Influence of Humidity on the Accuracy of Low-Cost Particulate Matter Sensors

Technical Report · October 2017

DOI: 10.13140/RG.2.2.21095.75683

1 author:

Norbert Streibl

Influence of Humidity on the Accuracy of Low-Cost Particulate Matter Sensors

Norbert Streibl

October 2017

Contents

1 summary

Air borne particulate matter PM absorbs humidity from the air. Therefore the measured values of PM-sensors systematically increase, when the relative humidity rh is high[1]. By using a growth function $gf(rh)$ this condensation effect may be estimated by using the formula $PM_{wet} = gf(rh) \cdot PM_{dry}$. Low-cost sensors typically measure wet particulate matter PM_{wet}, whereas expensive sensors for professional purposes often measure dried particles PM_{dry}. For comparison of these values the growth function $gf(rh)$ should be known.

The relative humidity is mainly determined by the large-scale weather conditions; on top of this there exists a daily temperature effect: Usually it is colder in the night than in the day and on sunny days the PM-sensor may even be exposed to the sun. Therefore during the night the ambient air is nearer to the dew point than during the day. The weather dependent relative humidity is overlaid by a $24\,h$-cycle induced by temperature. This suggests a "big data approach" to identify the humidity dependent growth function: The growth function can be identified either by minimization of the overlaid $24\,h$-cycle of the PM-signal or by minimization of the correlation between the PM-signal and the rh-signal.

In this study growth functions were determined for PM-sensors from the OK-Lab sensor network[2] which consists of inexpensive NOVA SDS011 sensors[3]. According to specification these may be used for $rh < 70\%$. For $rh \approx 80\%$ the empirical growth factors scatter between 1.5 up to 3. For $rh \approx 90\%$ growth factors are in the range 2 to 5. Low-cost sensors for wet particulate matter overestimate the amount of particulate matter significantly compared to sensors for dry particulate matter. Therefore a correction of the condensation effect is necessary and it will be large and quite uncertain. Besides the measurement uncertainty of the PM-sensor and the measurement uncertainty of the rh-sensor a significant "correction uncertainty" must be considered. Nevertheless the correction brings the best low-cost sensors into the same range of magnitude as professional measurement stations[4].

The empirical growth functions seem to display significant seasonal variations: during spring time in April and May an increased growth was found. It may be speculated that this is caused by plant pollen. It has not yet been investigated, whether there is also a variation between city, suburbs and countryside.

We considered OK-Lab sensors firstly in the suburban countryside in and around the small city Leonberg $15\,km$ from Stuttgart and secondly at the periphery of Stuttgart located at Pragsattel near the edge of the Stuttgart basin. Thirdly the inner-city sensor operated by the government agency LUBW for dry particulate matter at Stuttgart Neckartor in the middle of the Stuttgart basin is used as a reference. The humidity corrected PM-values of OK-Lab sensors increase - as it may be expected - from the countryside to the suburbs and are lower than the inner city reference sensor. Without humidity correction the raw PM_{wet}-signals are on humid days higher than the reference sensor for PM_{dry}. At the beginning of 2017 strong episodes of pollution have been observed at the reference sensor near Stuttgart Neckartor. These are well visible with the low-cost sensors in the surrounding area, even on the countryside, due to large-scale meteorological conditions such as inversions. During late summer and in the fall the nightly humidity is near dew point and the air becomes dry on sunny days. Therefore a $24\,h$-cycle in humidity as well as in wet particulate matter PM_{wet} is distinctly observable and can be compensated by applying an empirical growth function. Adjacent sensors show comparable results after humidity compensation. However, the empirical growth function must be determined individually for each sensor, as there is no universally valid compensation function - the local environmental conditions as well as systematic deviations of the humidity sensors are individual. Further improvements may be achievable, if a time-variant (i.e. sliding) parametrization of the empirical growth function would be employed.

In summary: The raw measured values PM_{wet} from low-cost sensors require correction for humidity rh in order to become comparable to reference sensors for PM_{dry}. Due to the uncertainties of the relevant sensor signals and the strong dependency on humidity, however, the corrected results remain rather an indication than a highly accurate measurement. On the other hand overall trends are well reproduced.

[1] Bernd Laquai pointed this out to the author in the context of low-cost PM sensors
[2] https://www.madavi.de/ok-lab-stuttgart/
[3] hyperlink: datasheet NOVA SDS011
[4] https:(https://www.stadtklima-stuttgart.de/index.php?luft_messdaten_feinstaubwerte)

Figure 1: foggy and dusty sunrise on February 11th, 2017 near the sensor Leonberg Silberberg #2

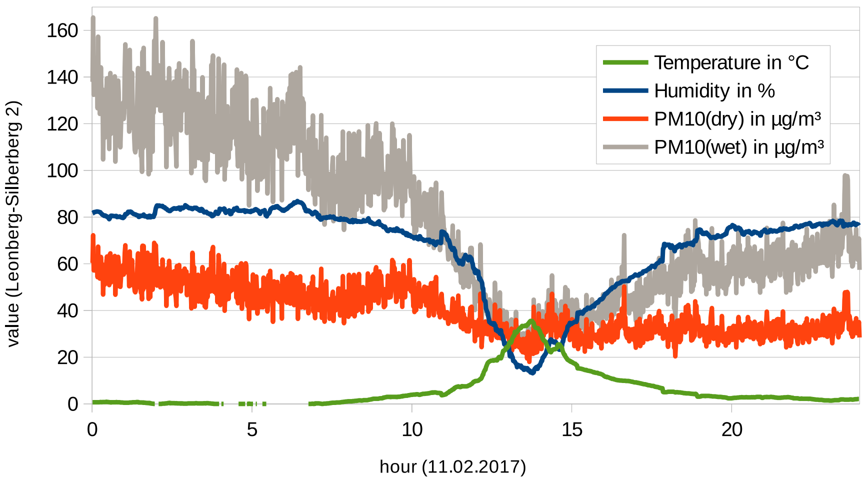

Figure 2: signals of sensor Leonberg Silberberg #2 on February 11th, 2017. Air temperatures (green) rise from nightly $0°C$ up to $35°C$ around $14:00$ hours while exposed to the sun. Humidity (blue) falls from 82% to 14% while $PM10_{wet}$ (gray) also decreases from $135\frac{\mu g}{m^3}$ to less than $40\frac{\mu g}{m^3}$. After correction with an empirical growth function $gf(rh) = (1-rh)^{-0.49}$ a slow continuous decrease of the estimated PM_{dry} is observed over the day.

