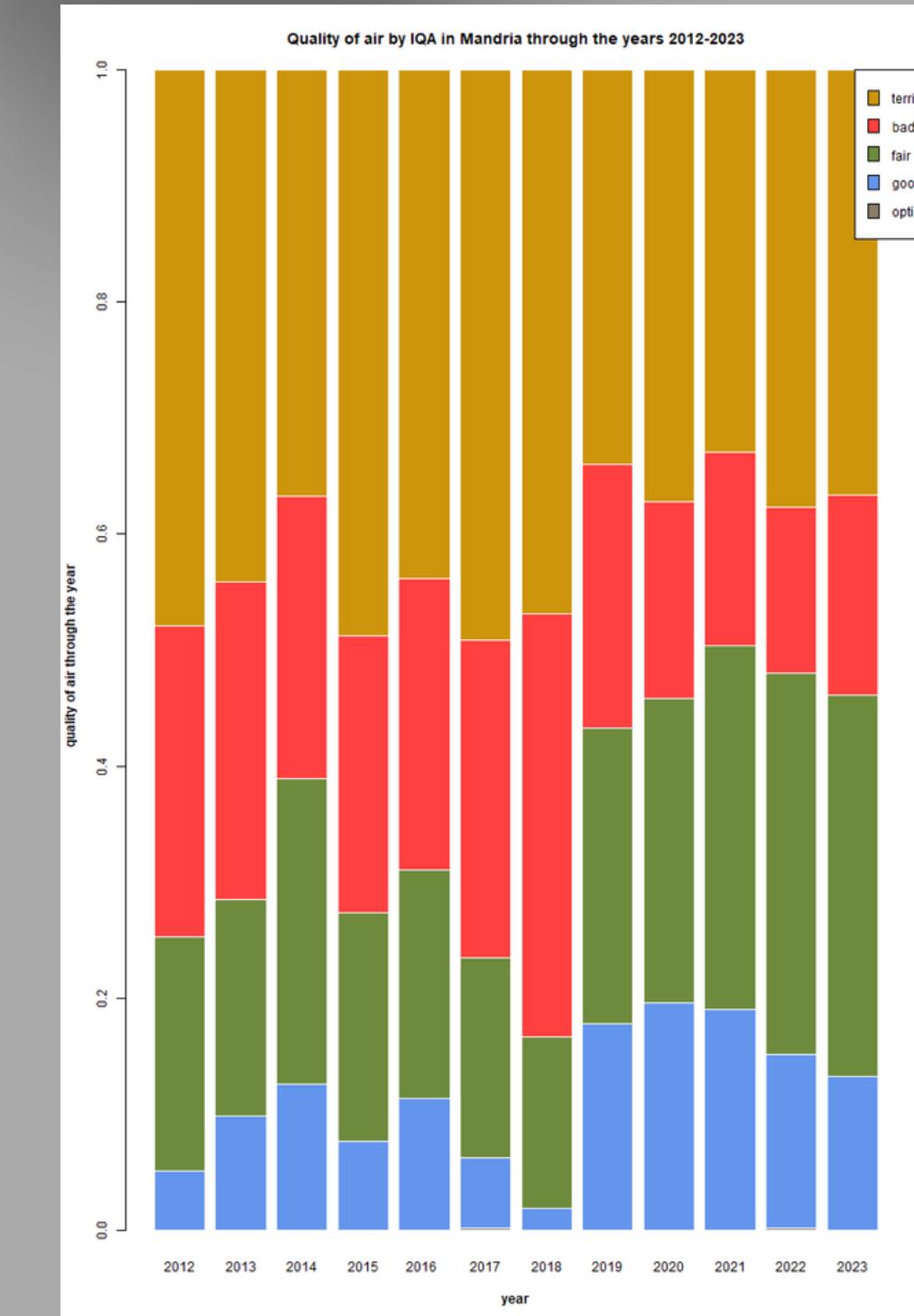
$$AQI_{pollutant} = \frac{pollutant\ data\ reading}{standard} \times 100$$

AQI from 2012 to 2023

ARCELLA



Terrible
Bad
Moderate
Good
Optimal



MANDRIA

QUESTIONS

1

How many times in the past years did we exceed law limits?

2

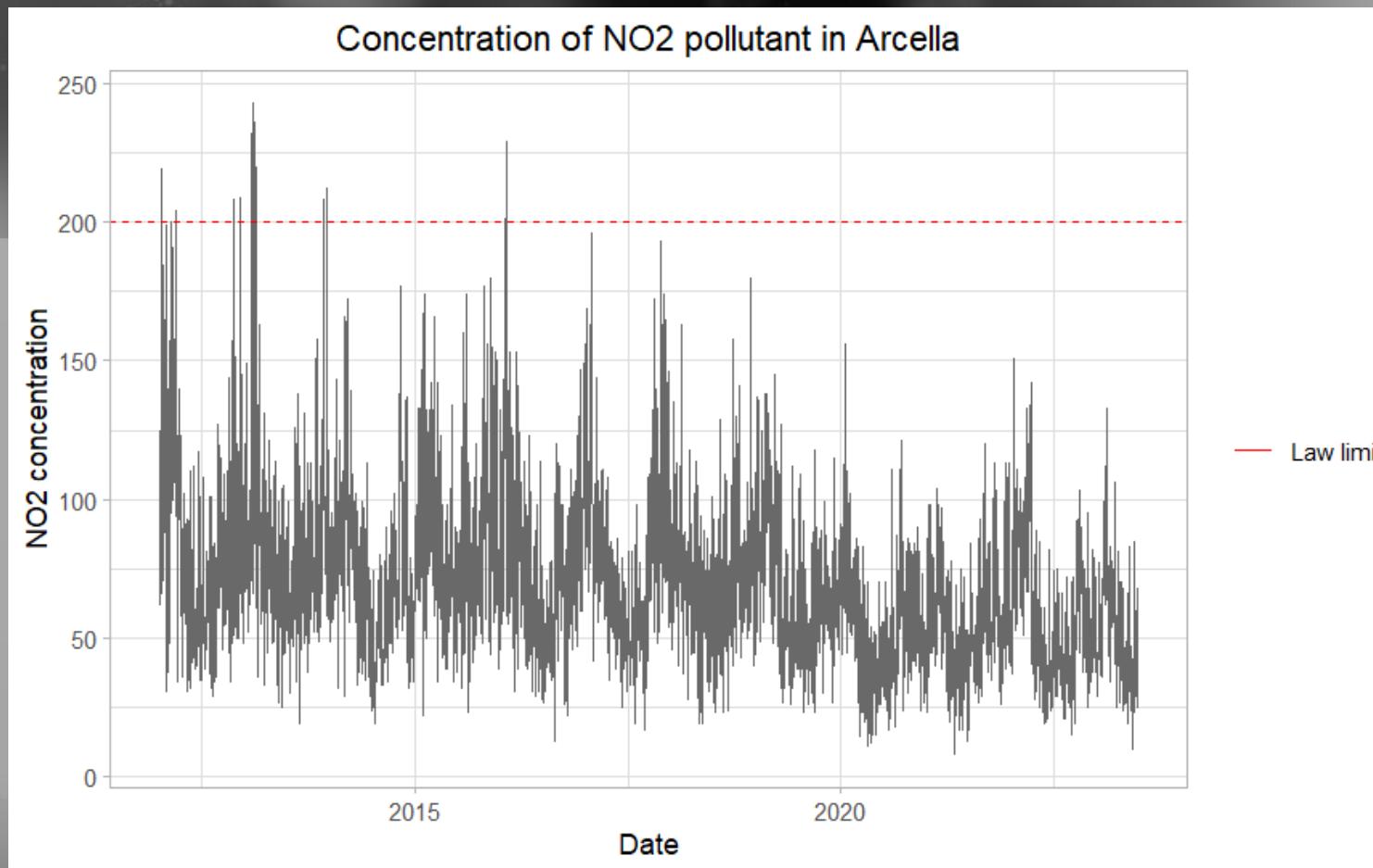
What are the trend and seasonality of the pollutants' concentration?

3

Can we predict the air quality index for the next one, two or six months?

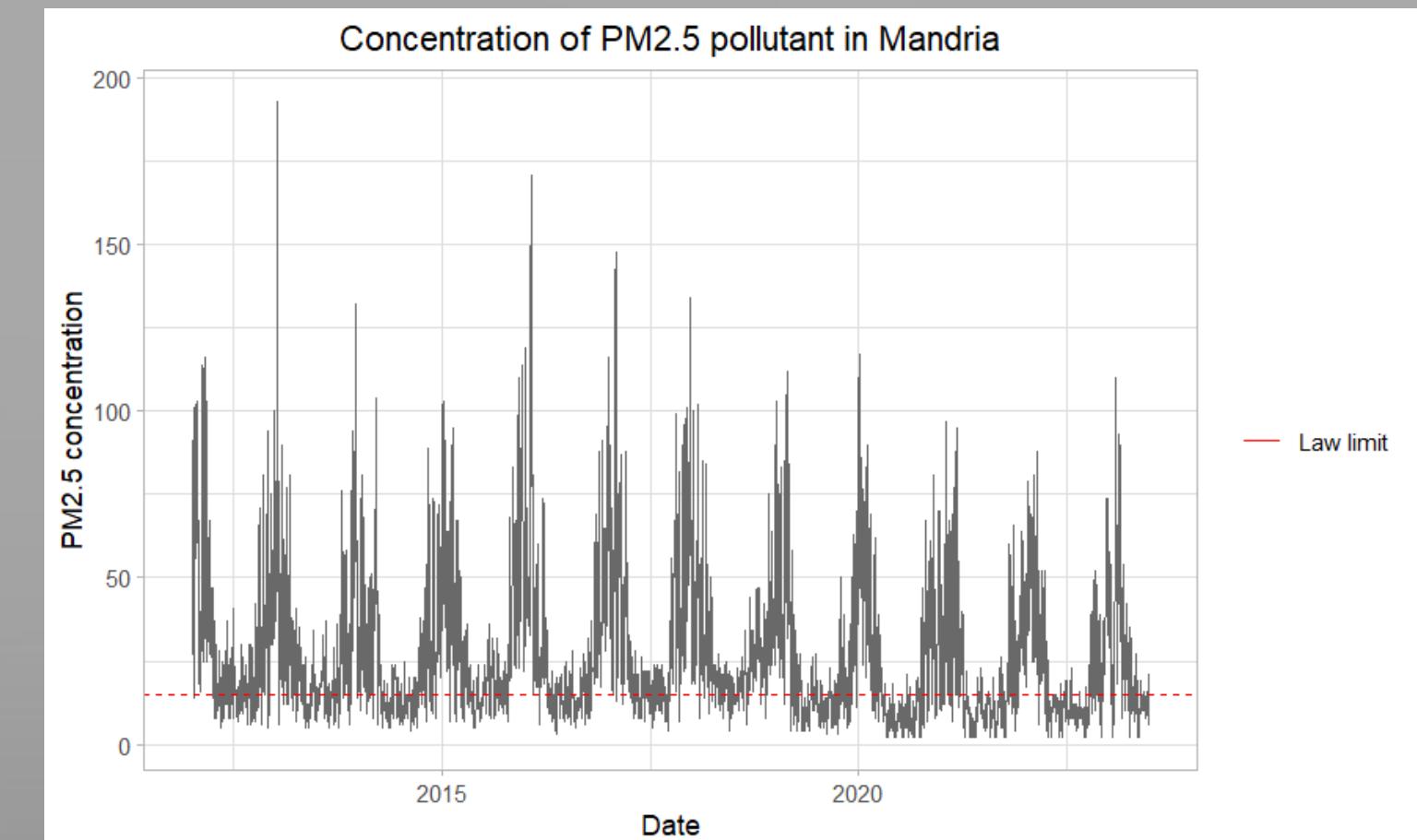
DATA VISUALIZATION

POLLUTANTS PLOTS



Example of a pollutant that exceeds law limits very frequently

Example of a pollutant that does not exceed law limits frequently

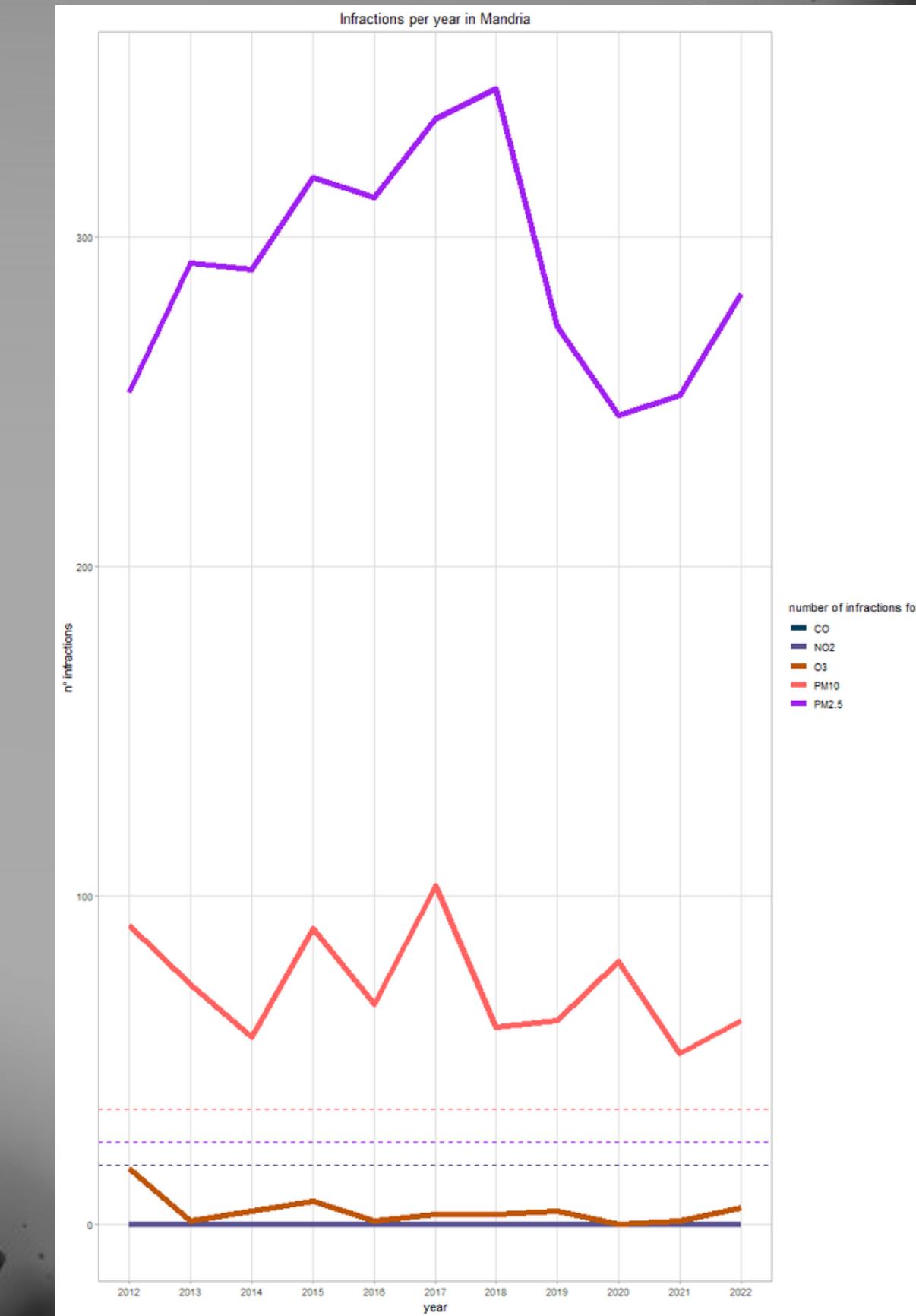
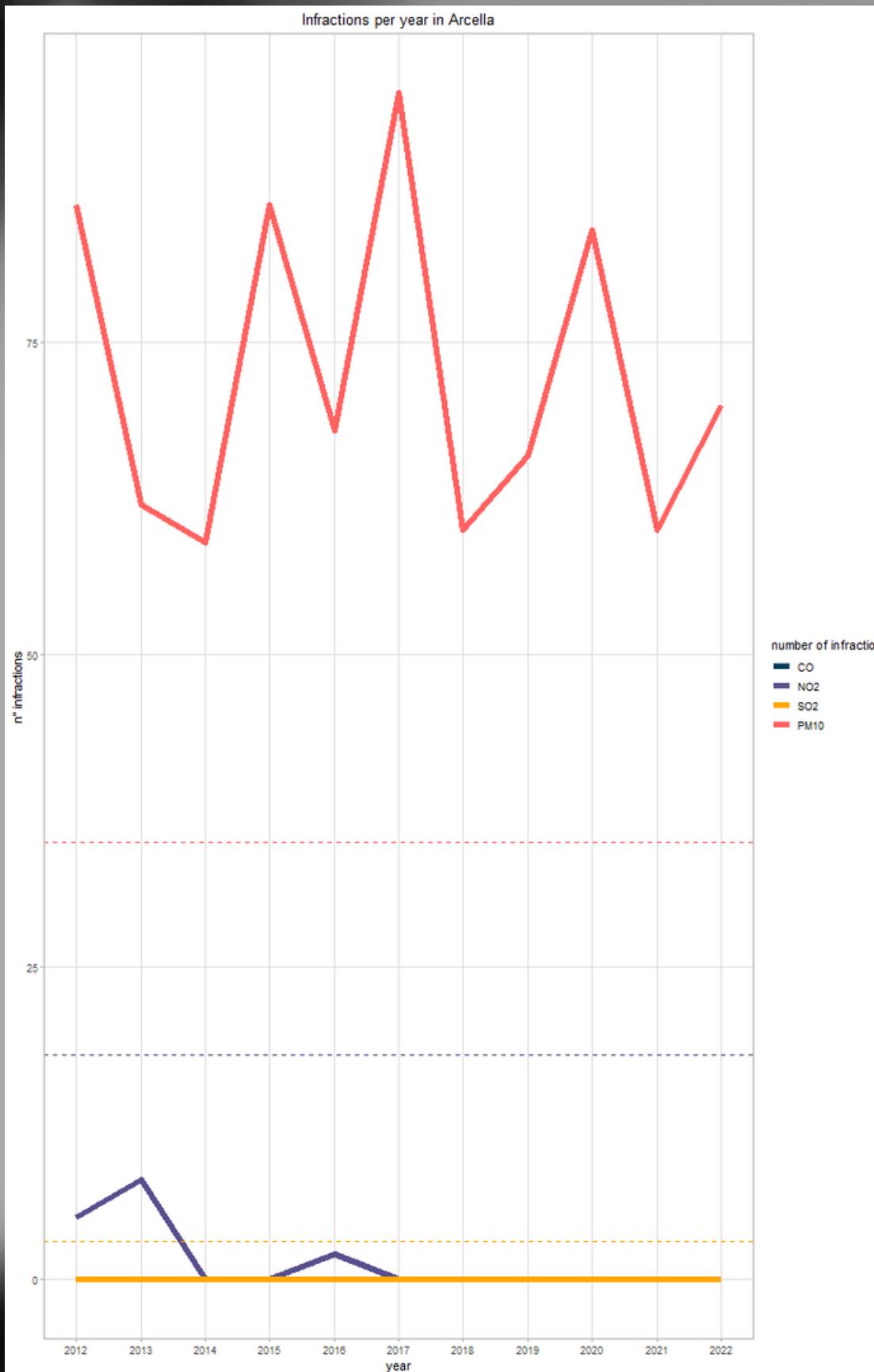


