Validating PM2.5 sensors using existing data

Plume Plotter
ash@plumeplotter.com

Introduction

The Luftdaten device (https://luftdaten.info/) is a simple device that comprises a cheap particulate matter (PM) sensor (usually the SDS011, but other sensors are supported) and a wifi component to upload the sensor readings directly to the Luftdaten server, where they are publicly available. These devices have been deployed widely, with many thousands of them in use (shown on https://maps.luftdaten.info).

However, although there have been several studies evaluating the SDS011 sensor and pointing out its weaknesses (especially its sensitivity to humidity), there seems to be no simple evaluation of the sensor that shows how well it agrees with reference instruments in monitoring ambient air. This has been done for commercial devices such as the Purple Air [1], Purple Air PA-II [2] and Aeroqual [3] sensors by colocating the device being tested with the reference instrument and measuring the correlation between their readings during some period of time. Results reported for these (for 1-hour means of PM2.5) are shown in Table 1. GRIMM and BAM are two types of reference instrument. R^2 is the coefficient of determination, which in this case is equal to the square of the Pearson correlation coefficient.

Table 1. Some previously evaluated sensors.

	Test period	**GRIMM R^2**	**BAM R^2**
Purple Air	19/02/2016 – 19/04/2016	0.91 – 0.93	0.78 – 0.79
Purple Air PA-II	08/12/2016 – 26/01/2017	0.94 – 0.98	0.87 – 0.92
Aeroqual	21/12/2017 – 15/03/2018	0.80 – 0.87	0.83 – 0.84

Is it possible to do a similar evaluation of the SDS011 by comparing historical PM2.5 readings from Luftdaten devices already installed with the corresponding readings from government PM monitors nearby? If so, there would be several advantages:

1. There is no cost for performing the experiments.
2. Thousands of devices can be evaluated, rather than a handful of selected devices.
3. Devices can be tested over different, longer, periods.
4. The devices are already deployed in a variety of real-world locations, rather than in a lab. (The term “field evaluation” in [1-3] seems to mean evaluation outdoors in ambient air but still at a controlled lab location.)

On the other hand, there are several difficulties:

1. The devices are not colocated: there is a distance between government monitors and the nearest Luftdaten sensors. Siting sensors apart, even a small distance, could dramatically change the readings because the pollution sources affecting each one could be different.
2. The Luftdaten sensors are installed in various unknown locations and conditions.
3. Data from Luftdaten sensors may be incomplete. As well as hours with no readings, there may be hours with few readings, which would reduce the accuracy of the hourly means.

All of these difficulties would seem to reduce agreement between the sensor readings, compared with the ideal case (colocated sensors). Therefore, we assume that:

> **The agreement (R^2) between any two sensors will be greater if they are colocated than if they are apart.** **(*)**

This should be true provided R^2 is high enough. (If R^2 is low, chance will be more likely to affect the readings.)

Methods

We investigate the correlation (R^2) between readings from government PM2.5 monitors and readings from various Luftdaten sensors nearby, and take the *maximum* of these as a lower bound on the R^2 that could be obtained if the sensors were colocated. This is justified if our assumption (*) is true.

We consider only 2018 data and include only those Luftdaten sensors in the UK that started in the first half of 2018. These sensors are all in or near three cities: Eastbourne, Bristol and Aberdeen. Each of these towns also has a government (DEFRA) PM2.5 monitor based on "TEOM FDMS" technology (which is different from both BAM and GRIMM). Table 2 lists the Luftdaten sensors included and the three DEFRA monitors. The fourth column shows (as a percentage) how many hours of 2018 have readings available. For the Luftdaten sensors, which normally upload readings every few minutes, the 1-hour mean for each hour was calculated as the average of all readings in the preceding hour.

Table 2. Luftdaten sensors and DEFRA monitors that have 2018 data available.

Town	ID	Earliest date	% hours
Bristol	DEFRA-B	01/01/2018	80
	3040	01/01/2018	96
	7675	01/01/2018	96
	7677	01/01/2018	92
	7685	01/01/2018	98
	8741	18/01/2018	90
	10179	21/02/2018	70
	10491	03/03/2018	78
	12711	27/04/2018	55
	14679	21/06/2018	49
	14787	25/06/2018	48
Eastbourne	DEFRA-E	01/01/2018	98
	4696	01/01/2018	97
	5523	01/01/2018	98
	5531	01/01/2018	80
	9044	25/01/2018	72
	11315	20/03/2018	72
	12585	21/04/2018	68
	13293	11/05/2018	53
	14369	15/06/2018	53
Aberdeen	DEFRA-A	01/01/2018	98
	5331	01/01/2018	98
	8733	18/01/2018	88

Comparing Luftdaten sensors

First we compare Luftdaten sensors with each other. We check every pair of Luftdaten sensors that are in the same town and which both have readings for at least 50% of the hours in the time period being considered. For each pair we calculate the R^2. We considered three time periods within 2018:

- 19/02 – 19/04 (used by the Purple Air evaluation [1]),
- 01/01 – 30/04 and 01/11 – 31/12 ("winter": coldest half of the year),
- 01/01 – 31/12 (whole year).