Location	Operator	$PM10_{wet}$	$PM10_{dry}$
Leonberg Silberberg #1	OK-Lab	$101\frac{\mu g}{m^3}$	$51\frac{\mu g}{m^3}$
Leonberg Silberberg #2	OK-Lab	$106\frac{\mu g}{m^3}$	$53\frac{\mu g}{m^3}$
Leonberg Gartenstadt #2	OK-Lab	$69\frac{\mu g}{m^3}$	$34\frac{\mu g}{m^3}$
Stuttgart Pragsattel	OK-Lab	$159\frac{\mu g}{m^3}$	$67\frac{\mu g}{m^3}$
reference: Stuttgart Neckartor	LUBW	-	$89\frac{\mu g}{m^3}$

Table 1: comparison of the median PM-values for several sensors on February 11th, 2017

2 correlation of mass of particulate matter and humidity

During fall season, when not much particulate matter pollutes the countryside, and when the nightly temperature is near dew point, and when many days are warm and sunny in Germany, then the temperature- and humidity-signals have a visible $24\,h$ periodical component. The humidity signal oscillates between 90% in the night and less than 30% when exposed to the sun.

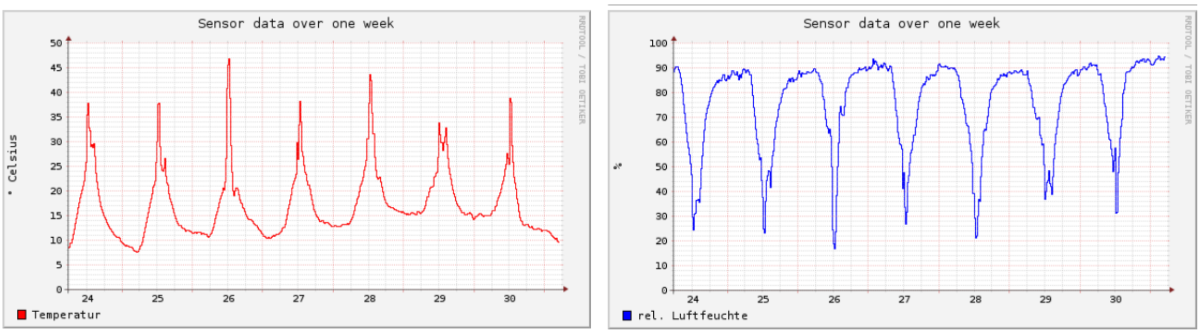

Figure 3: temperature and humidity between 24.09. and 30.09.2017 at sensor Leonberg Silberberg #2

The corresponding PM-signals also oscillate, although they look significantly more noisy. The values for dry air are in the single digit range $PM_{dry} < 10\,\frac{\mu g}{m^3}$.

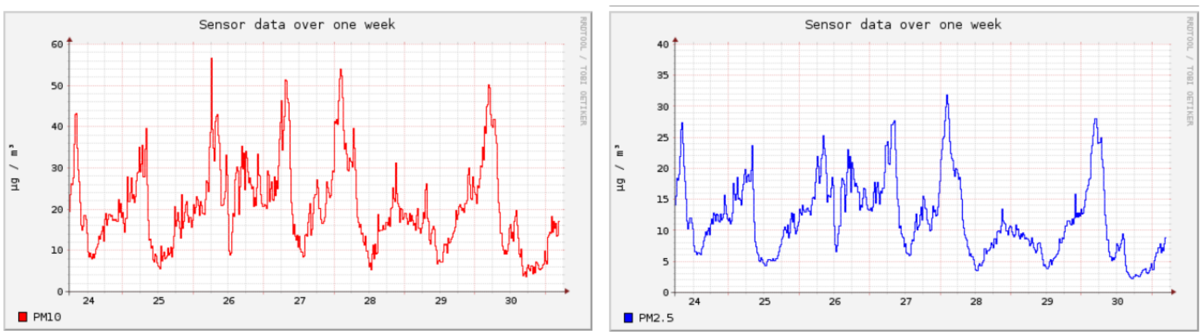

Figure 4: particulate matter between 24.09. and 30.09.2017 at sensor Leonberg Silberberg #2

The reference sensor downtown Stuttgart operated by the LUBW and located near the Neckartor [5] is exposed to a lot of street traffic and measures dry particulate matter of the class $PM10$. During the same week its daily average values range between 29 and $40\frac{\mu g}{m^3}$ and were mostly near $34\frac{\mu g}{m^3}$. Therefore single digit values out at the countryside seem quite plausible.

[5] see: https://www.stadtklima-stuttgart.de/index.php?luft_messdaten_feinstaubwerte

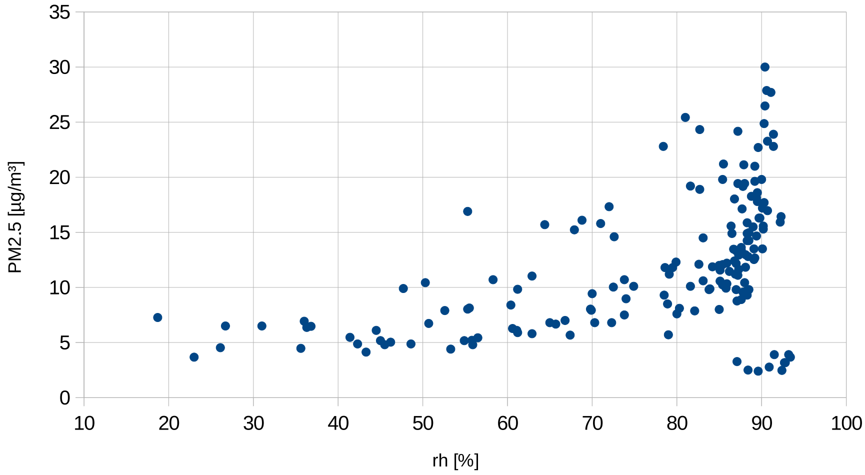

Figure 5: $PM2.5_{wet}$ versus rh between 24.09. and 30.09.2017 at sensor Leonberg Silberberg #2

The scatter plot of particulate matter against humidity shows a certain correlation, because the data points are not scattered chaotically in a cloud. They seem to accumulate along a curve, which is rather flat for humidities $rh < 50\%$. For higher humidity the curve increases strongly. It looks as if for $rh > 90\%$ fivefold higher PM-values may occur, than under dry conditions. This indicates a growth function with a steep slope towards higher humidity. In the lower right corner lies a cluster of measured values, where the PM-value is very low, while humidity is very high: these measurements were obtained after a strong rain had cleaned the air from particulate matter.