LAW LIMITS

EXCESS LAW LIMITS

POLLUTANT	ITALIAN LAW LIMIT	LAW LIMIT NUMBER OF DAILY INFRACTIONS IN A YEAR
CO	10 µg/m ³	not known
NO2	200 µg/m ³	18
SO2	125 µg/m ³	3
O3	180 µg/m ³	not known
PM10	50 µg/m ³	35
PM2.5	15 µg/m ³	25

ARCELLA



MANDRIA

PUBLIC INTERVENTIONS

Blocks of the traffic: three different levels of risk, related to the PM10 levels.

Effective in the reduction of some pollutants (especially NO₂ which is heavily related to fues consumption).

Heating regulations: from 2012 it has been approved a norm that imposes to all public offices to keep the temperature for heating lower than 20° C in order to reduce the fossil fuels consumption.

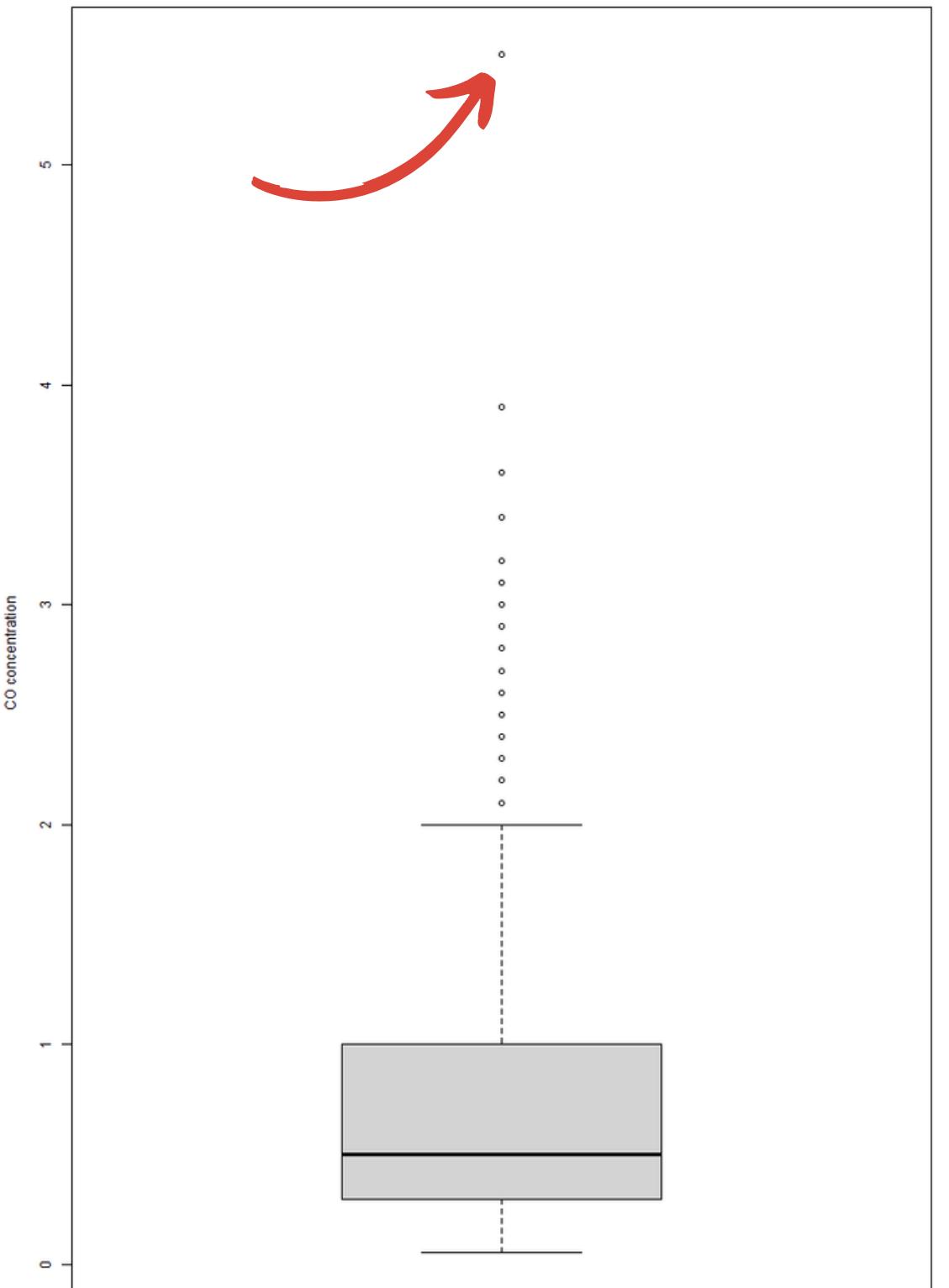
DATA MANIPULATION

OUTLIERS

The data presents **only one** very high outlier in the CO concentration, relating to Arcella's station, on the **16th of April 2021**.

Since it was not possible to attribute that anomalous data to any particular event that occurred that day, we replace that value with the **second highest** of that day.

Boxplot of CO concentration in Arcella



VARIABLES' SMOOTHING

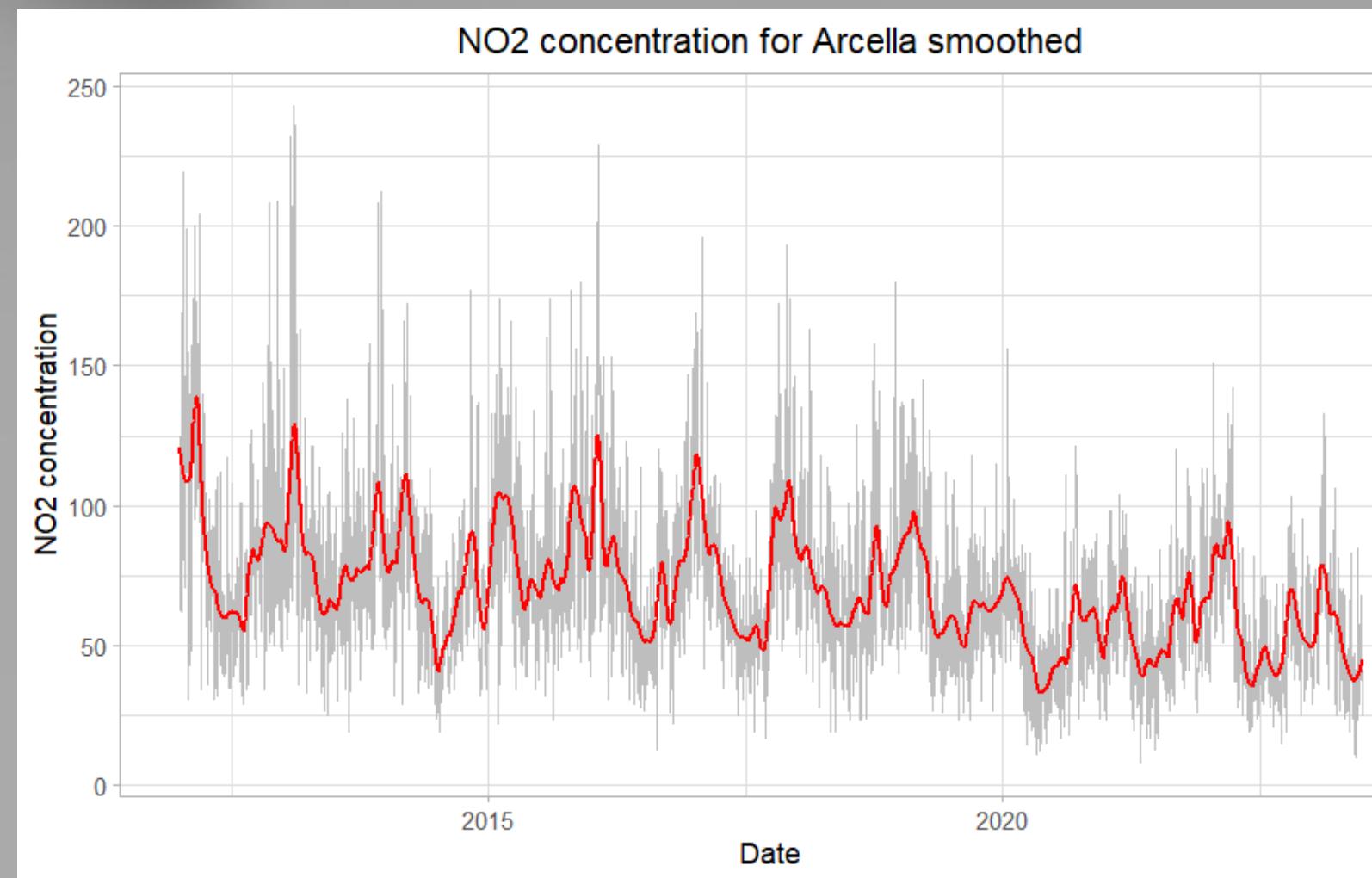
WHY?

Remove high noise from the original data

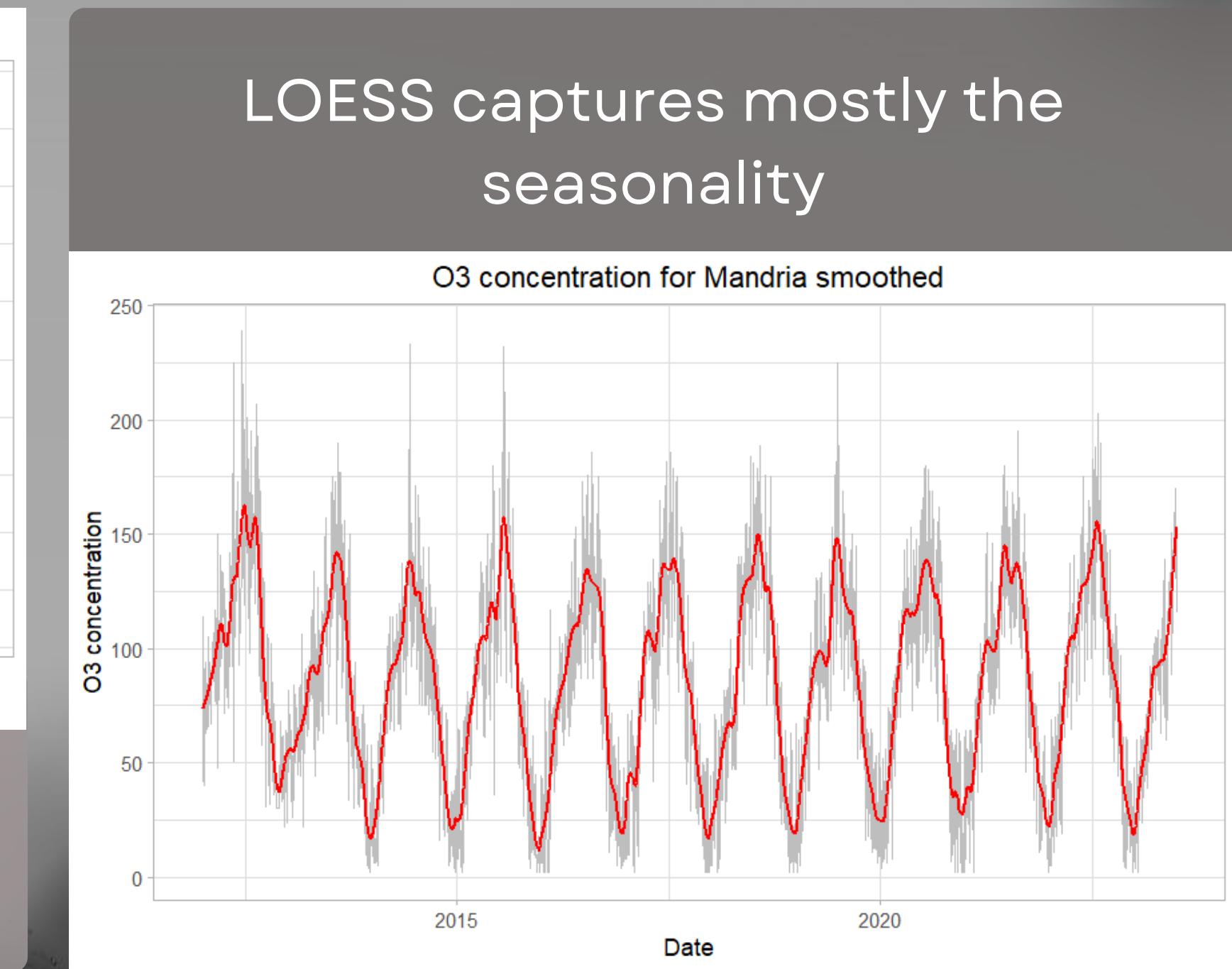
HOW?

