# Assessing the Impact of Driving Bans with Data Analysis

Lucas Woltmann[1] Claudio Hartmann[1] Wolfgang Lehner[1]

## 1 Introduction

Suspended particulate matter (SPM) is a significant problem discussed in current environmental research with an impact on the every-day life of many people. Our goal for the BTW 2019 Data Science Challenge (DSC) is to leverage information from available sensor data about SPM and assess the benefits and disadvantages of driving bans. Our application builds upon data of 57 sensors in the city of Dresden and 338 sensors in the city of Stuttgart. Each sensor tracks particle concentration, temperature, and humidity. Stuttgart has a particular interesting situation because of the driving ban for outdated diesel engines on roads in the inner city introduced in January 2019. This gives us the possibility to compare the effectiveness of driving bans not only over time but also between two cities. While we only analyze two cities exemplary in this report, we see high potential of applying our tools to other cities and scenarios. We think, this universality of our approach is an important factor in knowledge transfer. The applications are not limited to SPM analyses but can be extended for example to weather and climate research.

The following sections address the tasks and sketch our approaches. Section 2 discusses data cleaning and preparation. This includes a minute-wise aggregation and a linear interpolation of the data. Then, we show visualization techniques used to identify patterns for an assessment of driving bans (Section 3). Furthermore, we predict the future development of SPM with a technique called Cross-sectional AutoRegession (CSAR) [Ha19], developed in our group (Section 4). In Section 5, we condense the results of previous analyses and define no-go areas with the highest particle concentrations and also identify the reasons for particular no-go areas. We close our work with a short conclusion in Section 6.

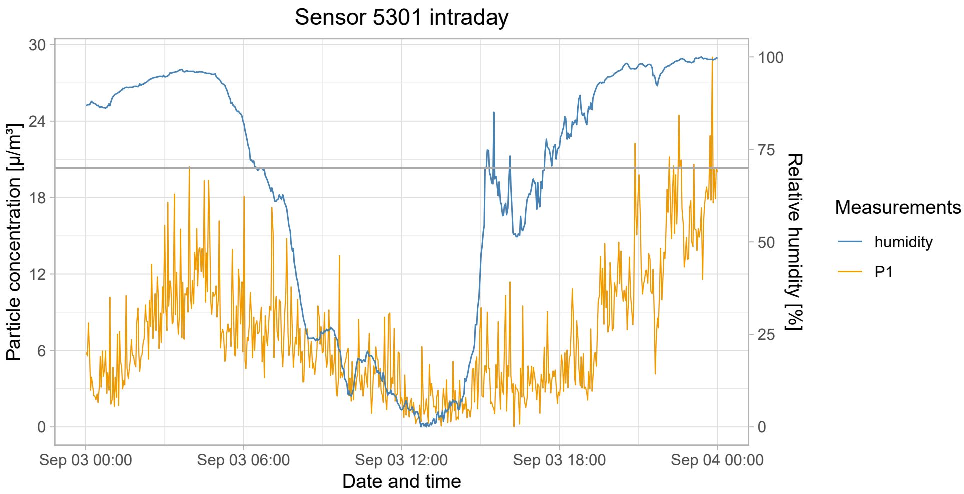
## 2 Data Cleaning and Preparation

First, the data of all sensors in Dresden and Stuttgart is collected. We use a nearest neighbor search around the city center of both cities with a radius of 10km to retrieve all sensors in both city areas.