3 possible choices for empirical growth functions

It is assumed that a growth function exists, which relates the measurements of wet and dry particulate matter:

$$PM_{wet} = gf(rh) \cdot PM_{dry} \tag{1}$$

The relative humidity is in the calculation formulas normalized $0 \leq rh \leq 1$, i.e. a value $rh = 1$ corresponds to 100% humidity. If wet particulate matter was measured and the growth function was somehow estimated, then a correction for humidity becomes possible:

$$PM_{dry} = \frac{PM_{wet}}{gf(rh)} \tag{2}$$

From the physics of condensation it is not obvious, that a growth function actually exists: the increase of the measured values depends on the size distribution of the particulate matter and how much of the finer dust becomes visible to the sensor after swelling due to the condensation of humidity. If the size distribution is irregular and there are peaks at certain sizes, then the growth will also become irregular. Nevertheless literature gives several parametrized formulas for empirical growth functions:

$$gf_{Hänel} = \frac{1}{(1 - rh)^{\beta}} \tag{3}$$

$$gf_{Soneja} = 1 + \frac{\alpha \cdot rh^2}{1 - rh} \tag{4}$$

$$gf_{combo} = 1 + \alpha \cdot \frac{rh^2}{(1-rh)^\beta} \tag{5}$$

$$gf_{Skupin} = \begin{cases} \frac{\alpha}{(1-rh)^\beta} & rh \geq 0.7 \\ \frac{1}{(1-rh)^{\beta - \frac{log(\alpha)}{log(0.3)}}} & rh < 0.7 \end{cases} \tag{6}$$

By properly choosing the parameters α and β these functions can be fitted to measured data. Common to all growth functions is that$gf(0) = 1$ and $gf(rh \to 1) \gg 1$. The growth functions according to Hänel and to Skupin were found in the thesis of A. Skupin[6], the function according Soneja from an scientific article[7] and the combination function is an own attempt to combine Soneja with Hänel.

4 empirical parametrization of growth functions

The measured values of the OK-Lab sensors are stored on a public server[8]. These signals were processed by taking the following steps:

- ▷ download time series from server, which consists of time stamp, $PM10$, $PM2.5$, temperature, humidity and some additional data
- ▷ eliminate non-plausible data, where values are outside a reasonable range and/or items are missing
- ▷ resample the measured data onto a regular time raster with 1 sample per minute (repeat last valid data set until a new and valid data set becomes available)
- ▷ process signals, e.g. apply humidity corrections or perform Fourier transformations to analyze spectral behavior or calculate correlations
- ▷ smooth data by median filtering and, if desired, resample to a more coarse raster with 1 sample per hour or per day

The identification of an empirical growth function is based on one of the two following hypotheses:

- ▷ temperature, humidity and the growth of particulate matter contain a periodic component with a $24\,h$ period. The absolute value of the corresponding normalized Fourier-coefficient becomes minimal, as soon as the influence of humidity is compensated in the best possible way. Therefore the parameters of the empirical growth functions are identified by minimization of the function

$$fom(\alpha,\beta) = 10 \cdot log_{10} \left(\frac{\left| \int_{-\infty}^{\infty} \frac{PM(t)}{gf(\alpha,\beta)} \cdot e^{2\pi i \frac{t}{24h}} \cdot dt \right|}{\int_{-\infty}^{\infty} \frac{PM(t)}{gf(\alpha,\beta)} \cdot dt} \right) \tag{7}$$

- ▷ the signals for dry particulate matter and humidity are not expected to be correlated. The normalized correlation factor between PM_{dry} and rh becomes minimal, as soon as the influence of humidity is compensated in the best possible way. Therefore the parameters of the empirical growth functions are identified by minimization of the function

$$fom(\alpha,\beta) = \frac{\sum_{i=1}^{N} (rh_i - \langle rh \rangle) \cdot \left(\frac{PM_i}{gf(\alpha,\beta)} - \left\langle \frac{PM_i}{gf(\alpha,\beta)} \right\rangle \right)}{\sqrt{\sum_{i=1}^{N} (rh_i - \langle rh \rangle)^2 \cdot \sum_{i=1}^{N} \left(\frac{PM_i}{gf(\alpha,\beta)} - \left\langle \frac{PM_i}{gf(\alpha,\beta)} \right\rangle \right)^2}} \tag{8}$$

whereby the brackets denote an arithmetic mean value $\langle x \rangle = \frac{1}{N} \sum_{i=1}^{N} x_i$

[6]Anne Skupin: "Optische und mikrophysikalische Charakterisierung von urbanem Aerosol bei (hoher) Umgebungsfeuchte", University Leipzig (2013)

[7]Sutyajeet Soneja et al.: Humidity and Gravimetric Equivalency Adjustments for Nephelometer-Based Particulate Matter Measurements of Emissions from Solid Biomass Fuel Use in Cookstoves; Int. J. Environ. Res. Public Health 2014, 11, 6400-6416

[8]https://www.madavi.de/ok-lab-stuttgart/

The growth functions according to Hänel and Soneja depend on one parameter each. Therefore the parameter can easily identified by plotting the figure of merit fom as a function of the parameter. More complicated functions are minimized numerically by using the simplex method, which is reasonably fast to compute.

5 growth function according to Hänel

Hänel gave the growth function $gf = (1 - rh)^{-\beta}$ wherein the “Hänel exponent” β serves as fitting parameter. For the sensor Leonberg Silberberg #2 different results are obtained depending on the time range of the measurements and depending on the figure of merit, i.e. whether the normalized Fourier-coefficient for the $24\,h$ period or the normalized correlation coefficient between particulate matter and humidity are minimized. In the following analysis the value for $PM2.5$ was used, because this signal looks less noisy than $PM10$.

