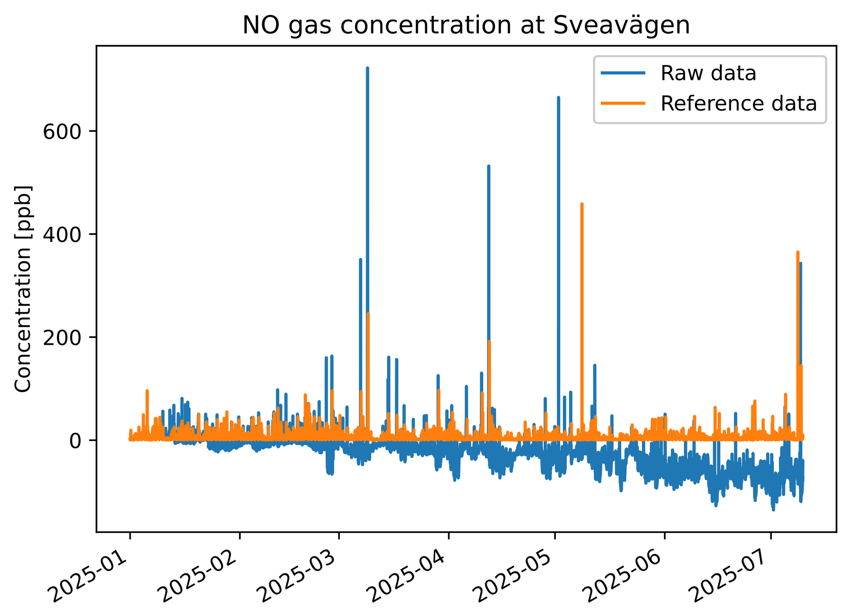
Kalman Filtering for Baseline Calibration

William Olsson and Ture Valtonen

Introduction

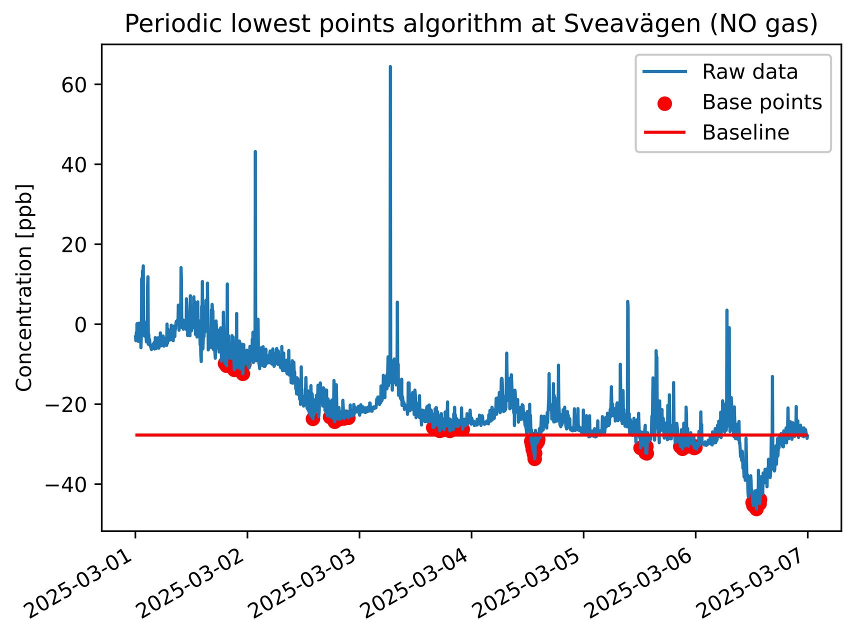
Sensorbee is a company in Linköping producing small gas sensors. Current gas sensors are large and expensive and these small sensors could serve as a compliment, providing data where we couldn't otherwise. Measurements aren't always reliable however; one problem is that the sensors tend to drift, as seen in the figure below. This is the problem we investigated this summer internship.