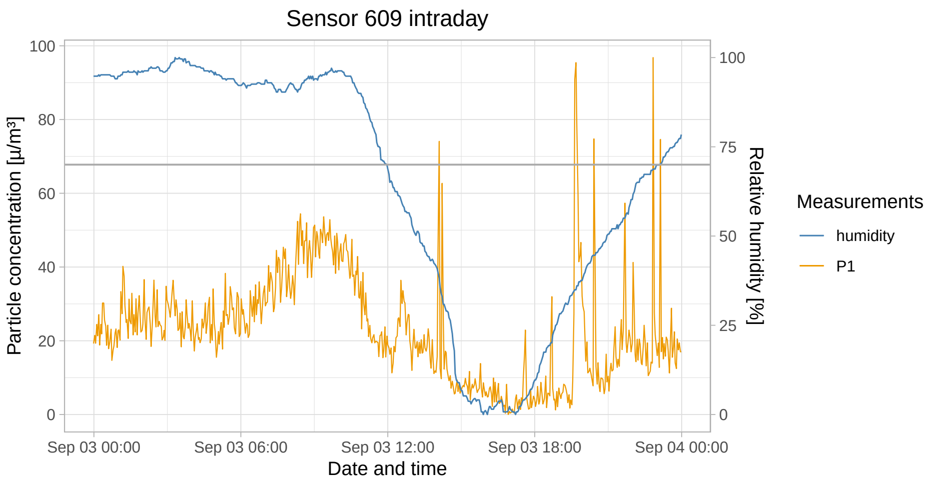
A general problem is the division of sensor types into SPM sensors and temperature/humidity sensors. To construct a common base for the analysis, we need to integrate the SPM sensors

[1] Technische Universität Dresden, Database Systems Group, 01062 Dresden, Germany
<firstname.lastname>@tu-dresden.de

doi:10.18420/btw2019-ws-31




(a) Comparison Dresden




(b) Comparison Stuttgart

Fig. 1: Particle concentration P1 compared to humidity.

and the humidity/temperature sensors into one data set. Therefore, we aggregate the data for each sensor type to a minute-wise time scale by averaging all measurement within every minute for each sensor. This provides a common minute granularity for all sensors. The next step is to merge the temperature and humidity data with the particle concentration data by sensor location and time. Note that this is only possible in a standardized minute granularity in both sensor type data.

From the resulting table, we remove all particle concentrations where the humidity is larger than 70% because above this value the particle sensors do not provide reliable reads [No15]. This leads to sparse data for both Dresden and Stuttgart because both cities have a rather humid river climate. Additionally, both cities are located in a valley. This problem is shown in Figure 1 which compares the particle concentration P1 to the humidity for one sensor in each city close to the river. The 70% limit for the humidity is marked with a black line.

Finally, we impute all missing values with linear interpolation. This has proven to be a very robust approach in the field of time series analysis [Mo15].

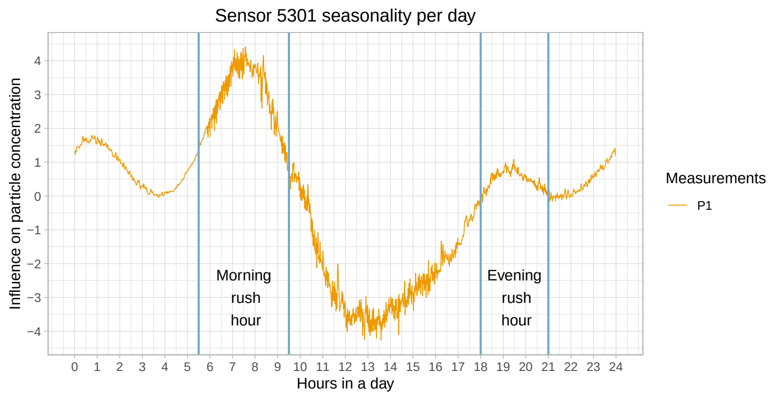
## 3 Visualization and Patterns

Visualization is a common approach to find interesting patterns. We present two different kinds of plots helping with the identification of structure in the data. First, all time series get split into trend, season, and residual components with time series decomposition [Cl90]. The seasonal component details peaks and valleys in the time series which occur in a regular pattern. This component shows the recurrent influence of each point in time on the time series, as shown on the y-axis.