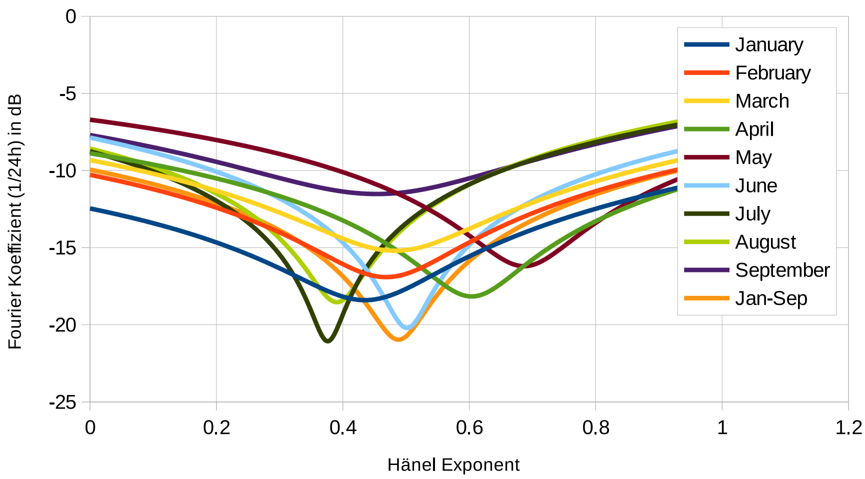

Figure 6: Fourier coefficient for the $24\,h$ period as a function of the Hänel exponent β

The normalized Fourier-coefficient typically changes by $10\,dB$, that is by an order of magnitude. Only in January the variation was less decisive, when the weather was humid and not very sunny and the PM-signals where dominated by strong pollution events. During every month, however, a clearly defined minimum occurs albeit at different Hänel exponents: they vary between 0.38 and 0.69. For the time range between 01.01.2017 - 30.09.2017 the optimum Hänel exponent takes the value $\beta = 0.49$.

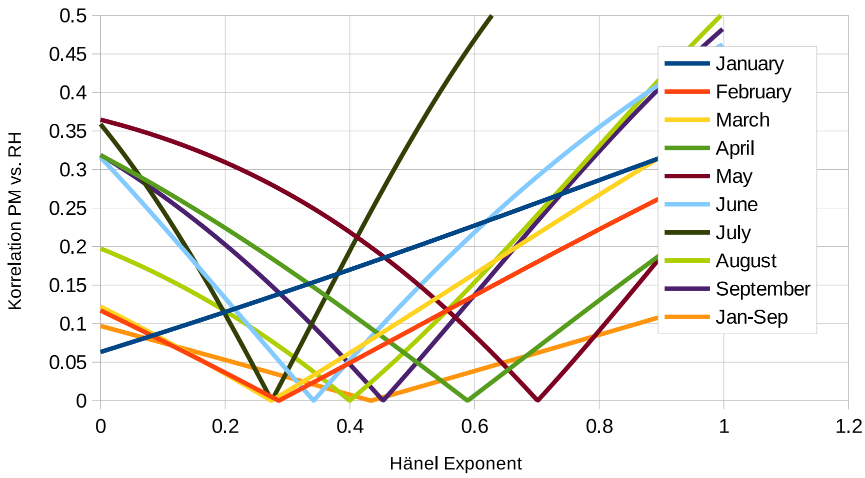

Figure 7: correlation coefficient as a function of the Hänel exponent β

The normalized correlation coefficient typically changes by more than an order of magnitude. However, in January no minimum occurs at all, probably because the correlation coefficient 0.06 between PM_{wet} and rh is already too small for a minimization. January was dominated by strong pollution events, such as the pollution by fireworks at new years day and a huge pollution episode caused by a meteorological inversion, that lead to an environmental alarm in Stuttgart ("Feinstaubalarm"). During the other months a clearly defined minimum occurs albeit at different Hänel exponents: these vary between 0.27 and 0.70. For the time range between 01.01.2017 - 30.09.2017 the optimal Hänel exponent takes the value $\beta = 0.43$.

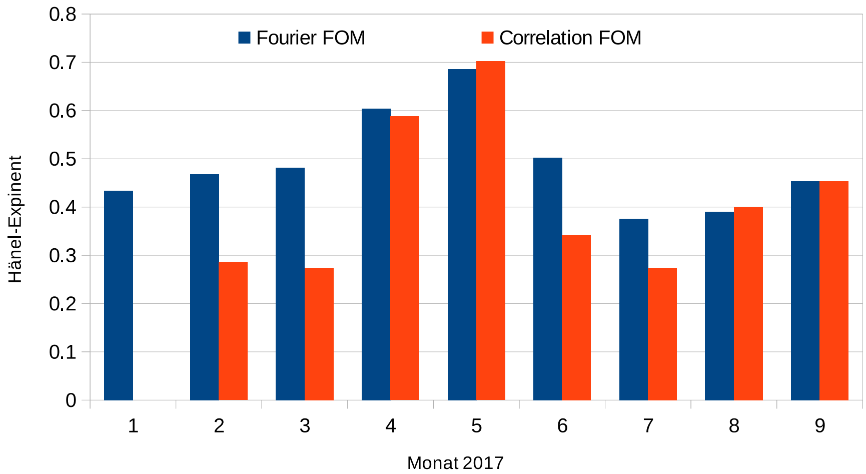

Figure 8: monthly variation of the Hänel exponent at sensor Leonberg Silberberg #2

The Fourier and the correlation method yield Hänel exponents in a similar range. Also the seasonal variation, which is clearly observable, looks similar. One might speculate, whether the spring maximum of the Hänel exponent around April and May is caused by plant pollen with a hygroscopic property different from ordinary dust.