span model, with the parameter *span* set to 0.02 for each pollutant

PLOTS



LOESS captures both the seasonality
and the (decreasing) trend



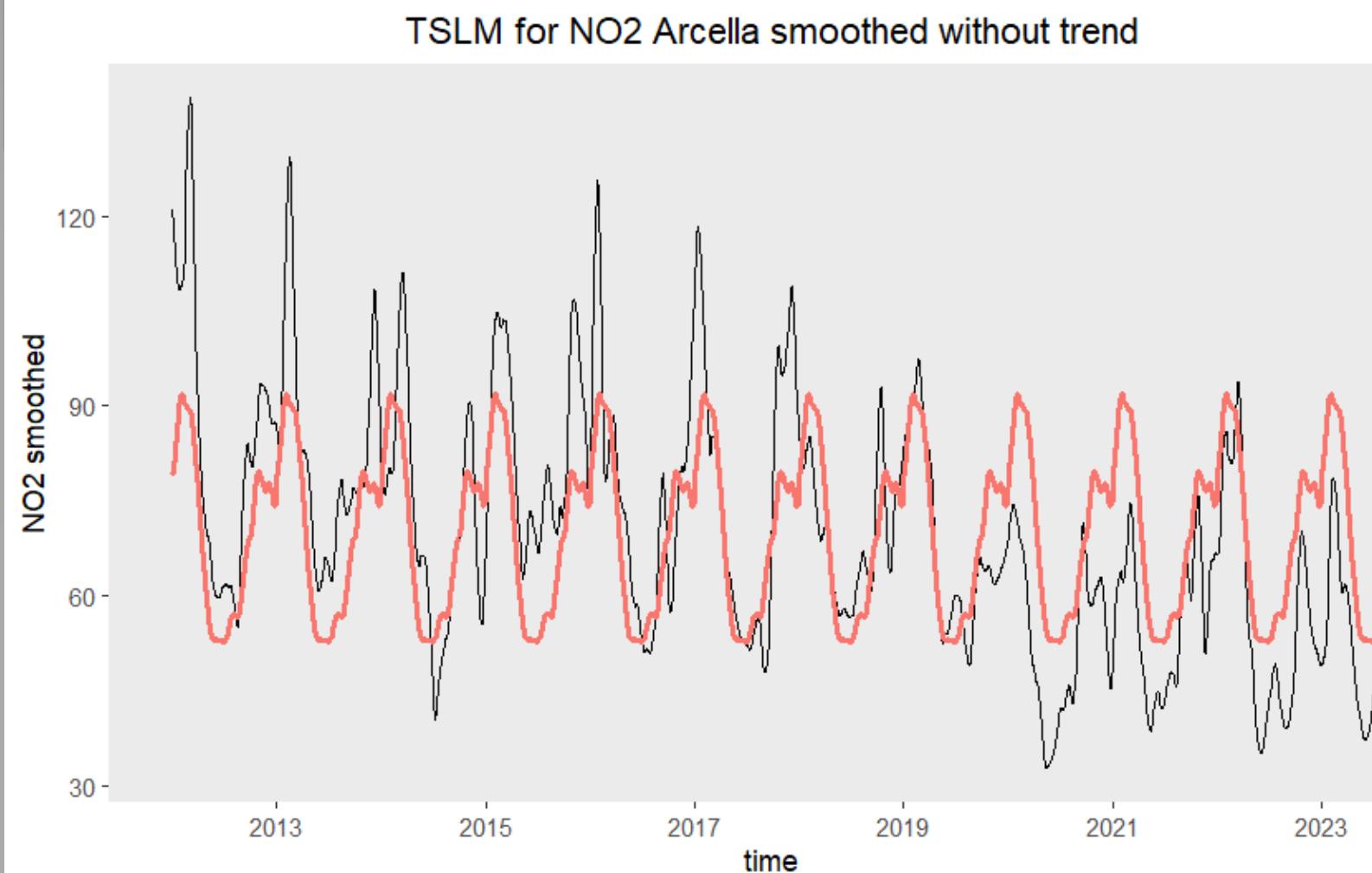
TIME SERIES LINEAR MODELS

TSLM

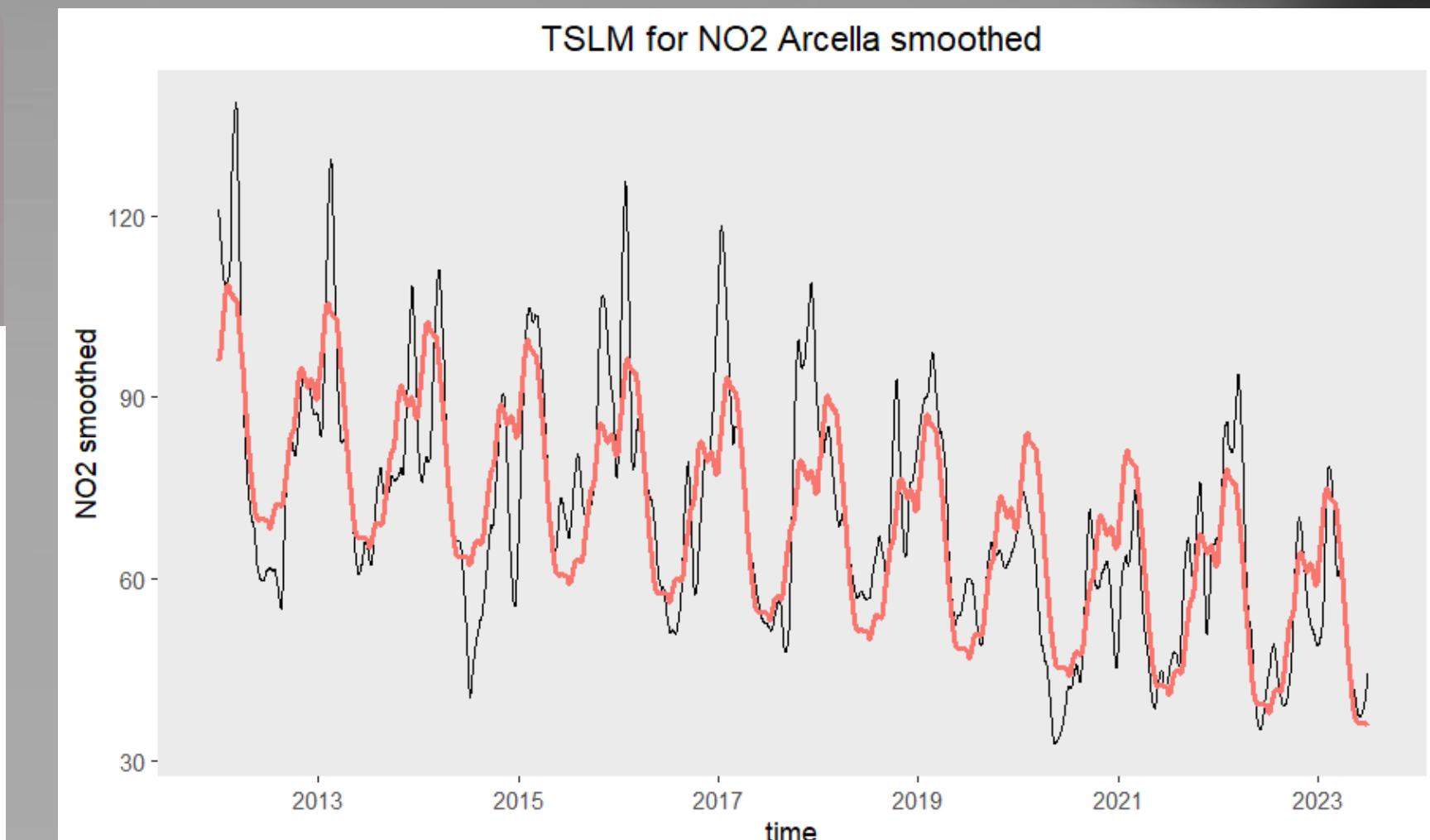
We applied the TSLM to each pollutant, with the aim of capturing the presence of a trend (if any), in addition to the evident seasonality.

		Multiple R^2			Adjusted R^2		
		Seasonality	Trend	Trend + seasonality	Seasonality	Trend	Trend + seasonality
ARCELLA	CO	0.9416	0.01243	0.9532	0.9361	0.01219	0.9487
	NO2	0.4456	0.2762	0.7147	0.393	0.276	0.6875
	SO2	0.6852	3,66E-02	0.6853	0.6554	-0.0002018	0.6553
	PM10	0.7856	0.004728	0.7894	0.7652	0.00449	0.7694
MANDRIA	CO	0.8345	0.07509	0.9075	0.8188	0.07487	0.8987
	NO2	0.6329	0.1383	0.7663	0.598	0.1381	0.7441
	O3	0.9311	0.0006326	0.9316	0.9246	0.0003944	0.9251
	PM10	0.7575	0.02402	0.7797	0.7345	0.02379	0.7587
	PM2.5	0.8034	0.02335	0.8249	0.7847	0.02312	0.8083

TSLM with only seasonality



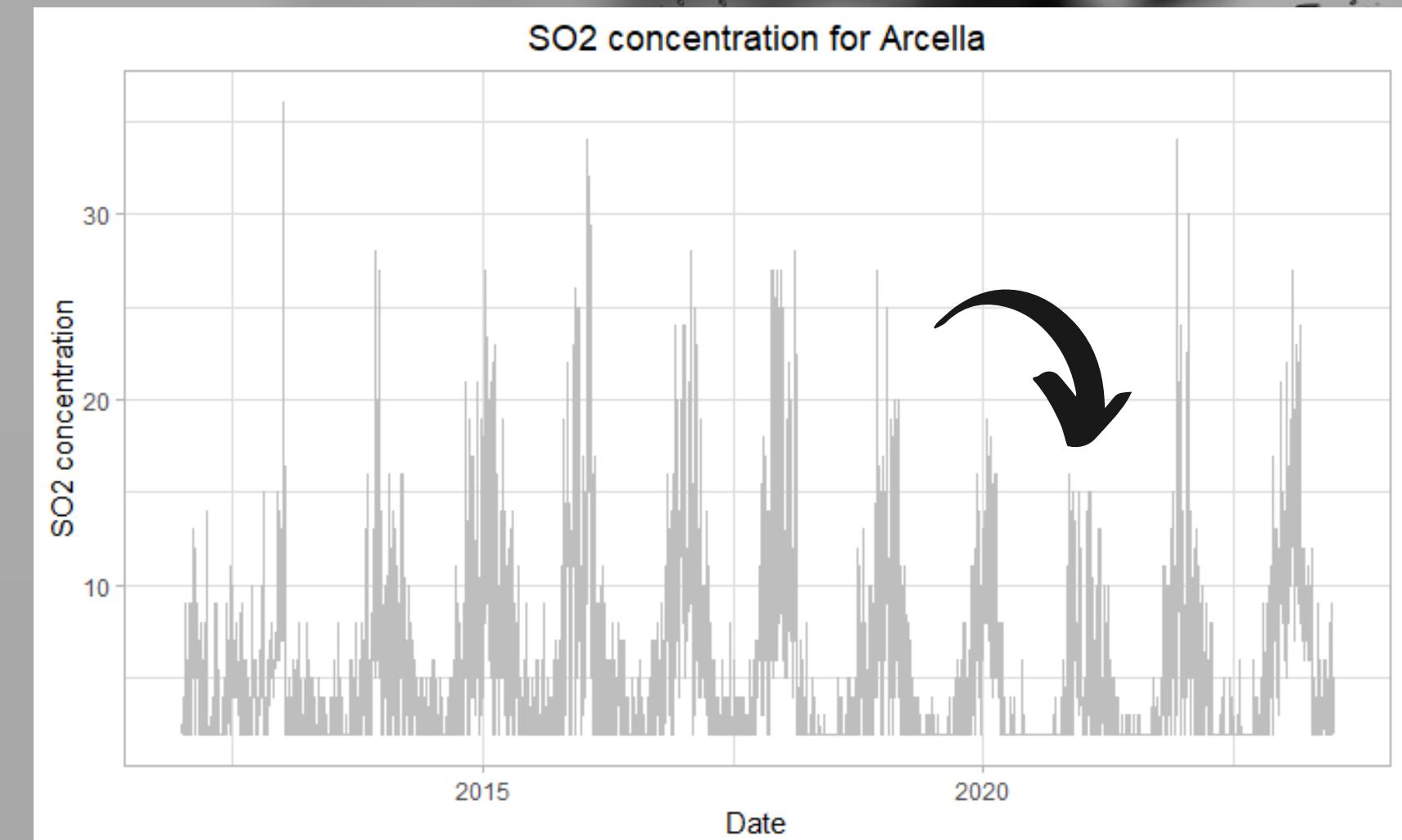
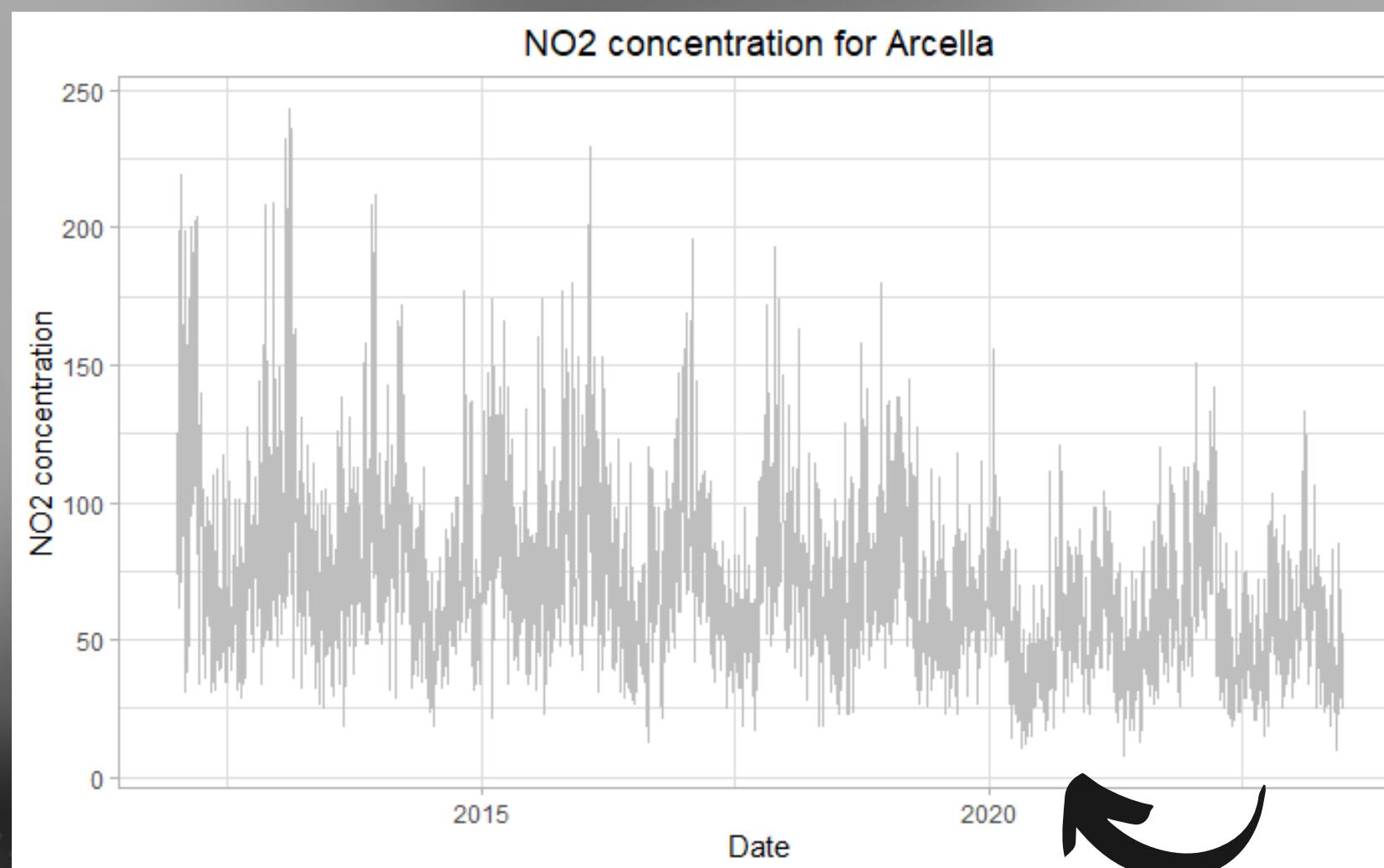
TSLM for NO₂ Arcella smoothed