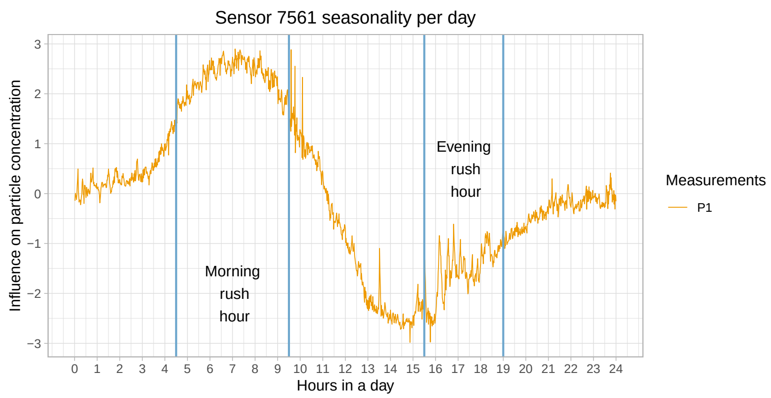
For example, we can clearly show the morning and evening rush hours as two peaks in the seasonal component of the sensor data in Figure 2. Both sensors are located in the respective center of the given city. Whereas the sensor in Dresden in Figure 2a has clearly visible peaks for these two events, the sensor in Stuttgart details the evening rush hour only in a noisy pattern for 2018. We can now compare the regular pattern of the rush hours before and after the introduction of the driving ban in Figures 2b and 2c. The last plot has no characteristic structure for neither morning nor evening rush hour. There are no recurring rush hour patterns in the data for January 2019. This can be ascribed to the driving ban because less cars in the city center also means less traffic during rush hours or even no rush hours at all. The sheer absence of peek concentrations can be seen as an improvement in air quality. But if we have a look at Figure 3, where we took a sensor near the border of the driving ban area, one can see a negative impact of the restriction. Following the introduction of the ban, the evening rush hour has a much larger impact on the sensor's surrounding area. Figure 3b details an influence value for the evening rush hour (y-axis) 10 times higher than in Figure 3a. This can be an indication that the driving ban does in fact make the inner city cleaner but only at the expense of flooding the surrounding areas with more cars and pollution. This negative effect can be observed especially during rush hours.

As a second visualization, we will map the distribution of particles over both cities. The plot in Figure 4 highlights particle hotspots in the cities. The darker the shade of an area, the higher the particle concentration. For visualization, the concentration is smoothed using cubic splines [Pe84]. The black points on the map are the locations of the SPM sensors. Furthermore, we will use this plot as one of the tools to identify no-go areas of high particle concentration in Section 5.

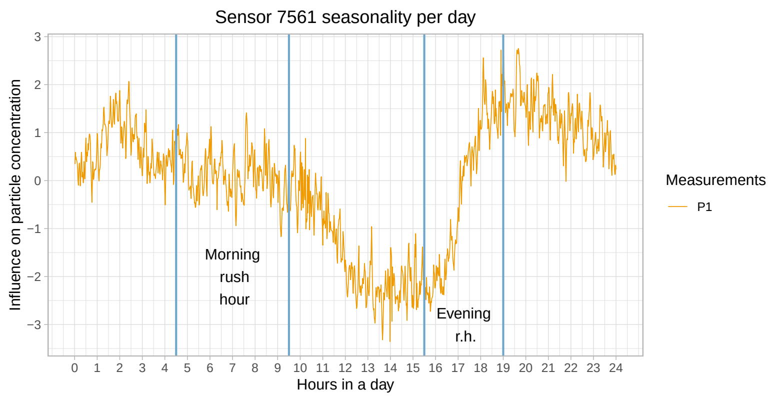
For now, we can compare the distribution of particles over Stuttgart before and after the ban. Figure 4c compared to Figure 4b gives the same intuition as our last visualization. The general pollution got less with the ban. This can be seen via the lighter hue of the complete city area and the smaller areas with dark hue. Whereas the particle pollution in the inner city got better, the surrounding areas suffer from heavier pollution. The hue of the concentration coloring is darker on the edges of the city of Stuttgart. As mentioned before, the reason