Baseline Estimation

One feature of gas concentrations is that they don't fall below a certain background level, often zero. In many cases they reach that background level regularly, often once a day. If we can find a baseline of lowest measurements from our sensor we can assume this baseline corresponds to the background level and calibrate accordingly.

An intuitive way to find a baseline is to look at the lowest point recorded during a day. However, this is sensitive to outliers and instead we look at the lowest 5% of points recorded.



Another factor to consider is that we expect the sensor to drift slowly over time, but the actual daily baselines vary a lot. Therefore we look at a rolling weekly average of daily baselines.

Kalman Filtering

Using the baseline estimation described in the previous section we get an estimate for the baseline given the previous week of data. Imagine we want to recalibrate again tomorrow, could we use today's data to improve our estimation?

The answer is yes, and the way we go about it is first making a guess about tomorrow's value. In our case we simply guess it will be the same as today, and quantifying how sure we are of our guess. When we get tomorrows measurement we weigh this against our guess to get a new estimate.

The actual algorithm used is a Kalman filter, with measurement and time updates defined as below. The value of Q is chosen by the user and R is the variance of the points used for baseline estimation.

$$\begin{cases} x_{k+1} = x_k + w_k, \; Cov(w_k) = Q \\ y_k = x_k + v_k, \; Cov(v_k) = R \end{cases}$$

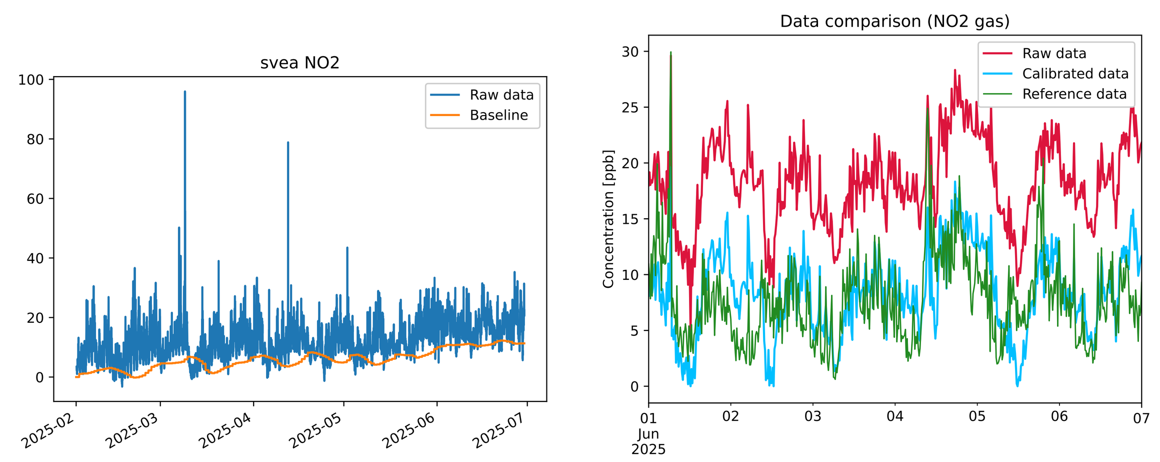
Results

In the figures below an example of an estimated baseline and an example of calibrated data is shown. The results varied between different gases and sensors. Overall we had most success calibrating NO_2 sensors. A comparison with error measurements are shown below.

Svea raw data			Calibrated data		
MSE	σ	Corr	MSE	σ	Corr
58.87	6.03	0.51	15.11	3.87	0.74

A foundational requirement for the algorithm is that the gas concentrations regularly reach a background level. When looking at O_3 data for example, we found that this wasn't the case and our algorithm performed poorly.

We also saw clear improvements when using our algorithm instead of basic automatic baseline calibration where we calibrate according to the latest baseline estimate.



Acknowledgements

Thanks to Oscar and David at sensorbee for all assistance and insights. Also, thanks to SLB-analys for providing us with reference data.

LINKÖPING UNIVERSITY

Sensor informatics and Decision-making for the Digital Transformation

UPPSALA UNIVERSITET