TSLM with both trend and seasonality

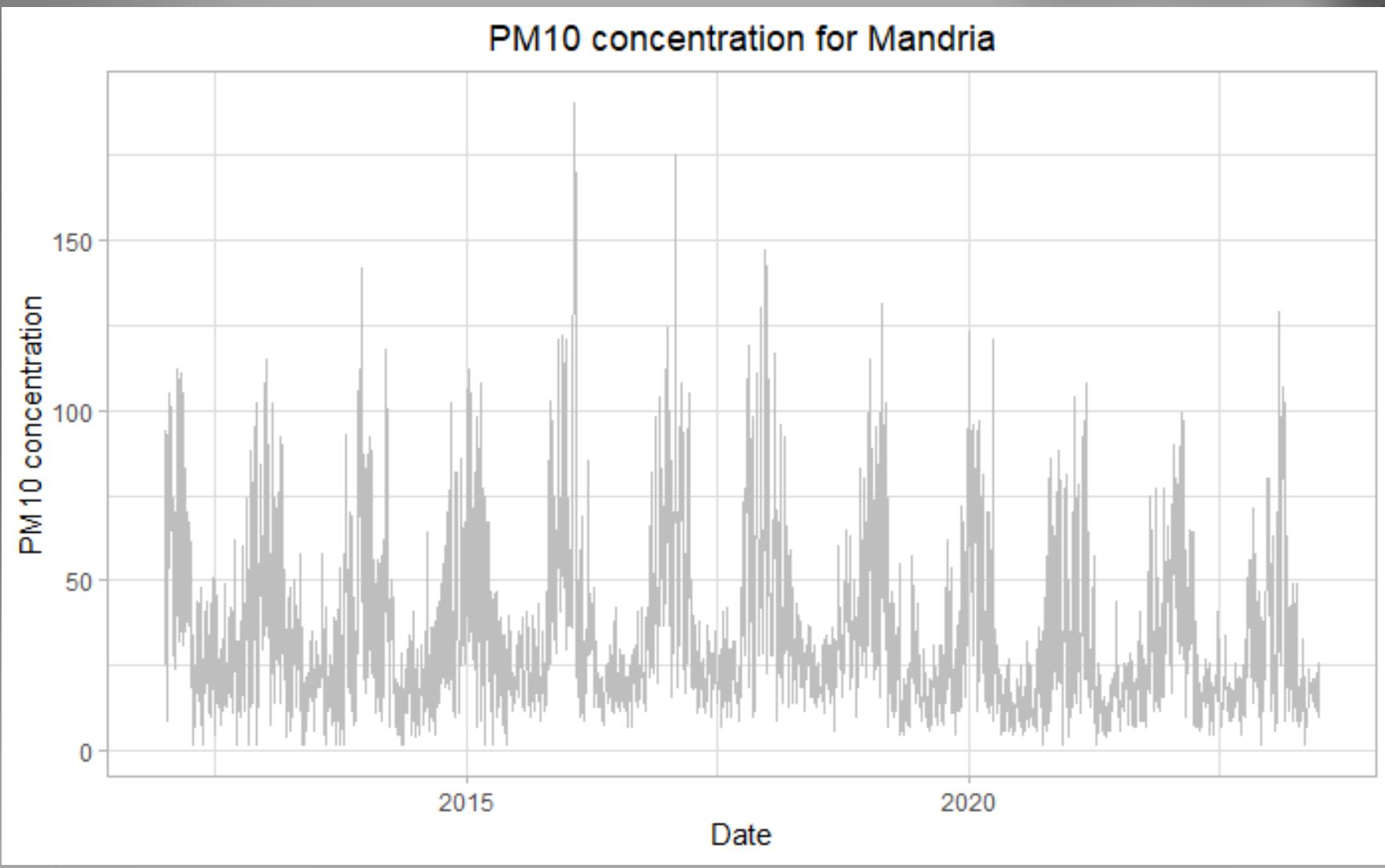
EFFECTS OF COVID-19

There are some pollutants that reached their historical minimum during the first wave of COVID-19, and therefore during the first lockdown. This is the case, for example, of NO₂



On the other hand, there are other pollutants that reached their "minimum peak" in winter 2021, as in the case of SO₂

EFFECTS OF COVID-19



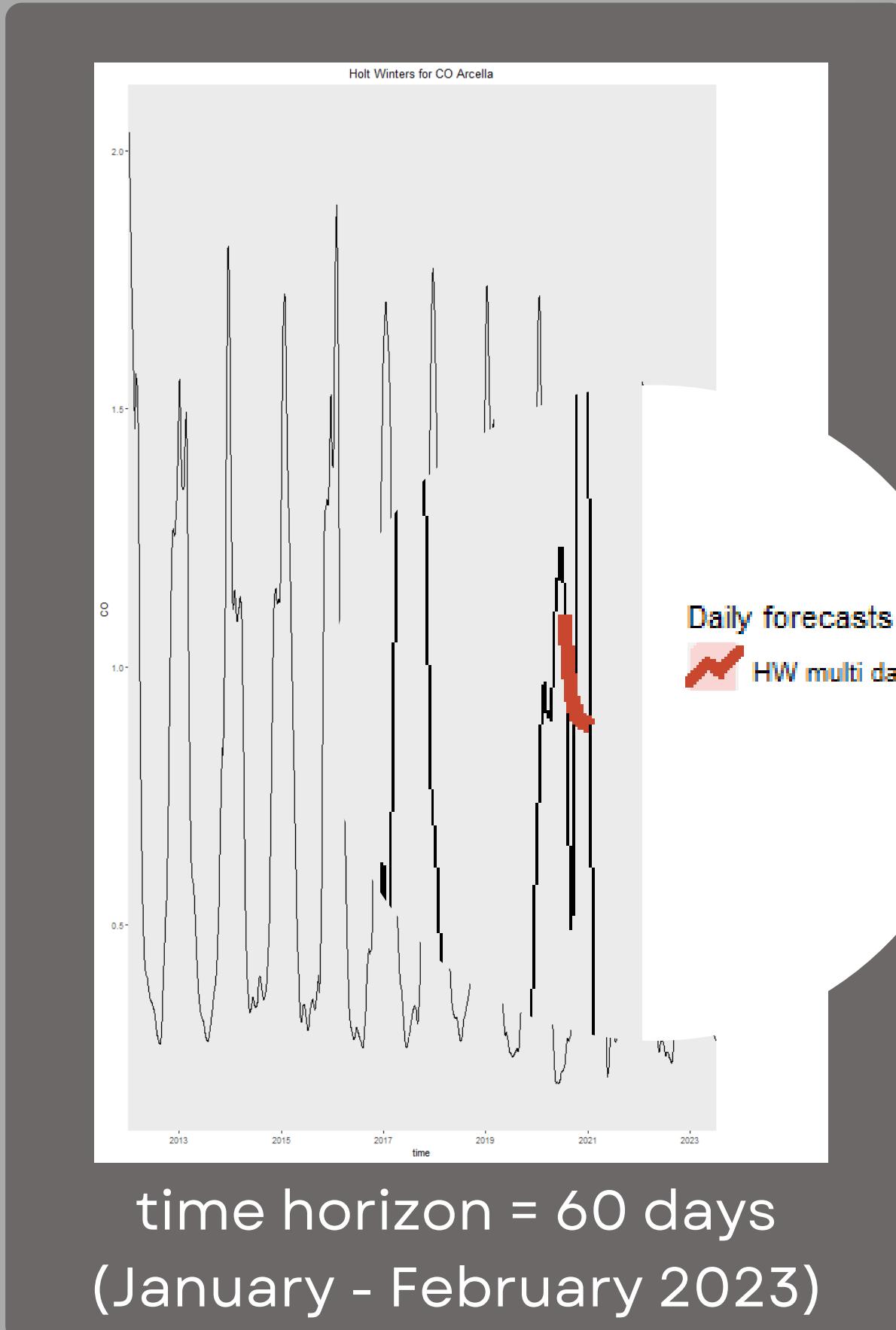
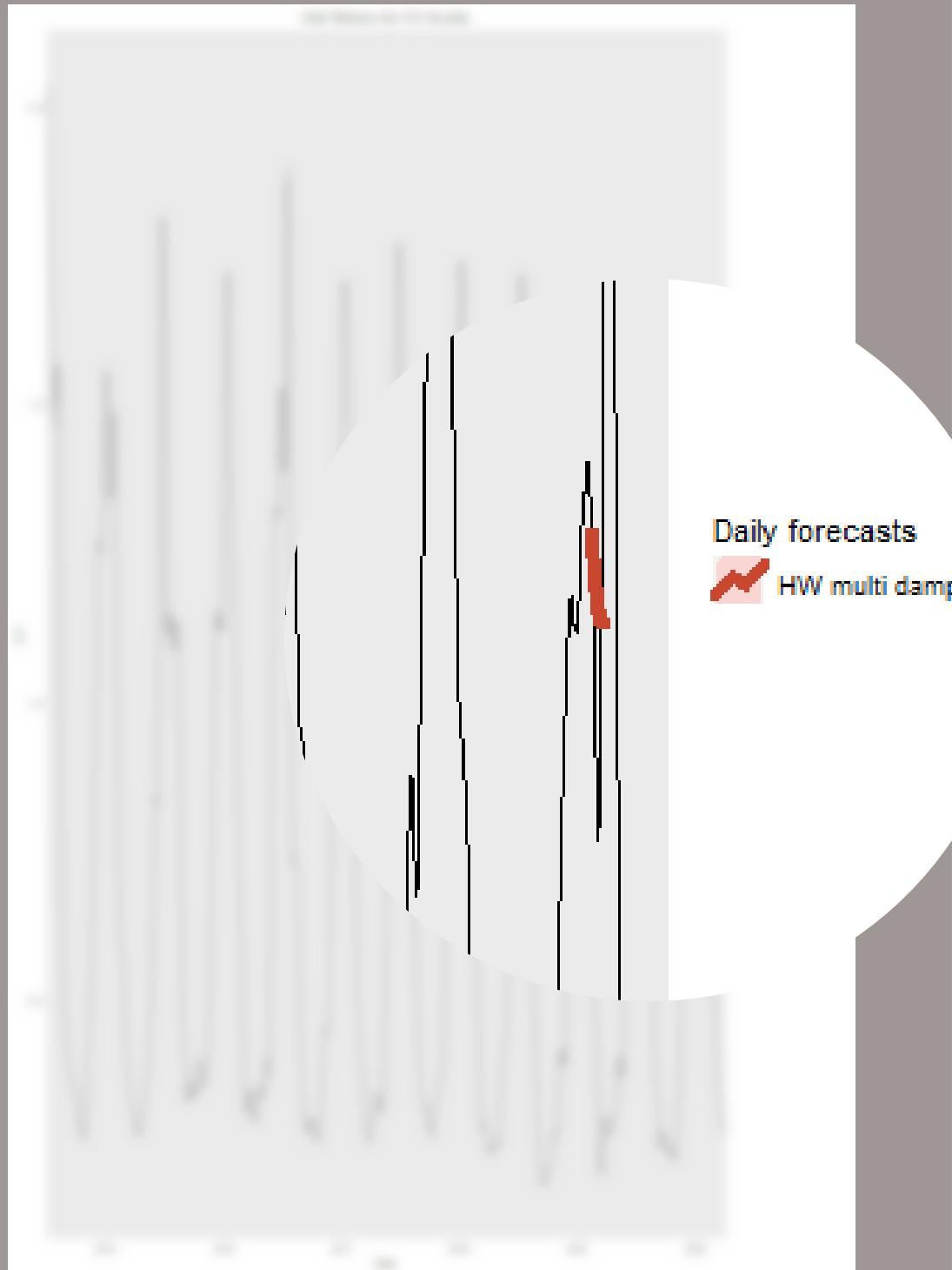
Finally, there are some pollutants whose data seems not to have been influenced by COVID-19, like PM10

This is probably due to the fact that the decrease in some activities (e.g. traffic) has led to an increase in others (e.g. domestic heating)

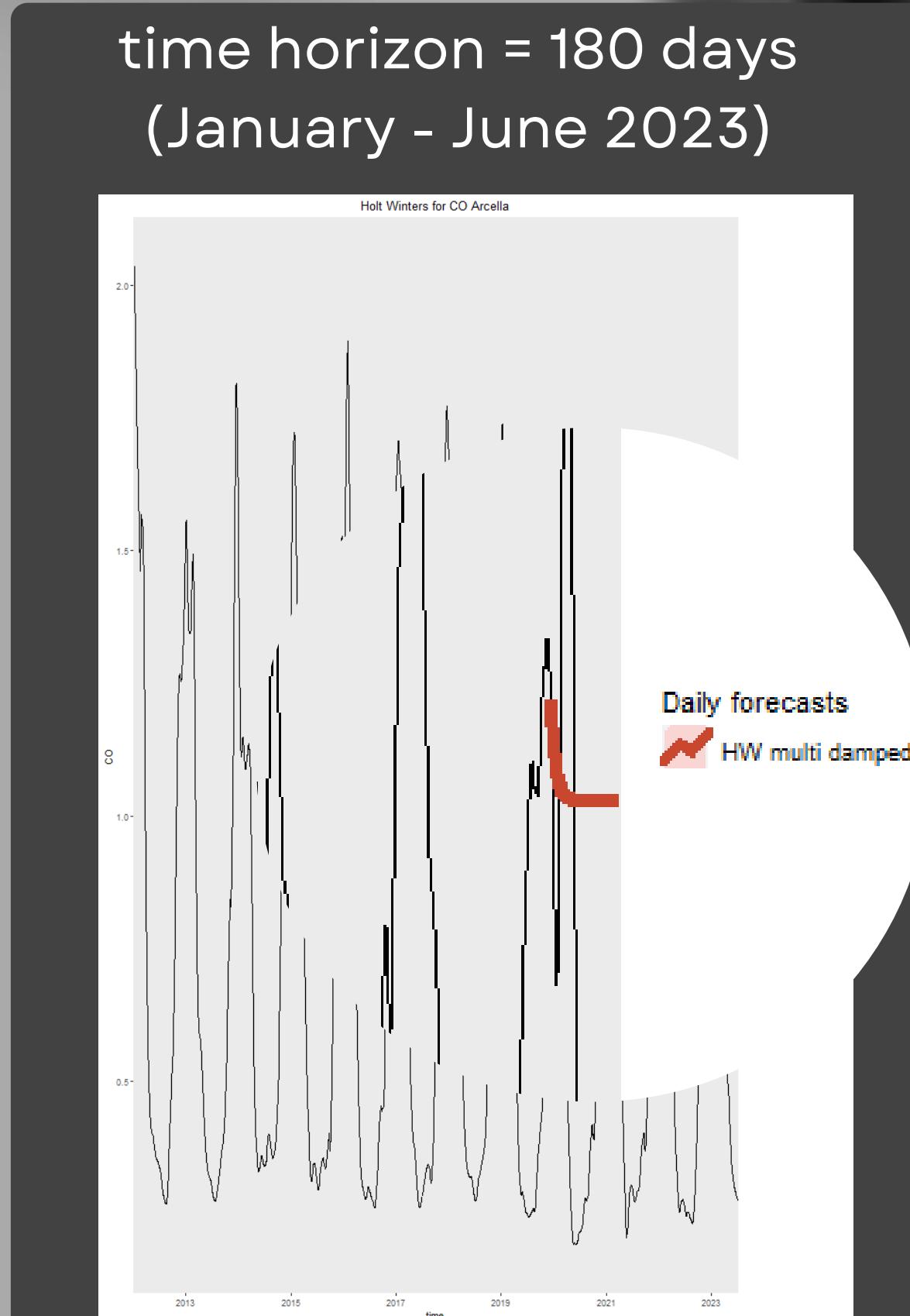
HOLT-WINTERS' MODELS

MODEL VISUALIZATION

time horizon = 30 days
(January 2023)