(a) Rush hours in Dresden




(b) Rush hours in Stuttgart in 2018

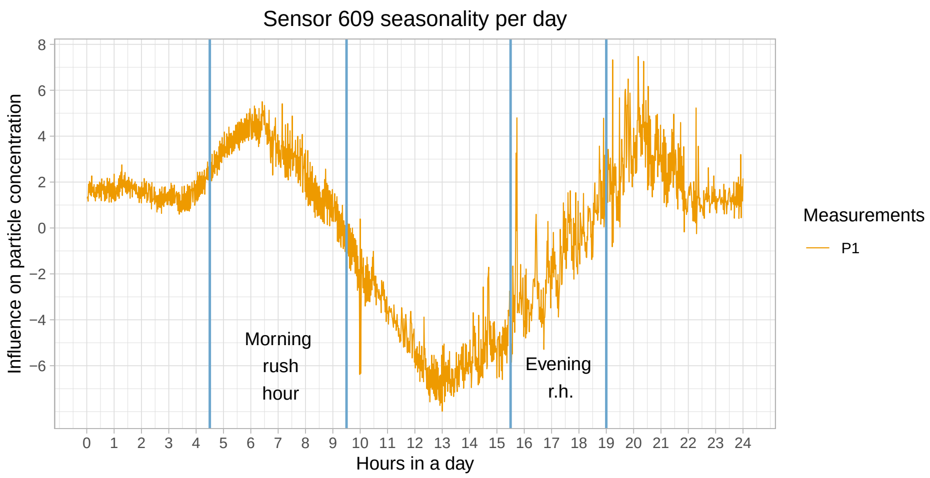



(c) Rush hours in Stuttgart in January 2019

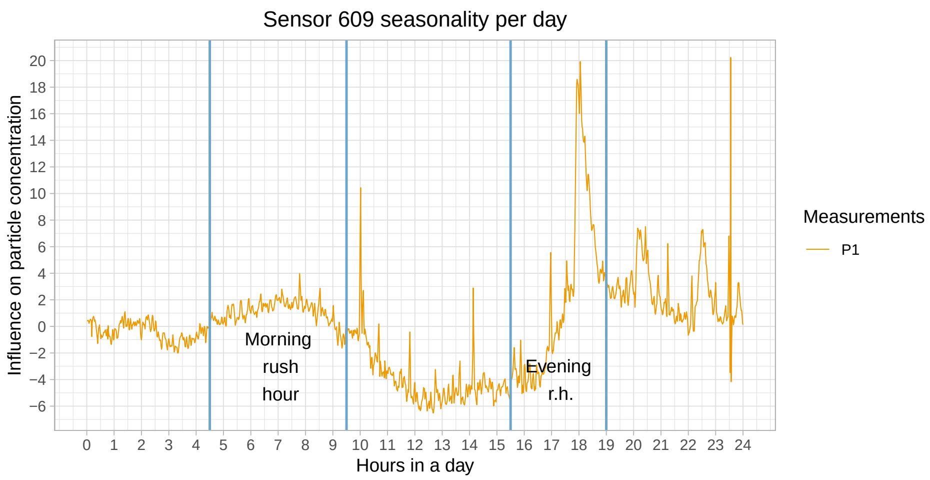
Fig. 2: Rush hour patterns for both cities.

for that might be the diverted traffic now filling the streets around the city creating more pollution there.

If we compare Dresden, a city without a driving ban, to Stuttgart, we only need to take a look at the data for 2018 because there is no active change in pollution in Dresden. Dresden has a lighter coloring and therefore less pollution because of having less traffic in the city area than Stuttgart [IN17]. The data in Stuttgart is less sparse because there are more sensors available. This leads to a finer gradient in color for the visualization and a more detailed view on pollution compared to Dresden. Nevertheless, the similar topology of both cities cause the same problems. Both cities have large areas of congestion on their access roads to the Autobahn (see Section 5). The valley position of both cities can dictate long access road to the Autobahn because building interstate roads near or in the valley might be either very expensive or very polluting. Long access roads mean more potential for heavy traffic.