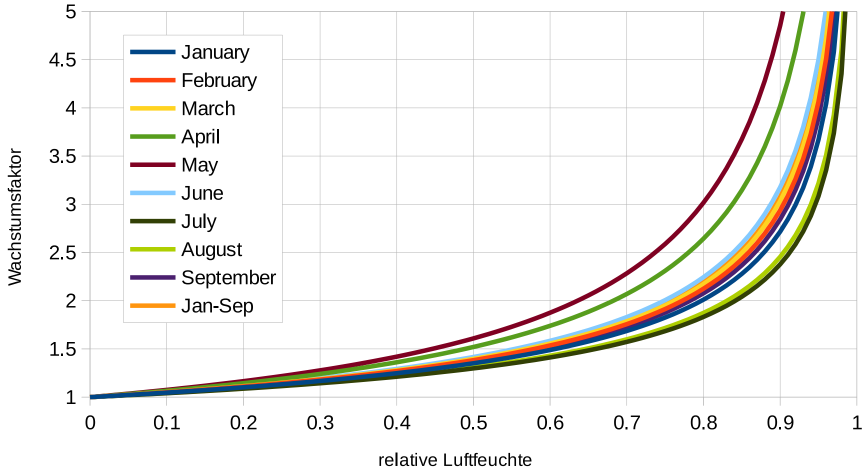

Figure 9: empirical growth functions according to Hänel for the different months

The empirical growth functions vary from month to month. Below a humidity $rh = 70\%$ the growth is below a factor of 2. Incidentally the sensor NOVA SDS011 is specified only for humidities $rh < 70\%$. For humidities $rh > 90\%$ growth factors of 5 may occur. Under humid conditions the sensor has a huge systematic deviation.

6 growth function according to Soneja et. al.

A similar analysis was conducted for Soneja's growth function $gf_{Soneja} = 1 + \frac{\alpha \cdot rh^2}{1-rh}$ with the "Soneja weight" α as fitting parameter:

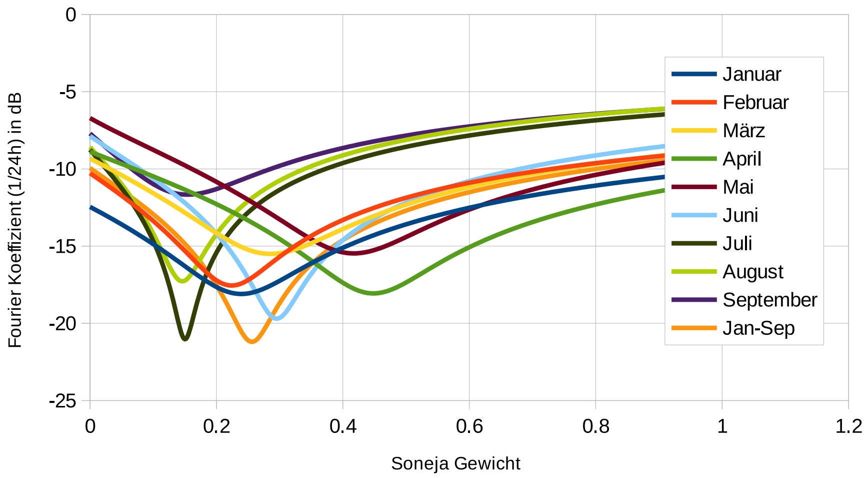

Figure 10: Fourier coefficient for $24\,h$ period as a function of the Soneja weight α

The normalized Fourier-coefficient typically changes again by $10\,dB$, that is by an order of magnitude. Only in January the variation was less, when the weather was humid and not very sunny and the PM-signals where dominated by strong pollution events. During every month, however, a clearly defined minimum occurs albeit at different Soneja weights: they vary between 0.15 and 0.45. For the time range between 01.01.2017 - 30.09.2017 the optimal Soneja weight takes the value $\alpha = 0.256$.

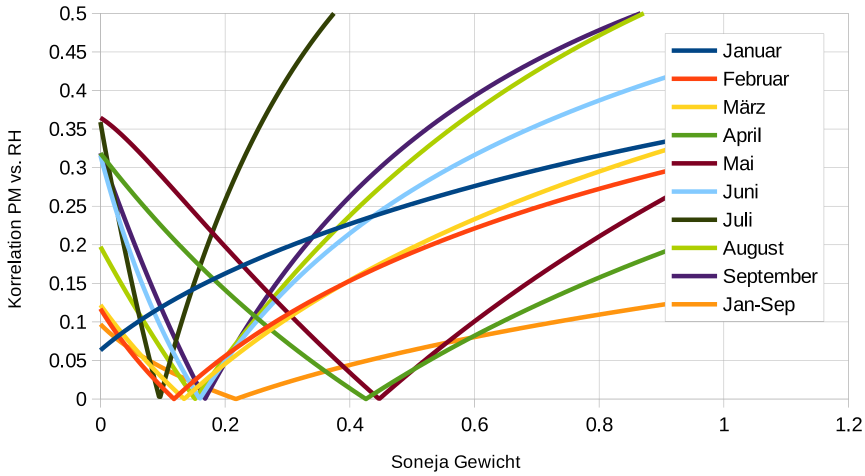

Figure 11: correlation coefficient as a function of the Soneja weight α

The normalized correlation coefficient typically changes by more than an order of magnitude. However, in January no minimum occurs just as in the case of the Hänel growth function. During the other months a clearly defined minimum occurs albeit at different Soneja weights: these vary between 0.1 and 0.45. For the time range between 01.01.2017 - 30.09.2017 the optimum Soneja weight takes the value $\beta = 0.218$.

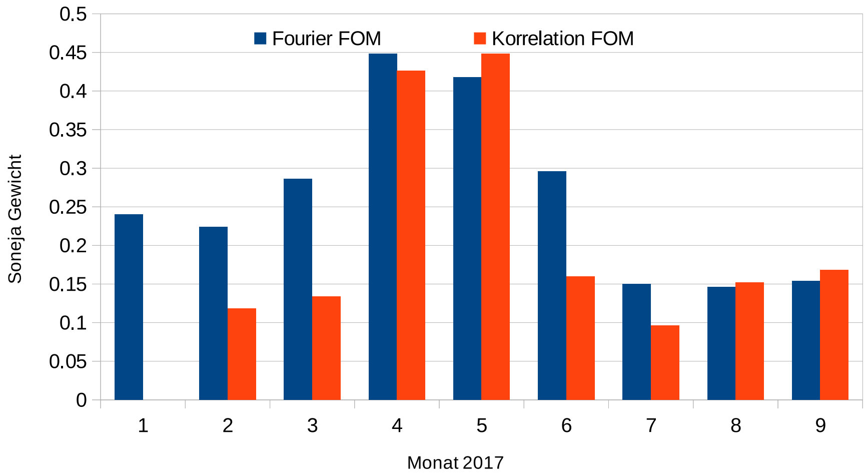

Figure 12: monthly variation of the Hänel exponent at sensor Leonberg Silberberg #2

Comparison of the monthly values for the Soneja weights shows a behavior similar to the Hänel exponent with a spring time peak around April and May.