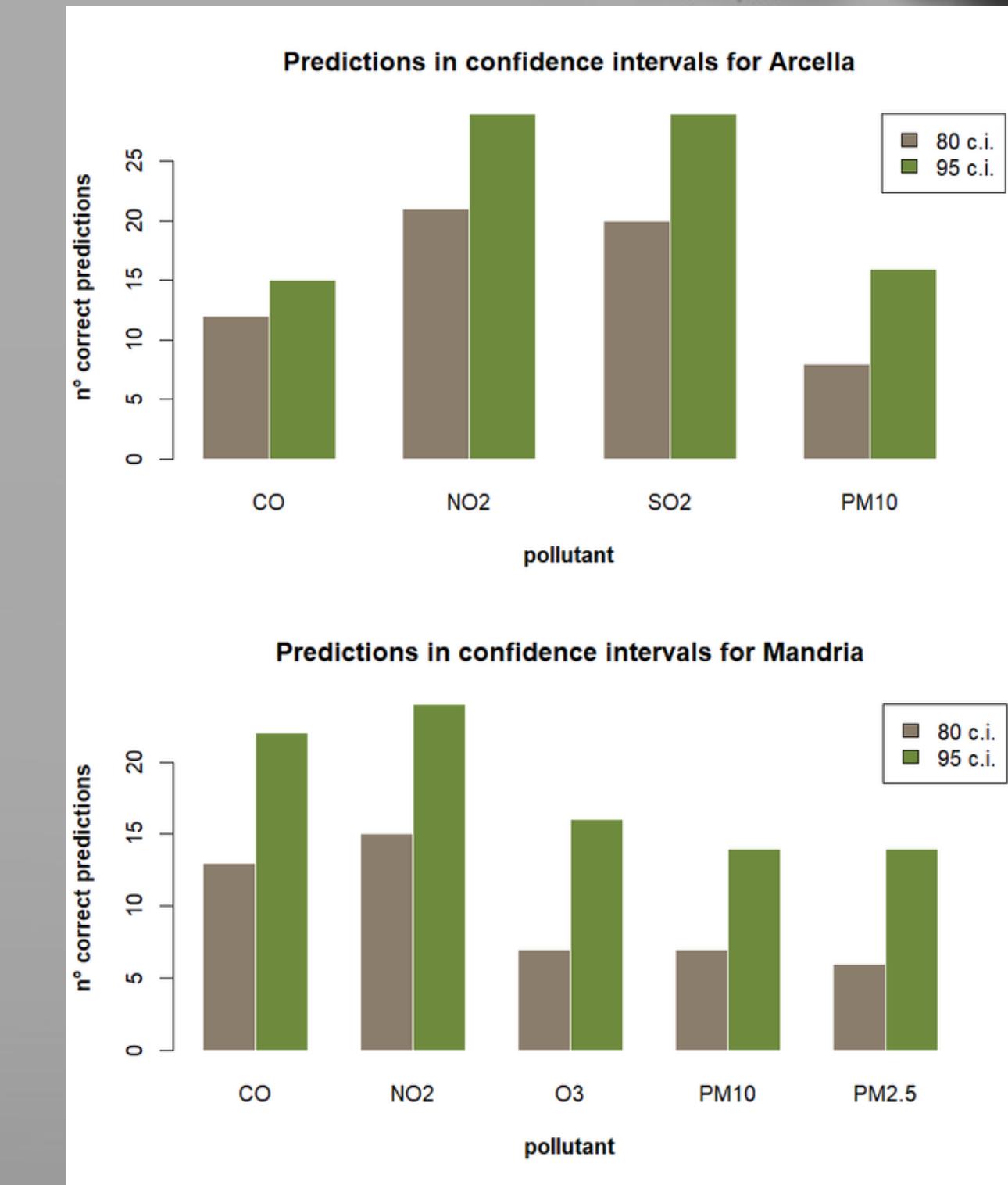
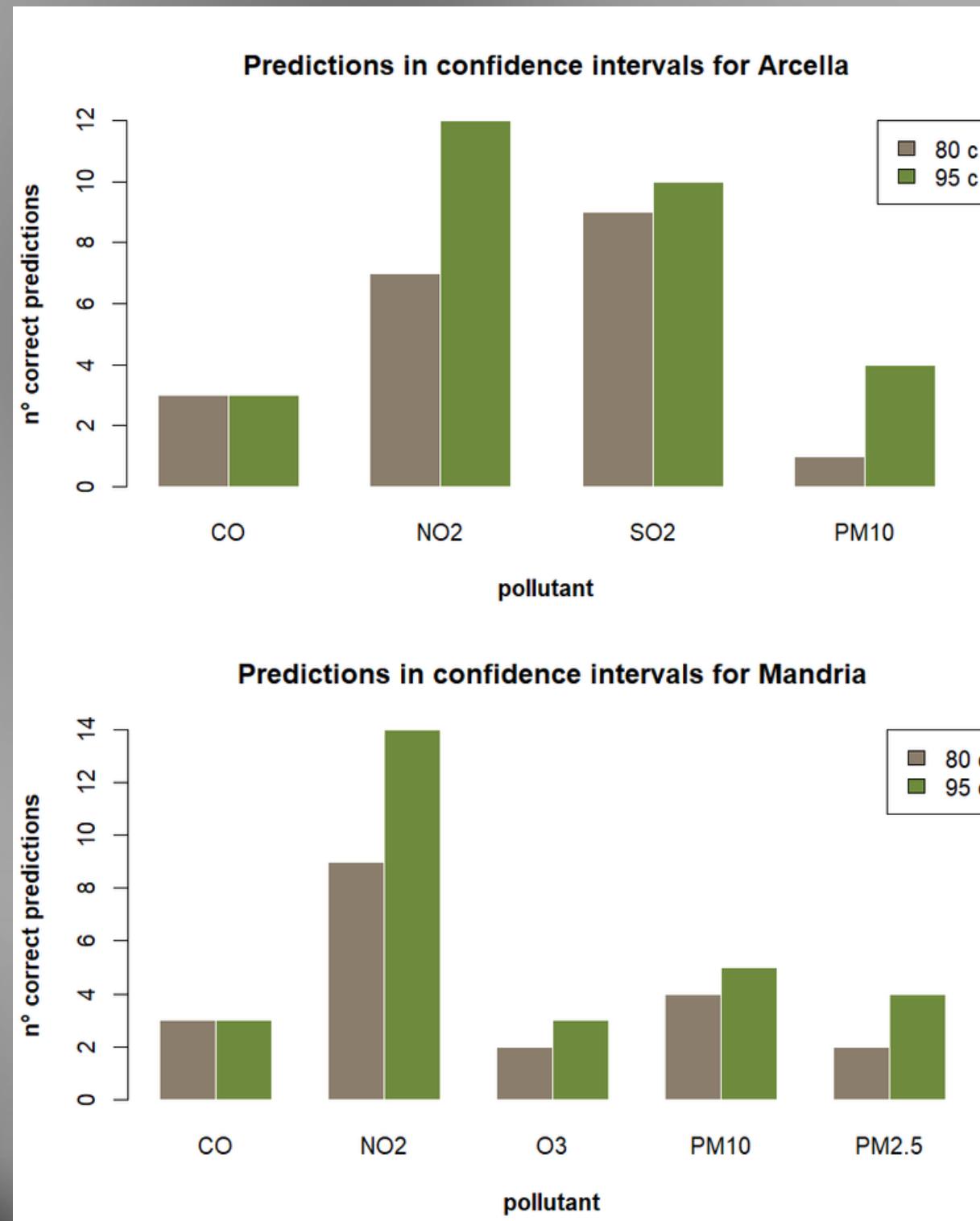
time horizon = 180 days
(January - June 2023)



PREDICTIONS INC.I.

REAL
DATA

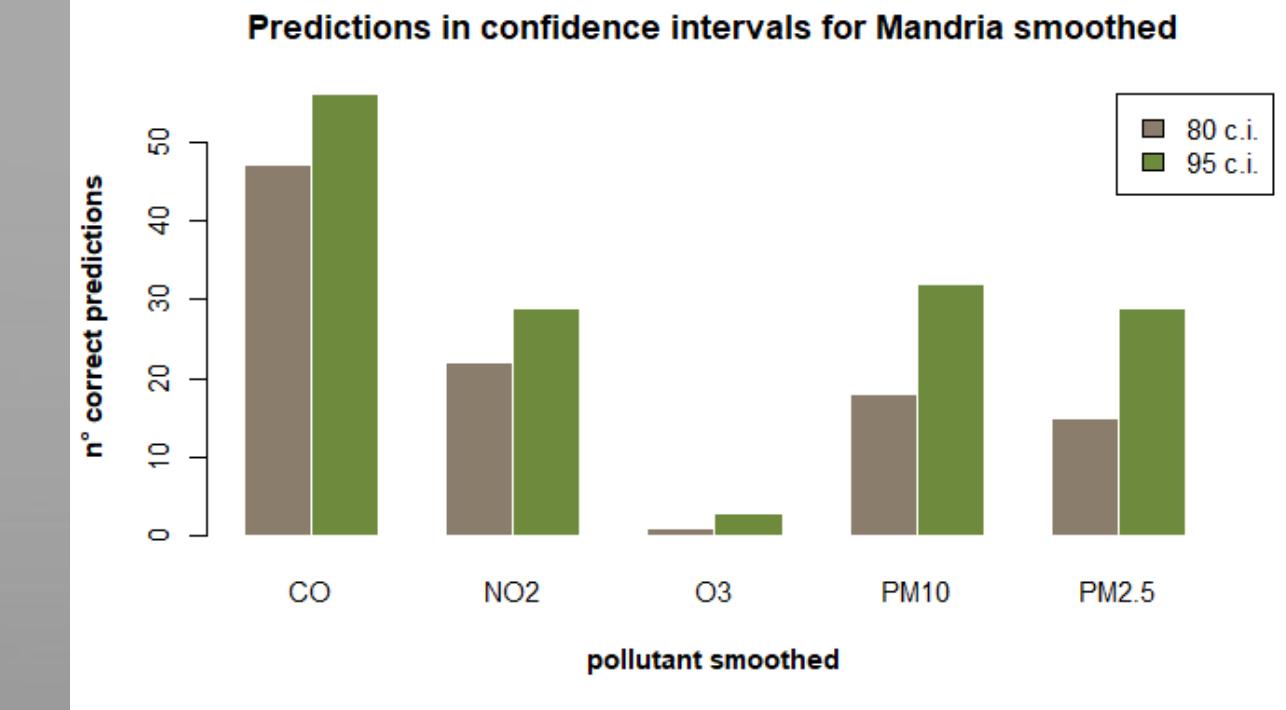
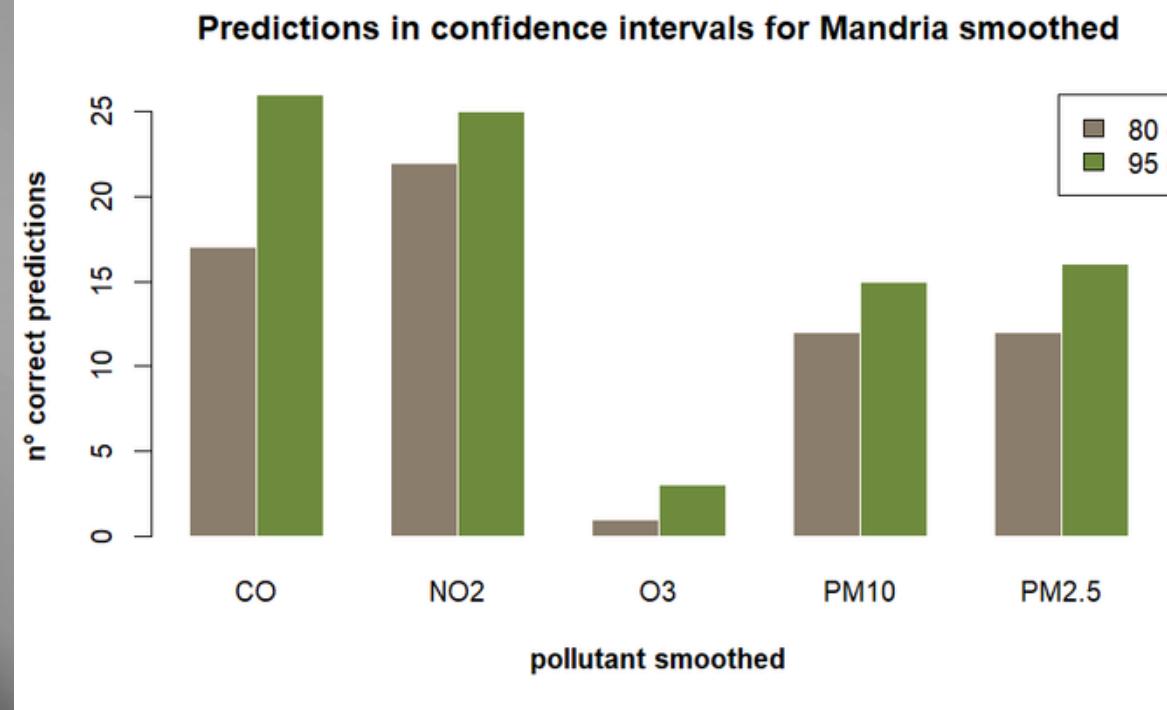
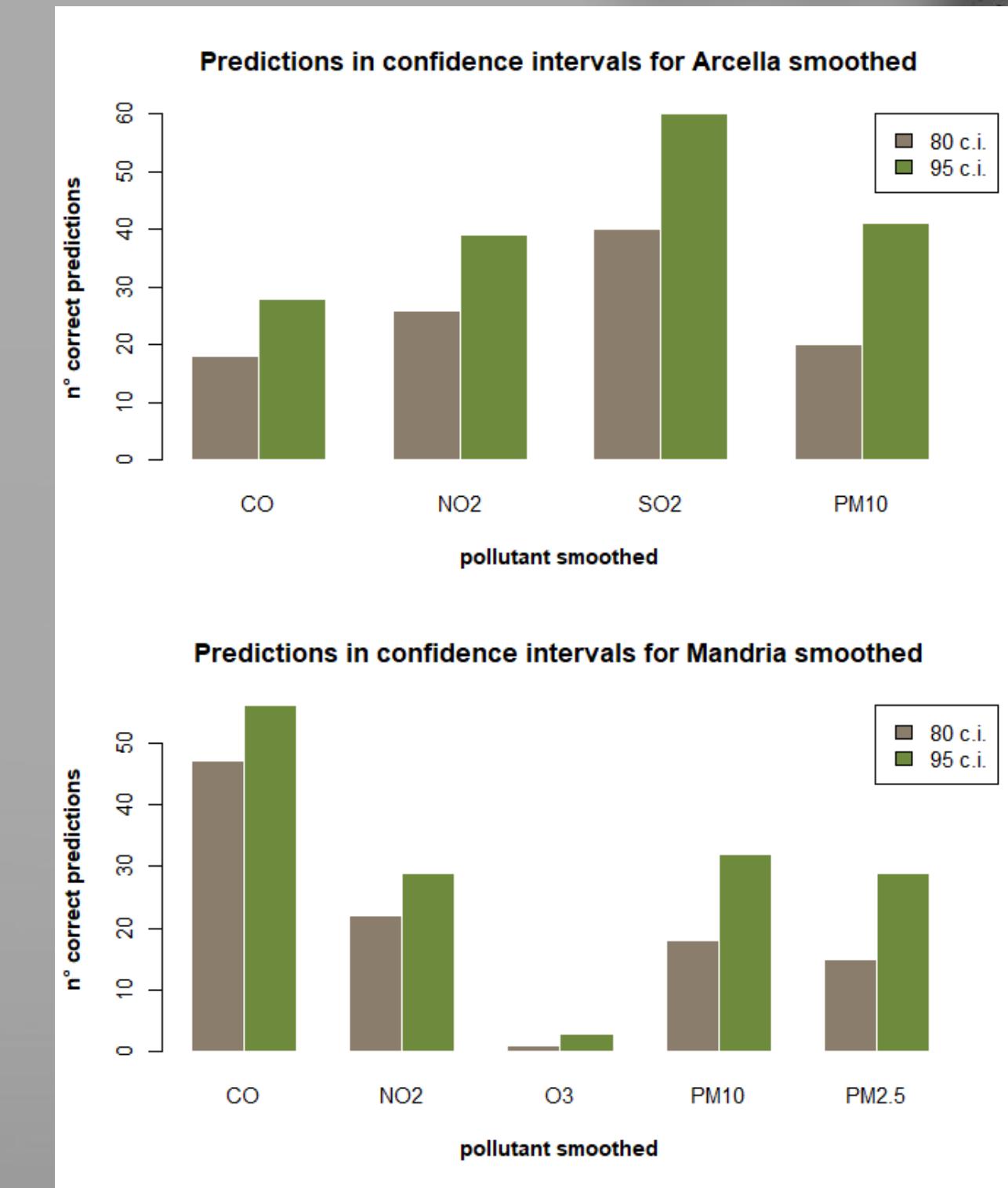
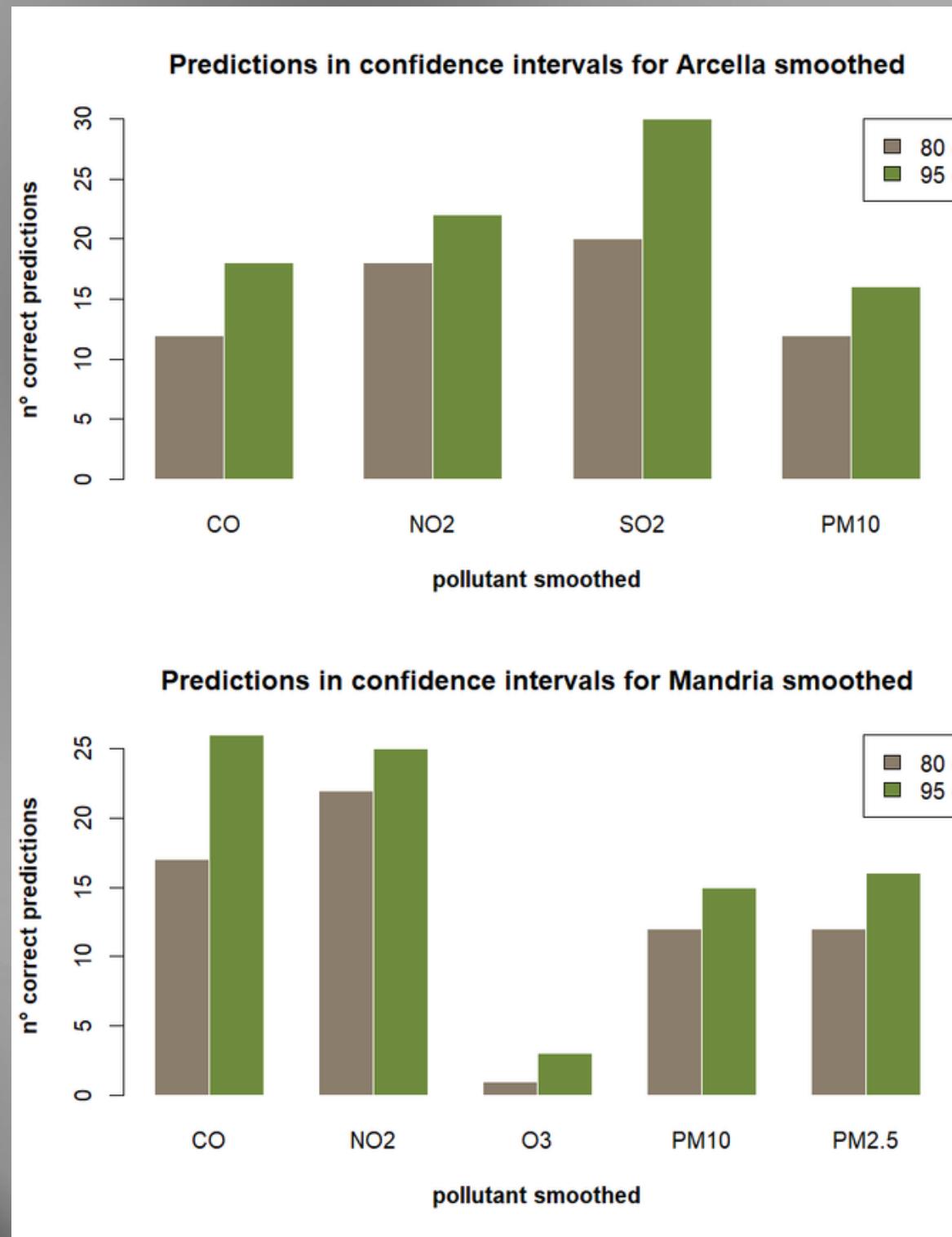
30 DAYS