Table 3 shows the number of pairs of sensors satisfying the criteria for each time period. It also shows the maximum R^2, the percentage of hours present, the two sensors in the pair, and the distance between them.

Table 3. Most correlated pair of Luftdaten sensors for each time period.

Time period	# pairs	Maximum R^2	% hours	Sensor 1	Sensor 2	Distance
19/02 – 19/04	28	0.98	91	5523	9044	303m
Winter	28	0.96	71	5523	9044	303m
Whole year	50	0.94	72	5523	9044	303m

Figure 1 shows the concentration reported by the most highly correlated pair of sensors (5523 and 9044) for each hour during the 19/02 – 19/04 time period. Figure 2 shows the linear correlation between these two sensors for each of the three time periods.

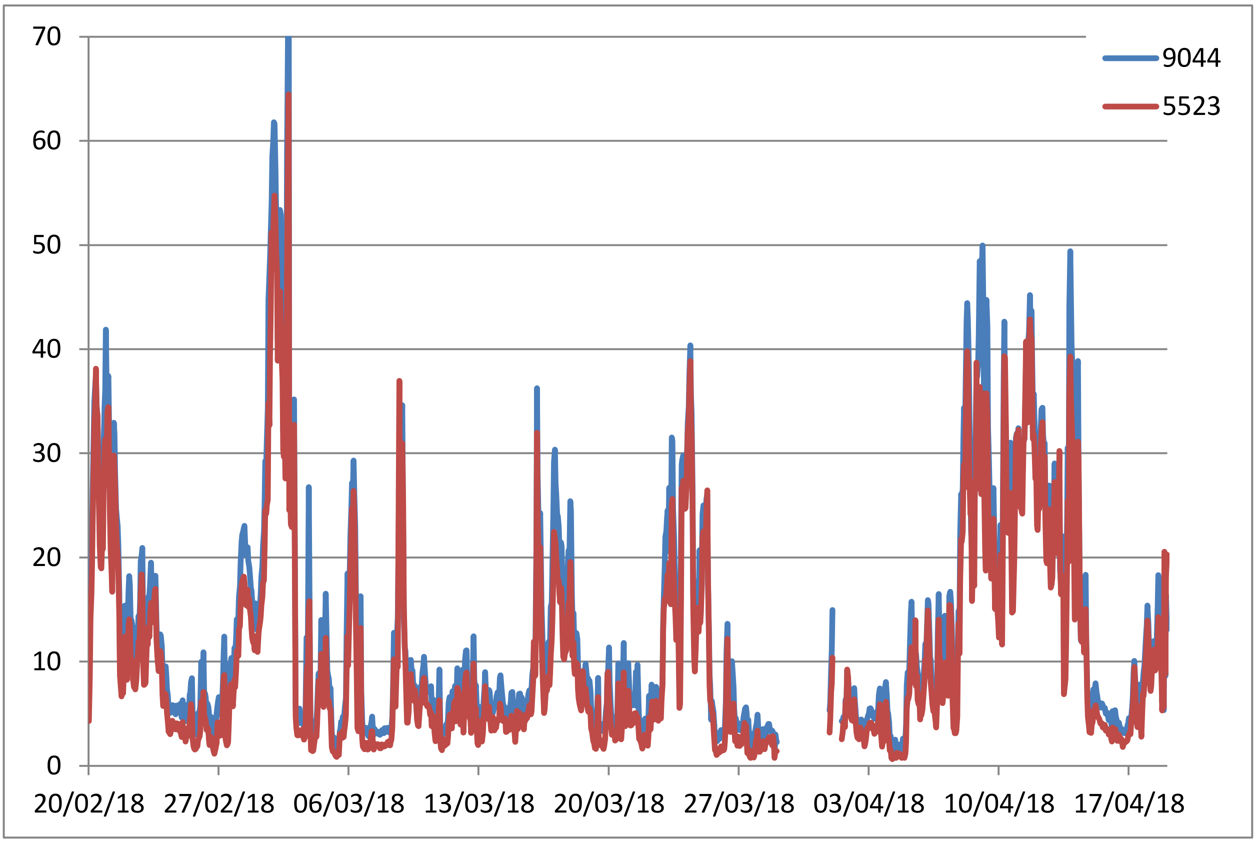


Figure 1. Luftdaten sensor 9044 vs. Luftdaten sensor 5523: PM2.5 1-hour mean (µg/m^3) during the period 19/02/18 – 19/04/18.