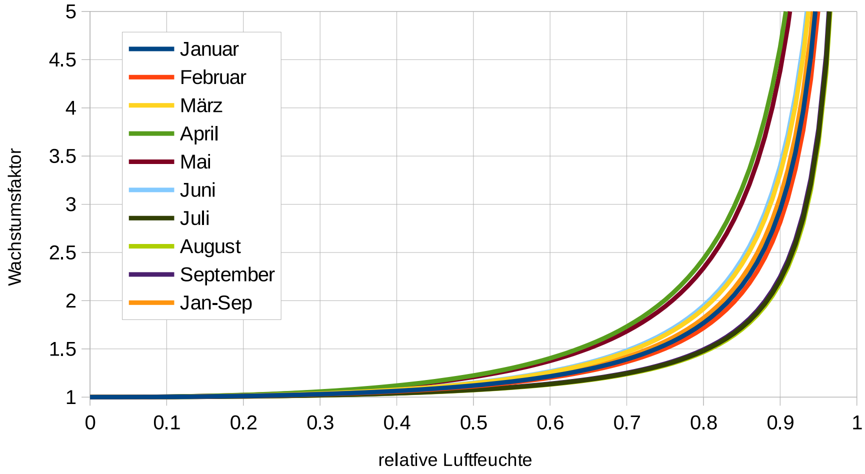

Figure 13: empirical growth functions according to Soneja et.al. for the different months

The empirical growth functions according to Soneja are quite similar to those after Hänel indicating a systematic deviation of a factor 5 under very humid conditions.

7 growth function according to Skupin and comparison

The two fitting parameters of the growth function according to Skupin were numerically identified by using the simplex method. However, the related growth curves look slightly less plausible than those according to Hänel or Soneja. A probable reason might be that the two fitting parameters are derived from noisy data with not too many data at low humidities $rh < 70\%$. A similar effect occurred, when it was tried to identify the two fitting parameters of the combo growth function. It seems that parameter identification for growth functions with more than one parameters is less robust.

Skupin investigated in her thesis large data sets derived from state of the art sensors in Leipzig and gives mean values and limits for the fitting parameters of her growth function. It is instructive to plot the resulting growth functions and compare to our empirical growth function according to Hänel.

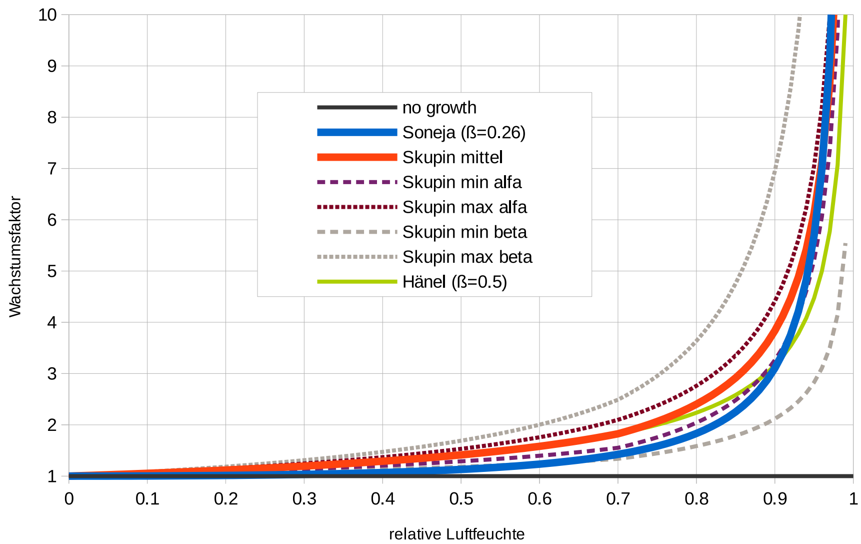

Figure 14: range of growth functions according to Skupin and Hänel

If the empirical growth functions according to Hänel and Soneja for the long time range from 01.01.2017 until 30.09.2017 and for the Fourier- and the correlation method are plotted, again a significant scattering of results becomes visible. Again a large deviation occurs for high humidities in all considered cases. If we use any empirical growth function for the correction of wet PM-signals, the correction uncertainty will be rather high due to this scatter.

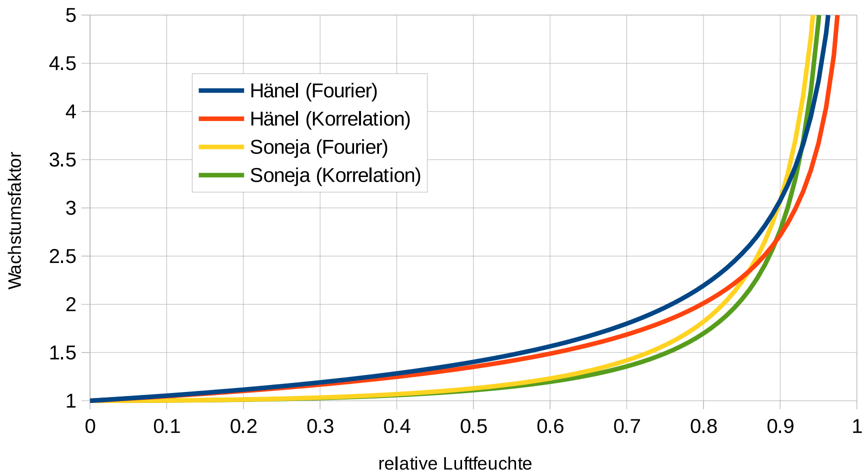

Figure 15: comparison of empirical growth functions for sensor Leonberg Silberberg #2

Around 80% humidity the growth factor may range between 1.5 and 3. Around 90% humidity the growth factor may range between 2 and 5.

8 results of humidity corrections

In the following study the growth function according to Hänel was chosen and the fitting parameter β was determined by minimization of the normalized Fourier-coefficient using the numerical simplex optimization method over the long time range from 01.01.2017 until 30.09.2017. The optimal Hänel exponents for different sensors were determined individually for each sensor. In this way humidity corrected time series for dry particulate matter $PM10_{dry}$ were estimated for exemplary sensors.