SAYA 09

SMOOTHED DATA

30 DAYS

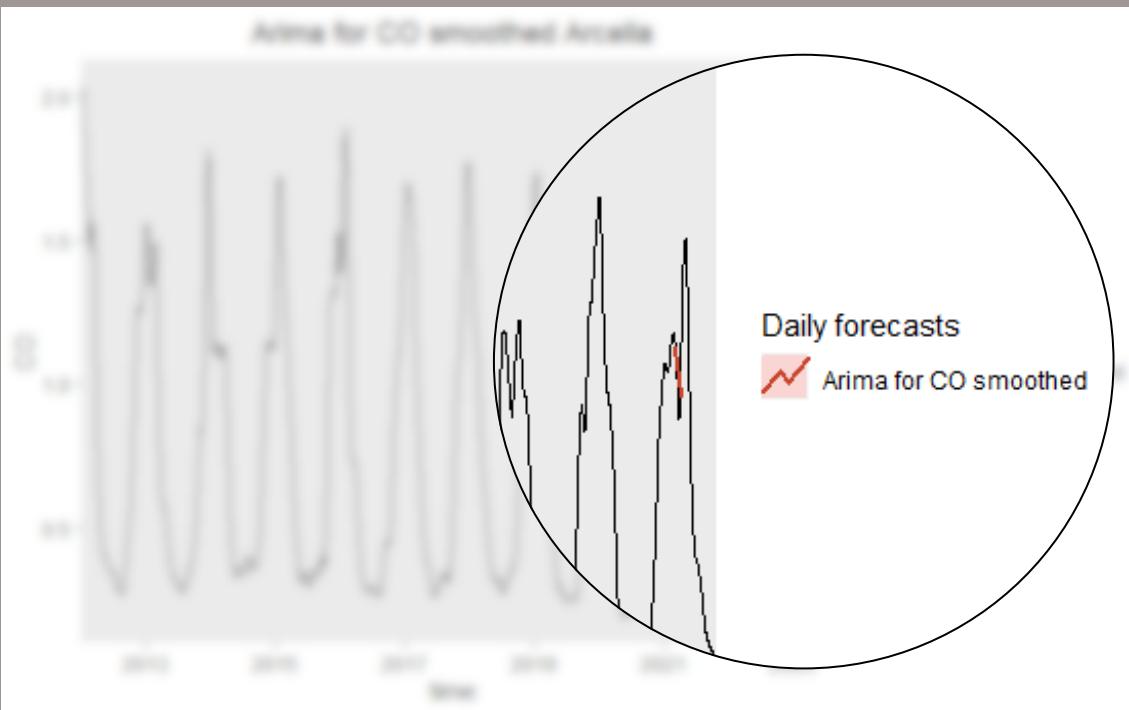


SAYING 09

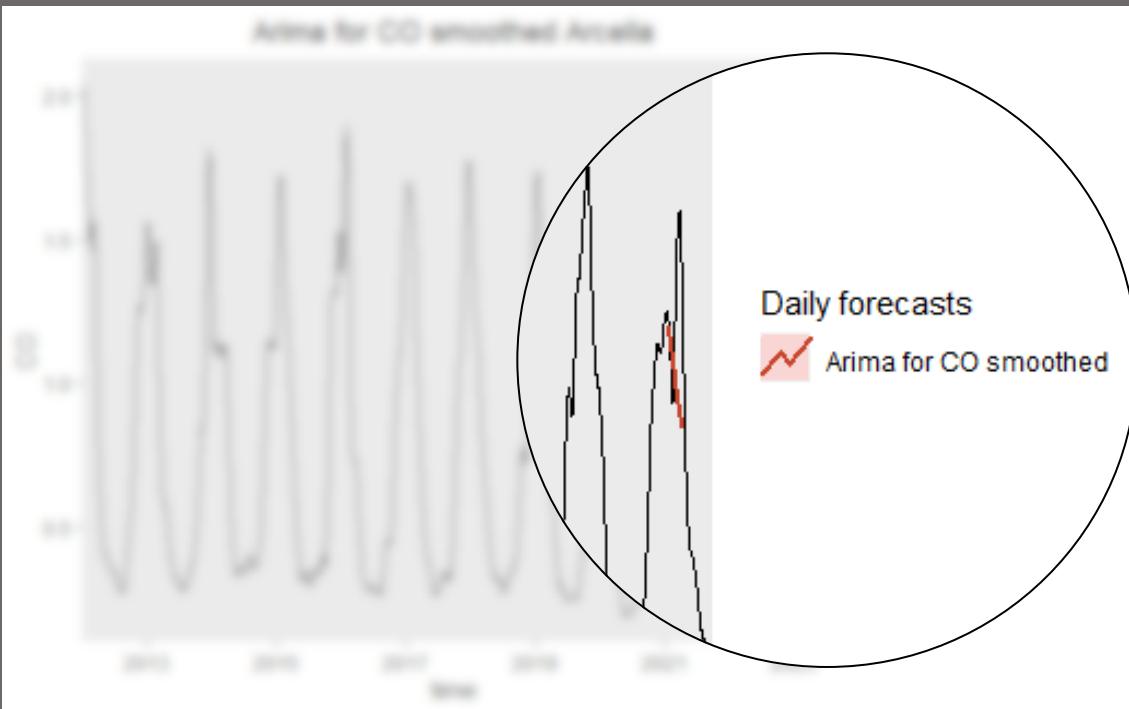
ARIMA MODELS

MODEL VISUALIZATION

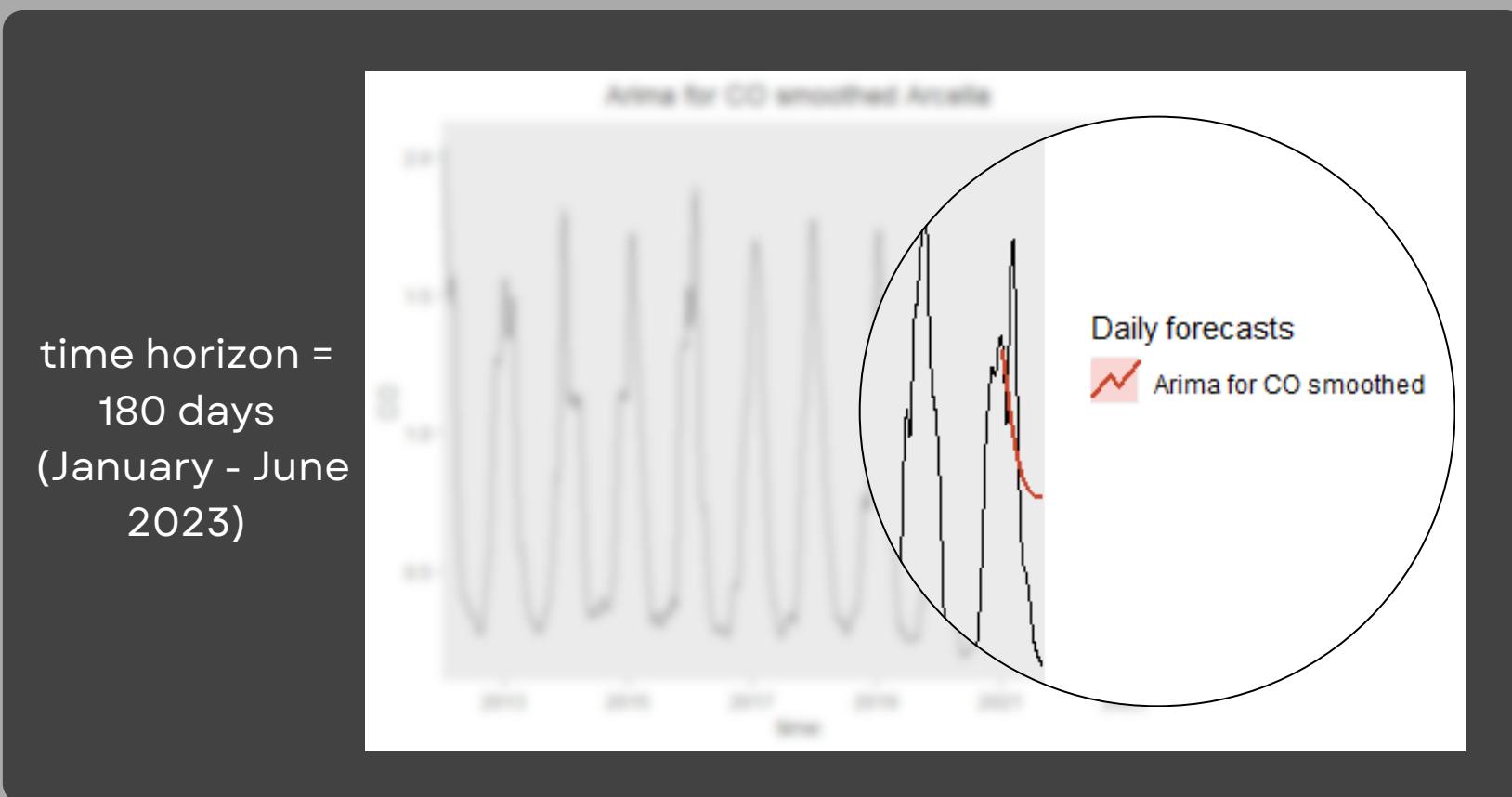
time horizon =
30 days
(January 2023)



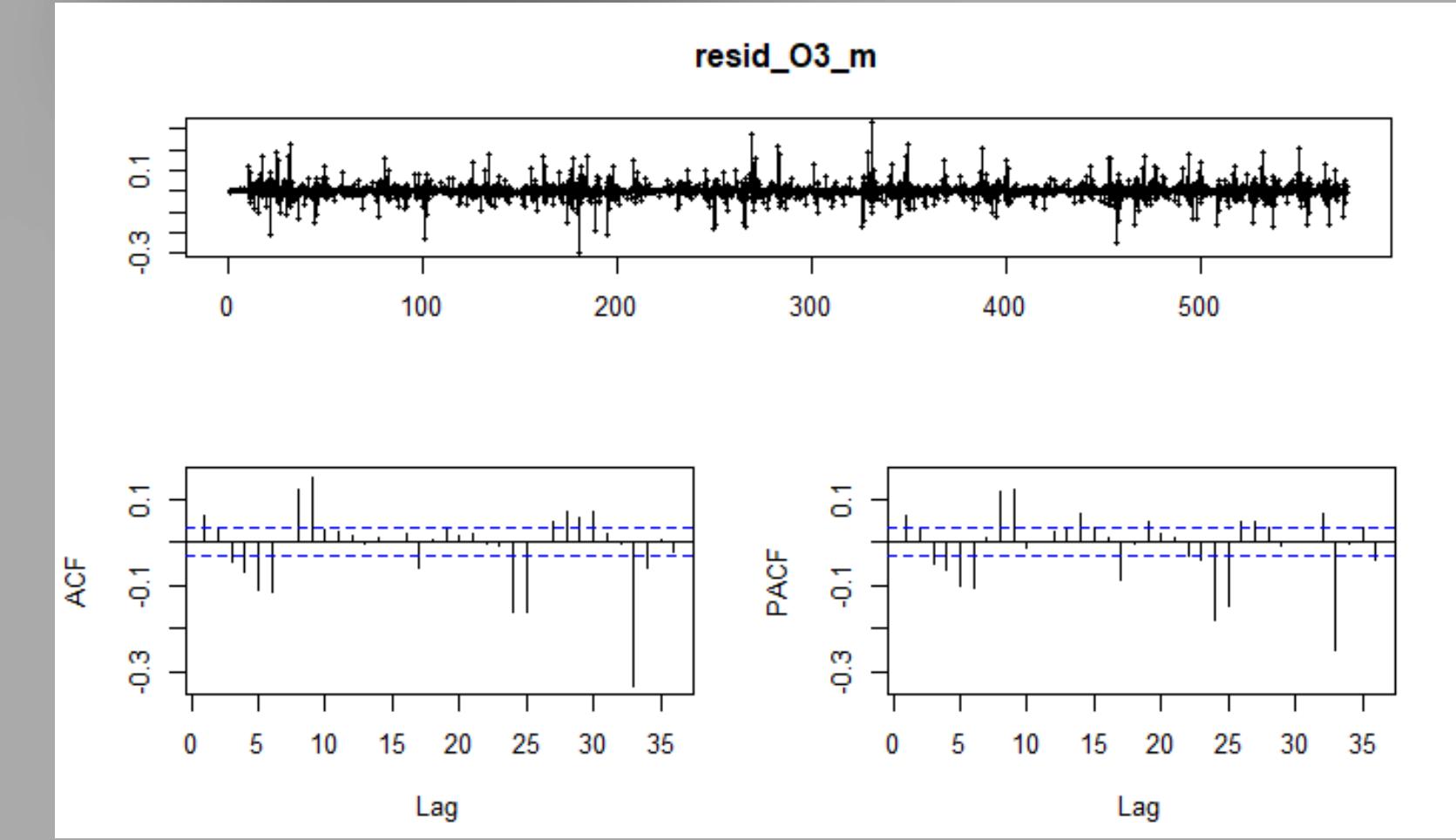
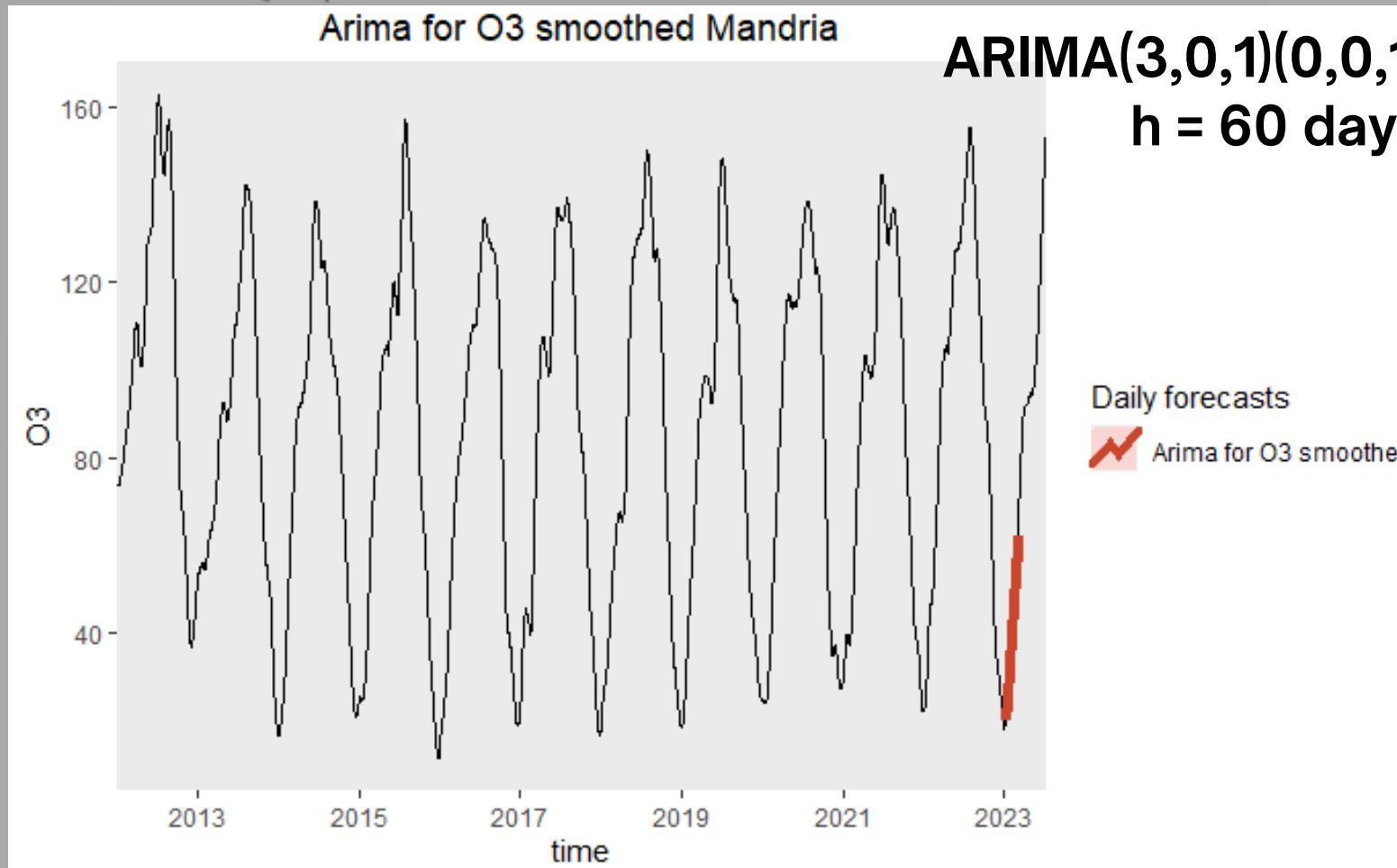
time horizon =
60 days
(January -
February 2023)