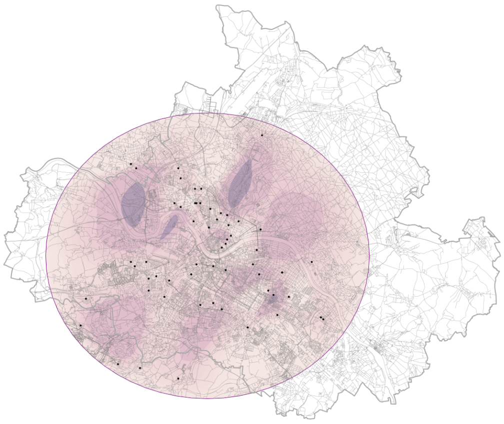



(a) Rush hours in 2018

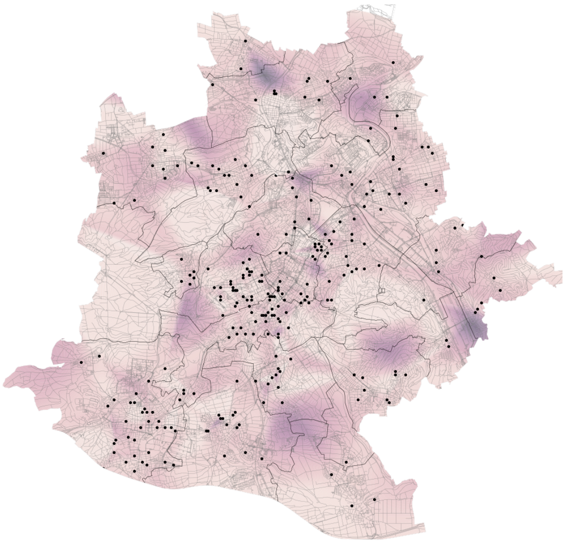



(b) Rush hours in January 2019

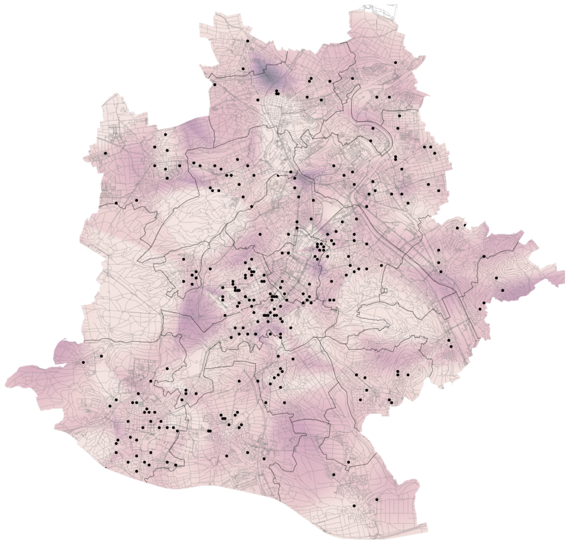
Fig. 3: Rush hour patterns for sensor outside the driving ban area.



(a) Particle distribution over Dresden in 2018.



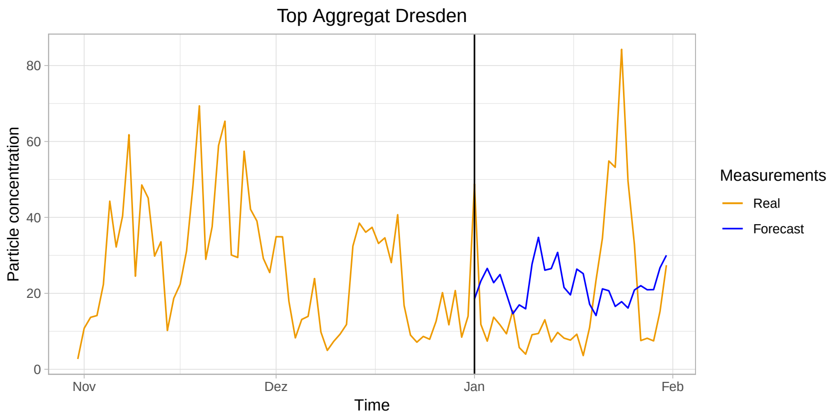
(b) Particle distribution over Stuttgart in December 2018.