The raw data $PM10_{wet}$ (gray) and the humidity corrected data $PM10_{dry}$ (blue) show at the sensor Leonberg Silberberg #2 in January a comparatively high pollution and in August a comparatively low pollution in the hourly median value.

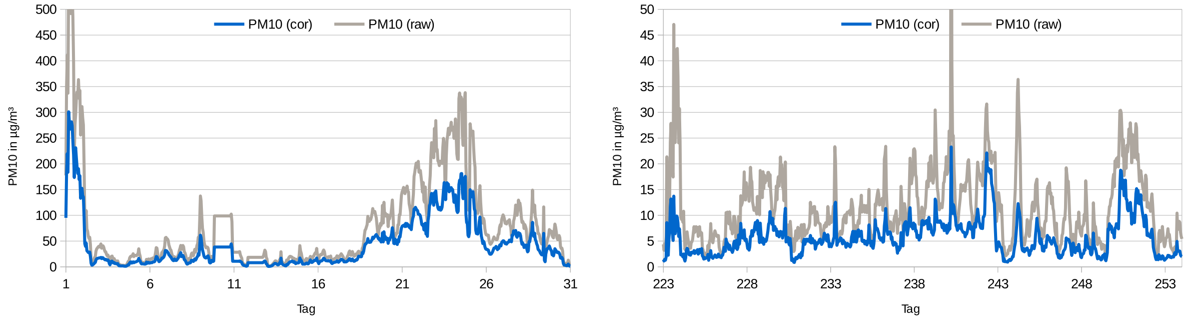

Figure 16: raw and humidity corrected $PM10$-signals at sensor Leonberg Silberberg #2 for January and August 2017

In January we see the aftermath of the new years day fireworks and a strong meteorological inversion beginning around 20.01.2017, both inducing heavy air pollution. The flat signal around 10.01.2017 was caused by a temporary sensor blackout. In August the pollution is generally low and overlaid by the above mentioned $24\,h$-cyclic growth of the particulate matter that leads to an oscillating signal. This periodic component is significantly reduced by the humidity correction. Actually a somewhat better compensation would be possible if the higher August value of the Hänel exponent would be used instead of the long term value.

Next the median of the PM-signal over one day was calculated for different sensors.

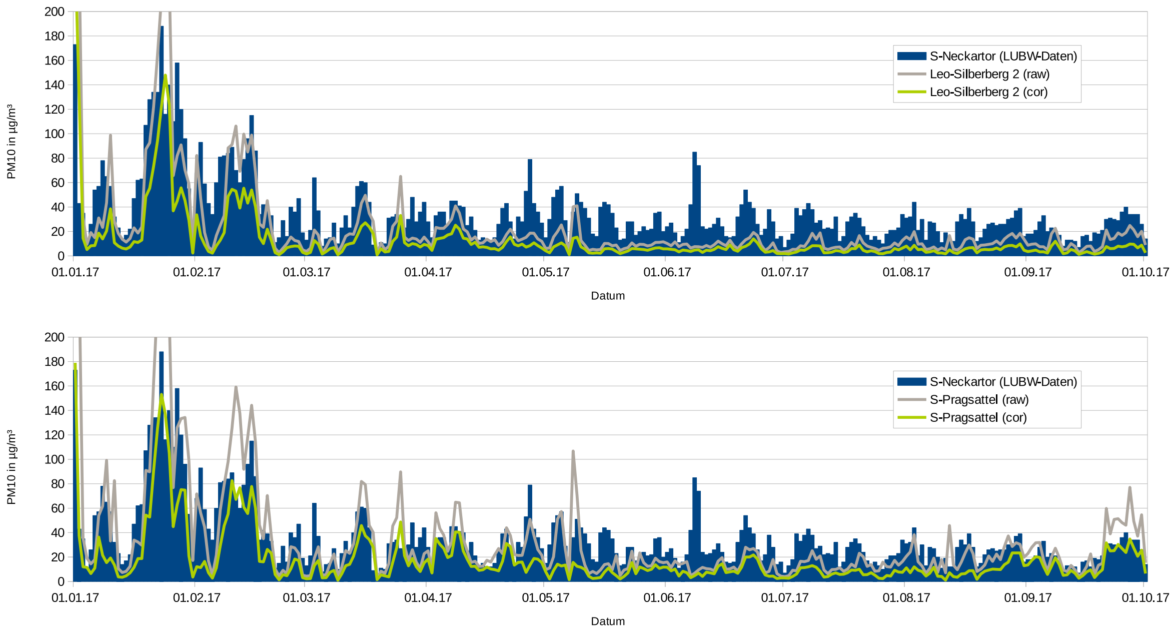

Figure 17: long time run (01.01.2017-30.09.2017) of the raw and humidity corrected PM-signals at the sensors Leonberg Silberberg #2 and Stuttgart Pragsattel and the reference sensor Stuttgart Neckartor operated by the LUBW

The blue bars show the reference sensor for PM_{dry} at Stuttgart Neckartor. It is expected to yield the highest signal, because of its location down in the Stuttgart basin and because of a high local traffic. The gray curves show the raw data PM_{wet} from the low-cost sensor Leonberg Silberberg #2, which is situated in the countryside outside the Stuttgart basin, and Stuttgart Pragsattel, which is situated at the edge of the Stuttgart basin. Both gray curves are often higher than the blue bars, although the pollution is expected to be lower at these locations, because of the growth due to humidity. The green curves show the humidity corrected values, which are always lower than the reference sensor at its highly polluted location. All curves reflect the strong pollution episodes at the beginning of the year 2017.