time horizon =
180 days
(January - June
2023)



RESIDUAL ANALYSIS



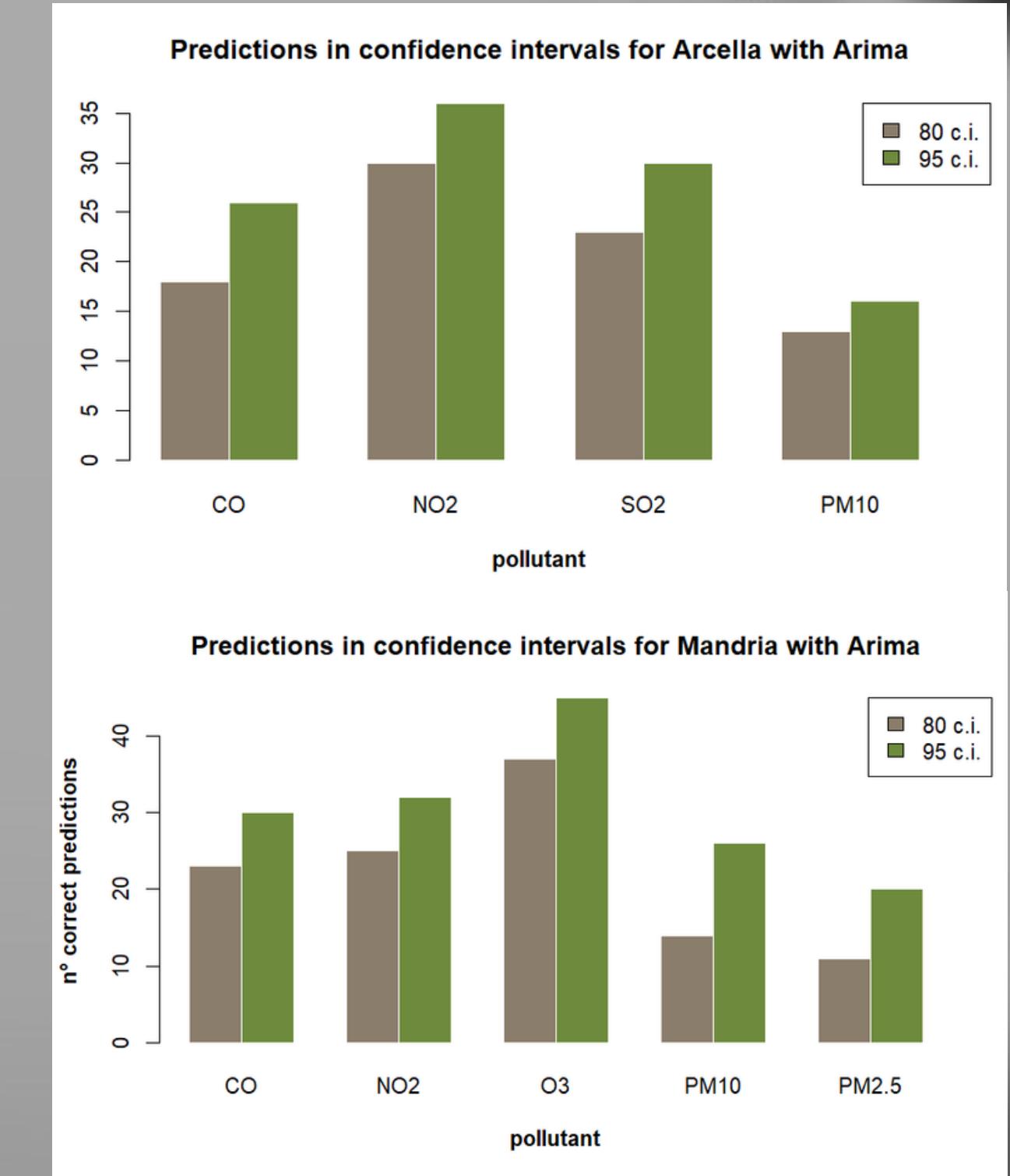
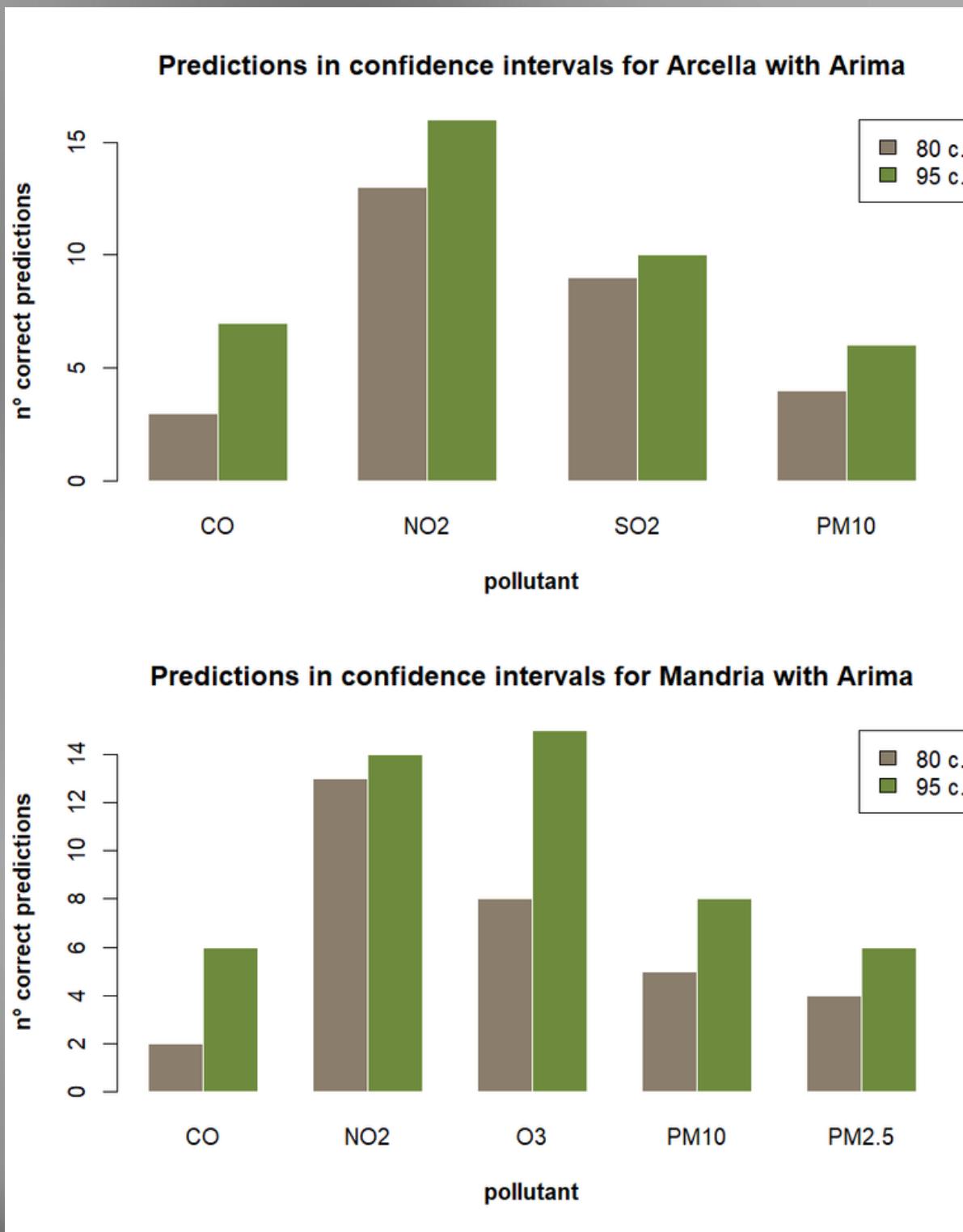
Even if, as in the case of the pollutant O₃ for Mandria, the prediction made by the arima model is very accurate, the residuals do not behave as white noise.

Probably due to the fact that there is a “minor” seasonality that the model is not able to capture.

PREDICTIONS INC.I.