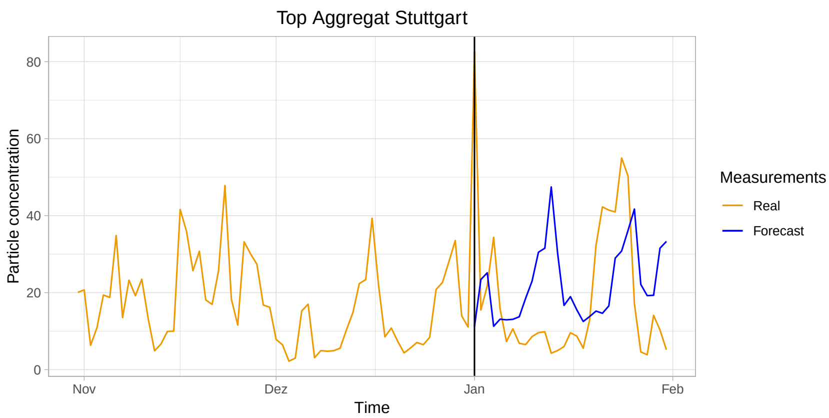
(c) Particle distribution over Stuttgart in January 2019.

Fig. 4: Particle distribution over the cities.

Even though the valley structure of both cities is hardly visible in the data visualization, one could argue that in Stuttgart one can see the southwest to northeast valley around the Nesenbach and the Neckar valley east of the city center in a darker hue. This assumption does not hold for Dresden, mainly due to the sparsity of sensors.




(a) Forecast January 2019 Dresden




(b) Forecast January 2019 Stuttgart

Fig. 5: Comparison of forecasts and actual sensor readings.

Given our analysis results, we can show that visualization is a powerful tool for assessing particle concentration and its impact on the environment. We see high potential for our approach to be applied to other cities and use cases. This could also help domain experts making decisions and drawing conclusions with a more profound background in environmental research than us.

## 4 Time Series Forecasting

Time series forecasting is a technique that allows computation of expected values for the future behavior of time-dependent measure values. In our previous work we designed a forecast technique called CSAR [Ha19] that focuses on the prediction of many time series that originate from the same domain, i.e. SPM sensor readings. Thorough predictions enable the comparison of predicted sensor readings without a driving ban and the actual measured values after the driving ban takes effect.

(a) No-go areas in Dresden in 2018.

(b) No-go areas in Stuttgart in 2018.

Fig. 6: No-go areas in both cities.

We predict the readings for all sensors in both cities using a dedicated CSAR model for Dresden and one for Stuttgart. The results of our experiment are shown in Figure 5. For the forecast, we aggregate the sensor readings to a daily granularity. Otherwise, the prediction of the entire January would end up with a forecast horizon too long to calculate reliably. For both cities, we see that the actual SPM levels (orange line) are slightly lower than the corresponding forecast values (blue line). For Stuttgart (Figure 5b) this might be attributed to the driving ban that took effect in January. The results for Dresden (Figure 5a), show that other external factor might influence the SPM levels too, such that there is a visible deviation.

However, the comparison of only two cities is too little data to draw a reliable conclusion from. More data and an in-depth analysis are required to assess the actual role of driving bans. Furthermore, other external influences that affect the SPM levels have to be identified and taken into account during the forecasting process. This would lead to more accurate forecast results and enable more reliable decisions on the effectiveness of driving bans.