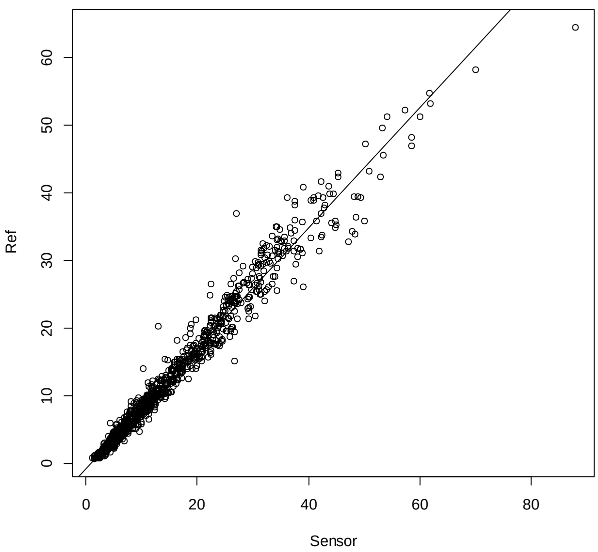
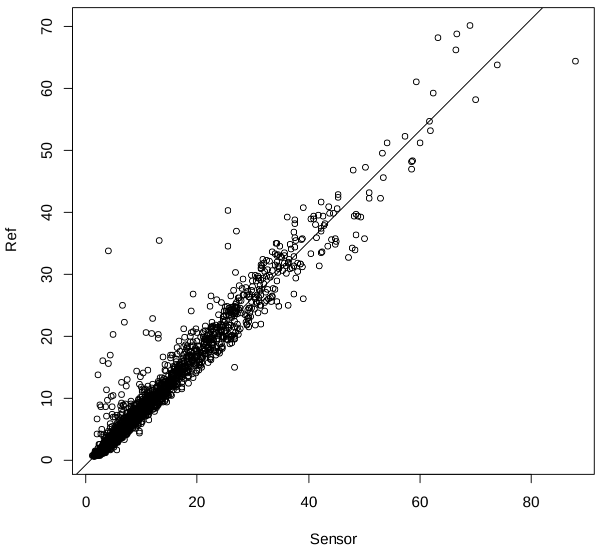
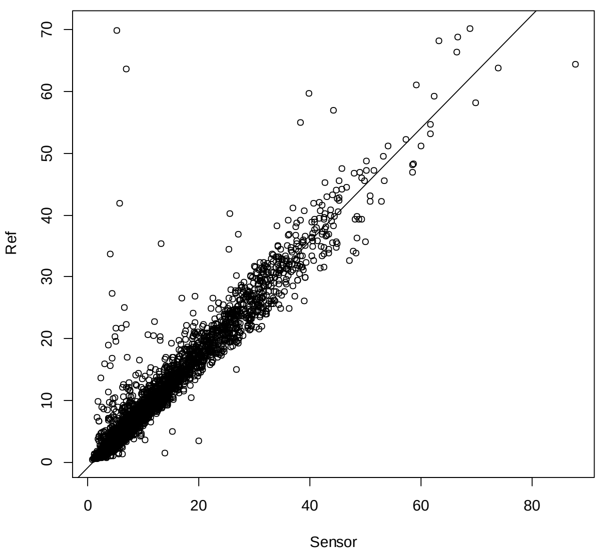
<table>
<tr>
<td>19/02 – 19/04</td>
<td>
</td>
<td>y = 0.8925x – 0.8919
R^2 = 0.9761</td>
</tr>
<tr>
<td>Winter</td>
<td>
</td>
<td>y = 0.9017x – 0.8472
R^2 = 0.9571</td>
</tr>
<tr>
<td>Whole year</td>
<td>
</td>
<td>y = 0.9142x – 0.8019
R^2 = 0.9369</td>
</tr>
</table>

Figure 2. Luftdaten sensor 9044 ("Sensor") vs. Luftdaten sensor 5523 ("Ref"): PM2.5 1-hour mean (µg/m^3).

Comparing Luftdaten and DEFRA sensors

Next we compare Luftdaten sensors with DEFRA (government) monitors. This time we check each DEFRA monitor against every Luftdaten sensor in the same town such that both have readings for at least 50% of the hours in the time period being considered.