REAL
DATA

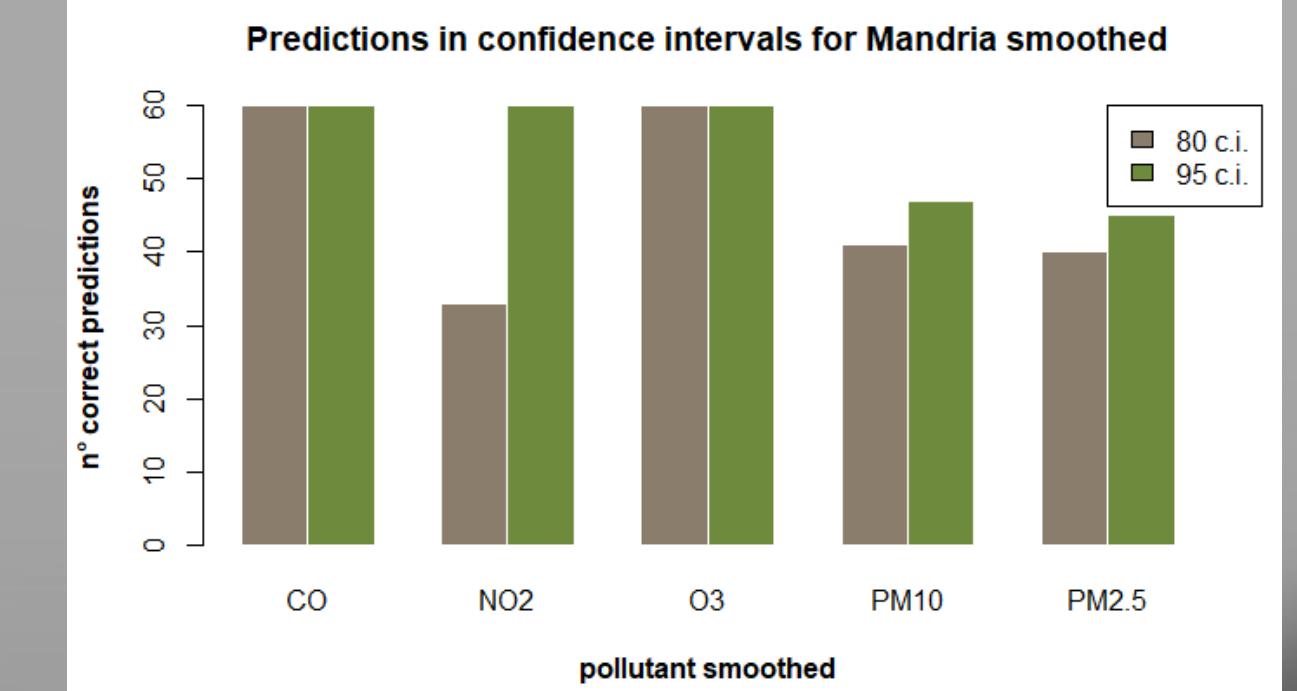
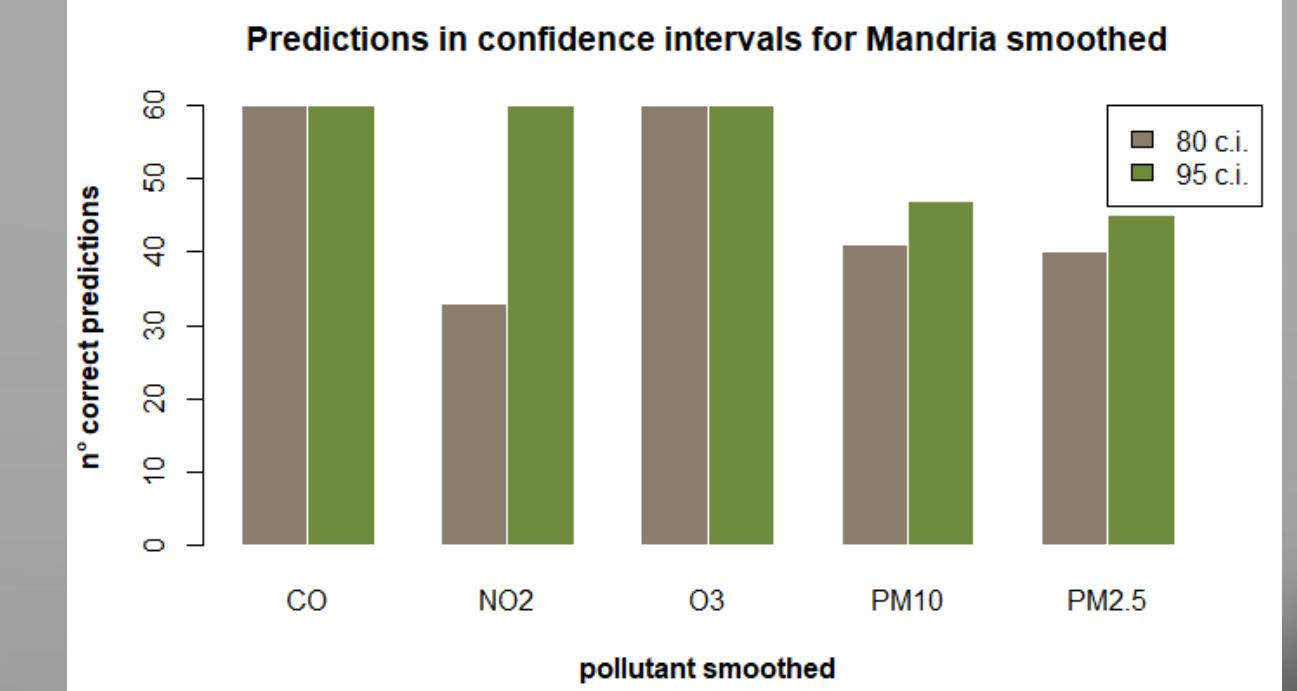
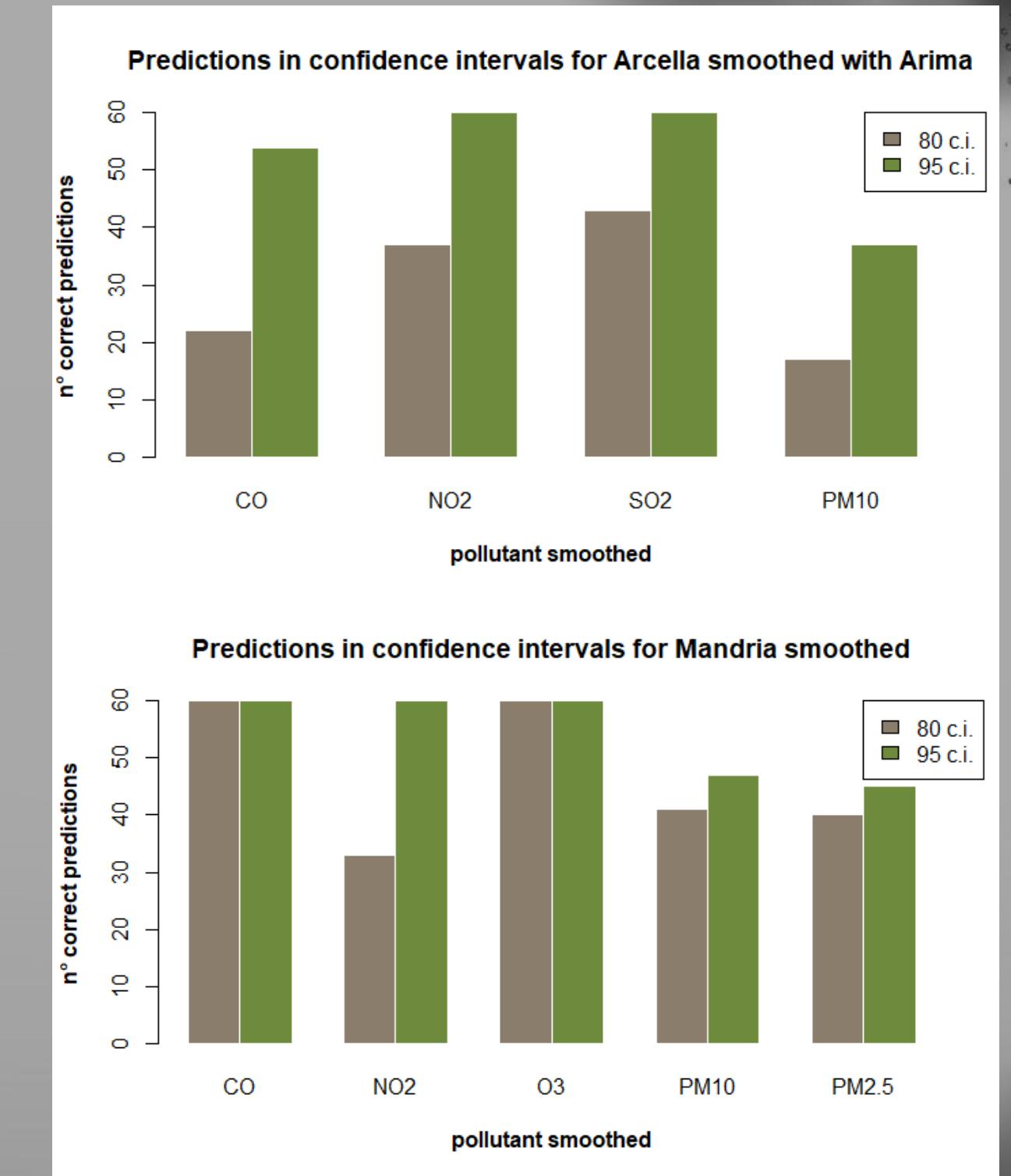
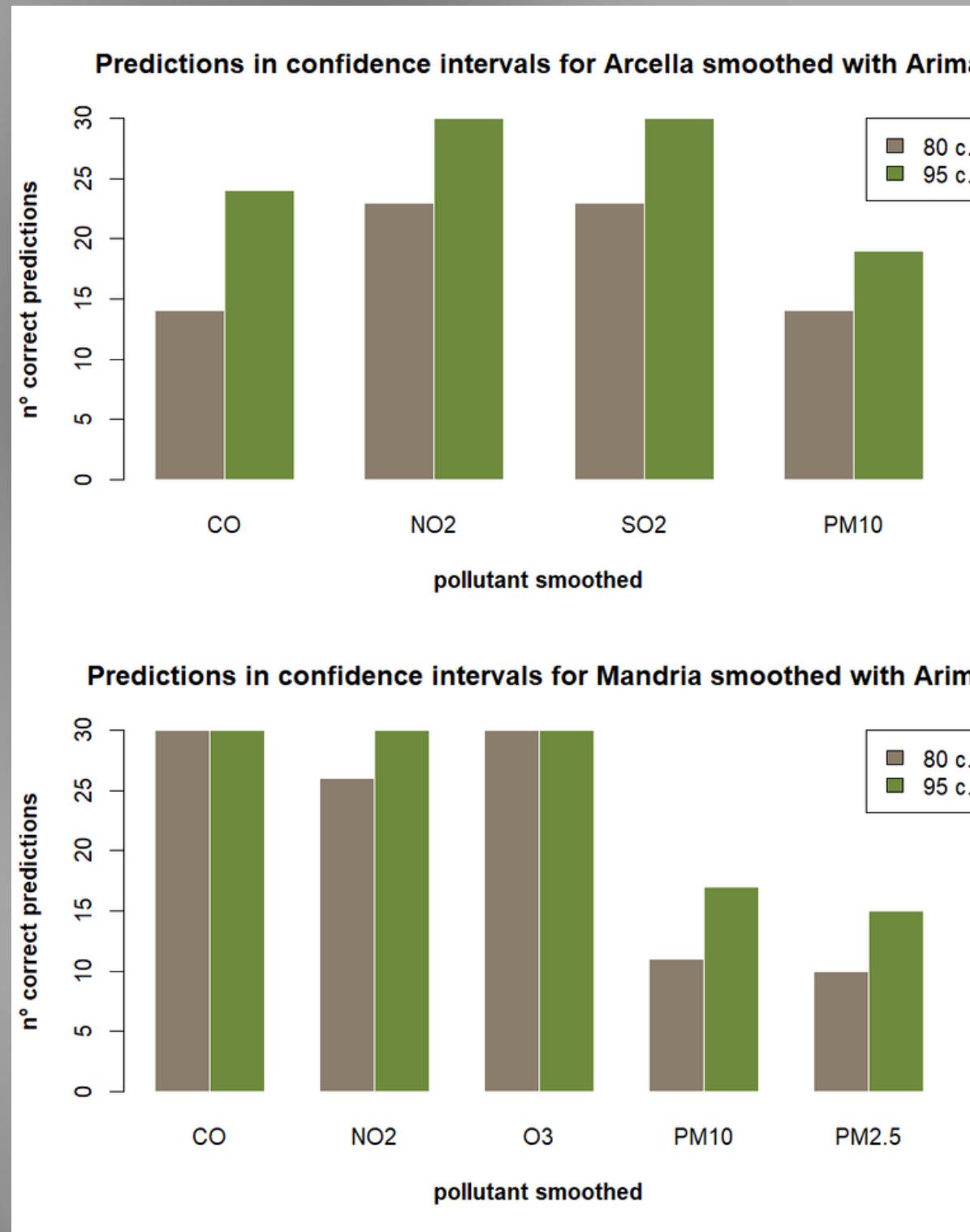
30 DAYS



SAG 09

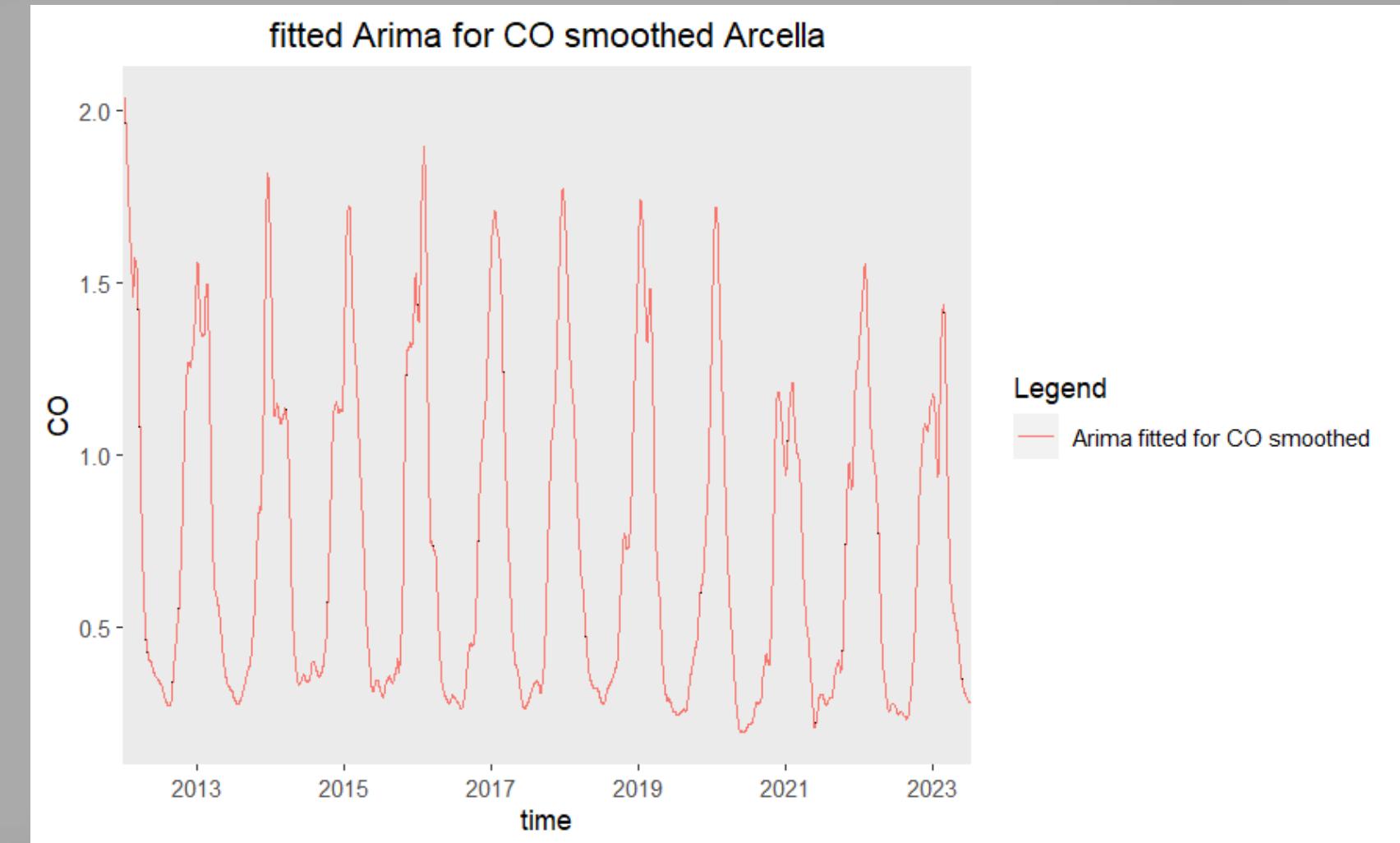
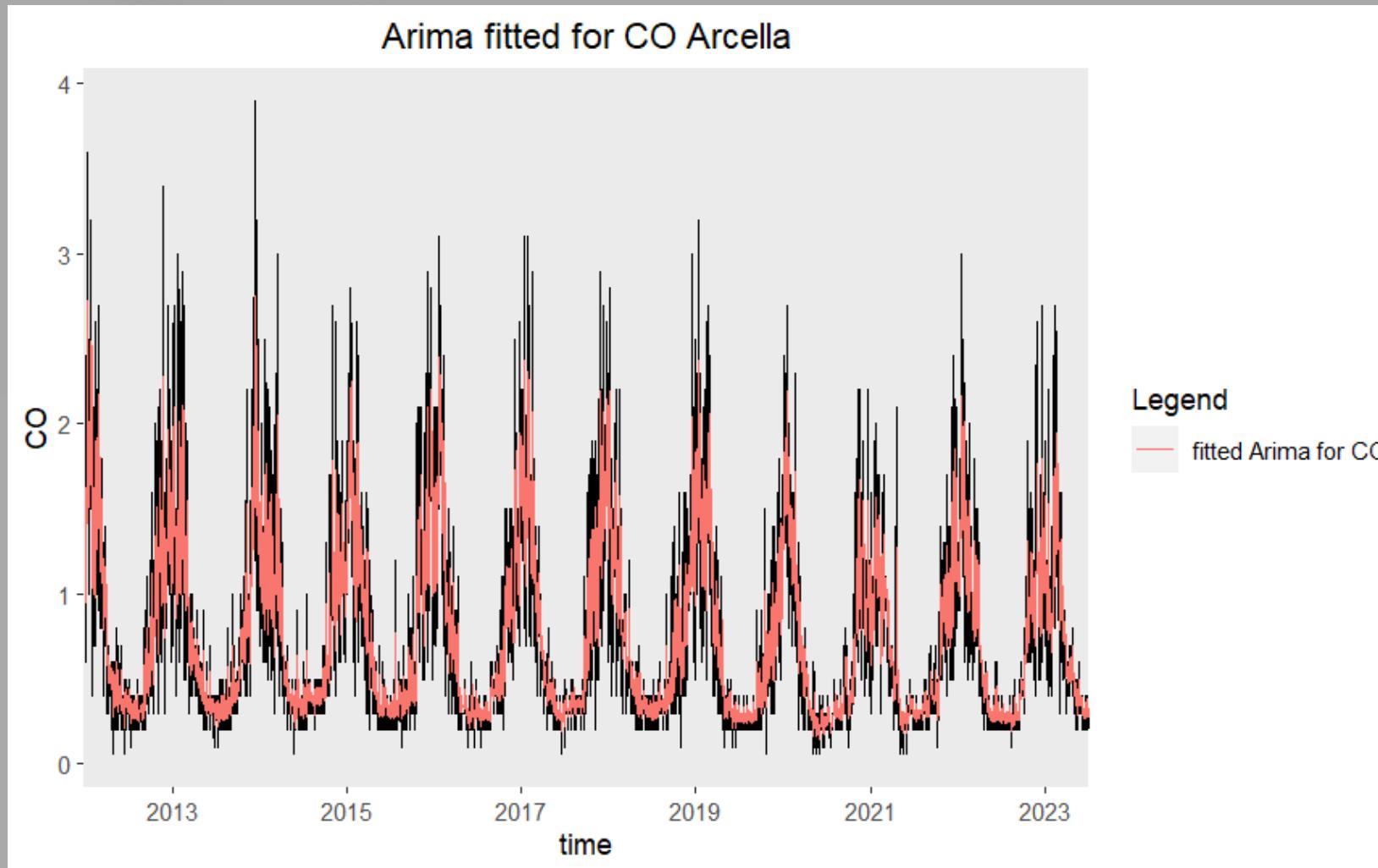
SMOOTHED DATA

30 DAYS



SAYA 09

ARIMA MODEL ON REAL DATA



We also try to fit the Arima model on the real data. This is the obtained model, in comparison the Arima applied to the smoothed data

ARIMA vs H-W

		Mean Square Error	
		ARIMA	Holt-Winters
ARCELLA	CO	0.4055	0.3624
	NO2	667.3033	695.9566
	SO2	28.7529	27.3162
	PM10	1302.119	1184.283
MANDRIA	CO	0.2169	0.2168
	NO2	622.7274	697.3313
	O3	181.9297	601.8722
	PM10	1019.565	1070.508
	PM2.5	865.3156	968.4596

CONCLUSIONS

1

Relevant number of infractions only for particulate matter (**PM10 & PM2.5**), polluting gases are often within the norms

2

Strong **yearly seasonality** for all pollutants. Slightly **decreasing trend** for some pollutants and a significant one for NO₂

3

Holt-Winters are able to capture correct behaviours for **~1 month**, **ARIMA** correctly captured up to **2 months**. **High noise** in the data makes **predictions uncertain** -> the models capture the trend much better than the specific oscillations

THANK YOU