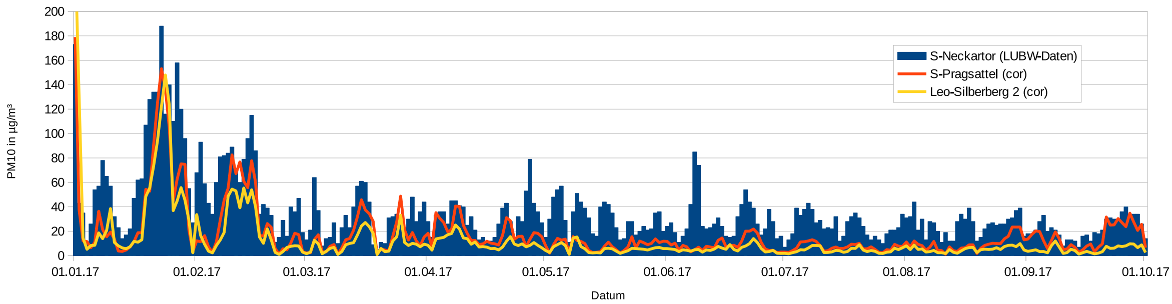

Figure 18: dry particulate matter in the countryside, in the periphery and downtown

Overexposing the curves for dry particulate matter PM_{dry} for the countryside sensor (Leonberg Silberberg #2), the sensor at the edge of the Stuttgart basin (Stuttgart Pragsattel) and the reference sensor down in the Stuttgart basin (Stuttgart Neckartor, operated by LUBW) shows several facts: In the bottom of the basin the pollution is highest. Outside the city the air becomes cleaner, because the air is exchanged more effectively and because there is less traffic. Nevertheless under specific meteorologic conditions, namely under inversion, the air pollution by particulate matter extends over a huge geographic area far out of the city.

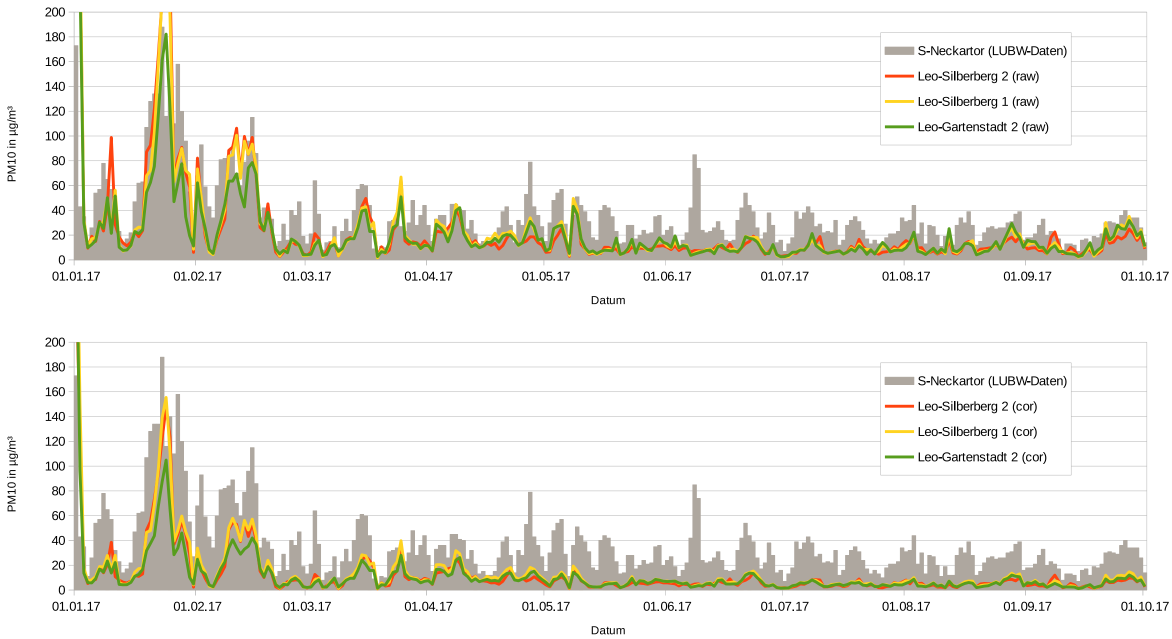

Figure 19: raw and humidity corrected PM-signals at adjacent sensors in Leonberg

Three sensors in Leonberg are quite near to each other: the distance between Leonberg Silberberg #1 and #2 is less than $100\,m$ and the sensor Gartenstadt #2 is about $3\,km$ away. After humidity correction the sensor signals are reasonably similar and significantly below the reference sensor Stuttgart Neckartor. However, not all OK-Lab sensors fit this picture. Therefore another calibration or a quality selection of the sensors might be considered. The sensor specific Hänel exponents are quite different and therefore cannot be transfered from one sensor to another:

location	Hänel exponent	$PM10_{wet}^{median}$	$PM10_{dry}^{median}$	$PM10_{wet}^{sigma}$	$PM10_{dry}^{sigma}$
Leonberg Silberberg #1	0.30	11.9	7.3	49.6	28.7
Leonberg Silberberg #2	0.49	11.9	6.4	43.9	23.9
Leonberg Gartenstadt #2	0.37	11.8	6.5	46.6	22.4
Stuttgart Pragsattel	0.28	19	10.8	58.1	22.9
Stuttgart Neckartor (LUBW)	1	n/a	28	n/a	28.2

Table 2: individual Hänel exponents and statistics (PM-values in $\frac{\mu g}{m^3}$) for some OK-Lab sensors from 01.01.2017 until 30.09.2017

The overall statistics of the humidity correction looks quite plausible: the long time median for $PM10$ decreases significantly after correction from wet to dry values. The medians increase from countryside over city periphery to downtown. Even more significantly, the standard deviations are about cut in half by the humidity correction and become comparable between the OK-Lab sensors and the reference sensor.

Another improvement in the correction of the growth caused by humidity may be possible, if not a long time value of the Hänel exponent is employed, but instead a sliding short time value.

9 possible directions for future research

▷ Are the seasonal variations in fig. 8 and 12 real? they seem to indicate that in spring the hygroscopy of particulate matter differs from other seasons

▷ Is there a dependence on location or composition of particulate matter or weather conditions? Probably the Hänel exponent varies also between countryside, inner city, locations nearer to sea (salt) etc. as well as on weather conditions

▷ Why do sensors differ so much in their Hänel exponents? One might suspect at first glance, that the Hänel exponents should not be so much a function of the individual sensor, but rather of the physics of condensation and the hygroscopy of the particulate matter. Or do we see here the normal scatter and all we can say about low-cost sensors is, that we must expect a Hänel exponent somewhere between 0.2 and 0.6?

▷ Is there a time lag between the time series of humidity and the increase of measurement values of the particulate matter due to swelling of the particles? One might expect some time constants due to the physics of condensation.

▷ Some sensors seem not to allow for plausible humidity corrections. Why? Are they just bad?

▷ How do inner city sensors, e.g. those near Stuttgart Neckartor, actually compare to the reference sensor?

▷ Would a sliding (time-variant) parametrization of the growth function yield better results?

It is quite simple to come up with many more interesting research questions ...