## 5 No-go Areas

As mentioned earlier, we will identify no-go areas with the highest particle concentrations using the generated maps. We calculate the centroids for all areas with a dark hue. The centroid gives us the possibility to analyze nearby points of interest which could generate such a high concentration. Currently, we already identified six of the top concentration accumulations in Dresden in Figure 6a and the top six in Stuttgart in Figure 6b.

In Dresden, the accumulation in the north and south are the access roads from the surrounding areas to the city center. This includes two long tunnels on the Autobahn producing a high concentration at their entrances. The northwestern area is an Autobahn exit which has a lot of traffic backup from a large shopping center located right next to the Autobahn. The one cluster near the city center is the river harbor. Ships are well-known for not having clean engines [Sm15].

The Neckar harbor in the eastern part of Stuttgart has a similar influence. A high particle concentration can be spotted there. Near the harbor, there is an industrial area which is visible in the concentration as well. The city center has also a high particle concentration due to a lot of traffic going through. A particular interesting point is the ridge around the Birkenkopf hill southwest of the city center (marked with the mountain symbol). The curvature of the ridge forms a barrier in the particle concentration. The reason for this phenomenon could be the agglomeration of particles in front of the ridge if the wind flows perpendicular to the ridge. Last, one can see two darker areas for the access roads to the western Autobahn in the north and northwest of Stuttgart.

## 6 Conclusion

We have shown a multi-tool workbench for assessing the impact of SPM and driving bans in city areas. We use visualization and forecasting to find interesting patterns. This is done completely with a pure data-driven approach. Our argumentation closely follows the findings from the data. So, we rely on the correctness and completeness of the data. Data analysis only can be as good as the data fulfilling these two criteria. In our case the sparsity of the measurements due to the high humidity in some areas and the sparsity of the sensor network itself are negative factors for the analysis. The sparsity introduces uncertainty which can not be modeled correctly or can not be modeled at all. Complete data would require a organized network of sensors in our use case. Whereas the idea of open data and citizen science to collect relevant data is a good one, we see potential in getting a more organized structure into the project. Given the severity of the SPM topic, this could extend into a government-organized data collection scheme.

We think our work can be extended to different cities but also to a different set of problems. Any data which can be represented both as a concentration distribution and a time series will be assessable with our framework. Topics closely aligned would be weather and climate research, market research, and energy research. We see our approach as an important tool for knowledge transfer between people from different areas of research. The main goal would be to give domain experts a robust analysis platform for their decision making.

## References

[Cl90] Cleveland, Robert B; Cleveland, William S; McRae, Jean E; Terpenning, Irma: STL: A Seasonal-Trend Decomposition. Journal of Official Statistics, 6(1):3–73, 1990.

[Ha19] Hartmann, Claudio; Ressel, Franziska; Hahmann, Martin; Habich, Dirk; Lehner, Wolfgang: CSAR: the cross-sectional autoregression model for short and long-range forecasting. International Journal of Data Science and Analytics, jan 2019.

[IN17] INRIX Global Traffic Scoreboard, http://inrix.com/scorecard/, Accessed on 08-02-2019.

[Mo15] Moritz, S.; Sardá, A.; Bartz-Beielstein, T.; Zaefferer, M.; Stork, J.: Comparison of different Methods for Univariate Time Series Imputation in R. ArXiv e-prints, October 2015.

[No15] SDS011 Laser PM2.5 Sensor specification, http://ecksteinimg.de/Datasheet/SDS011 laser PM2.5 sensor specification-V1.3.pdf, Accessed on 08-02-2019.

[Pe84] Peter Alfeld: A trivariate Clough-Tocher scheme for tetrahedral data. Computer Aided Geometric Design, 1(2):169–181, 1984.

[Sm15] Smith, T. W. P. et al.: Third IMO Greenhouse Gas Study 2014. International Maritime Organization, United Kingdom, 2015.