Table 4 shows, for each time period, the number of pairs of sensors satisfying the criteria. It also shows the maximum R^2, the percentage of hours present, the two sensors in the pair, and the distance between them.

Table 4. Most correlated Luftdaten sensor and DEFRA monitor for each time period.

Time period	# pairs	Maximum R^2	% hours	Sensor 1	Sensor 2	Distance
19/02 – 19/04	13	0.84	94	4696	DEFRA-E	3308m
Winter	13	0.82	95	4696	DEFRA-E	3308m
Whole year	17	0.74	95	4696	DEFRA-E	3308m

It is worth mentioning that, although each of these is the maximum R^2 over all pairs of sensors, they are not outliers. For example, the second greatest R^2 for 19/02 – 19/04 is 0.8324 (for sensor 7675 and the Bristol DEFRA monitor, which are 1284m apart).

Figure 3 shows the concentration reported by the most highly correlated pair of sensors (Luftdaten 4696 and the Eastbourne DEFRA monitor) for each hour during the 19/02 – 19/04 time period. Figure 4 shows the linear correlation between these two sensors for each of the three time periods.

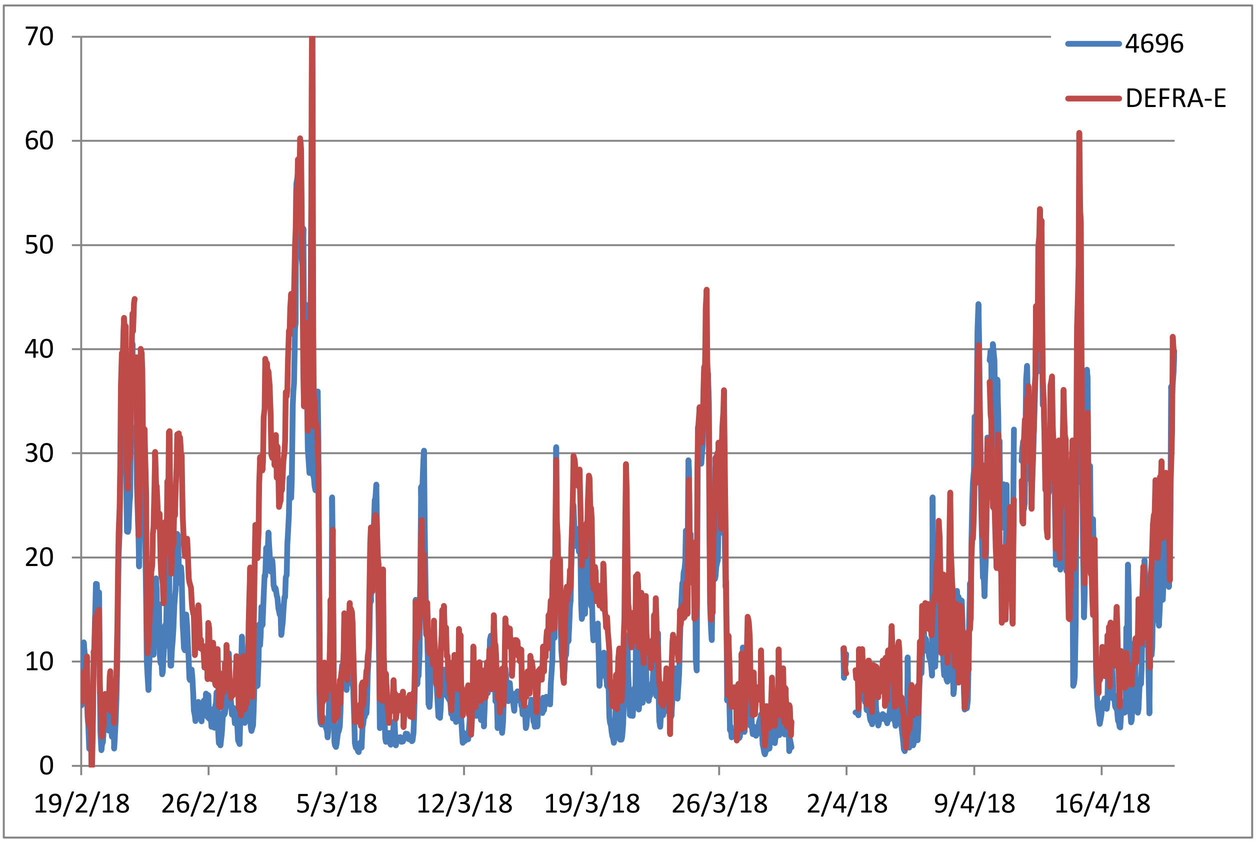


Figure 3. Luftdaten sensor 4696 vs. DEFRA Eastbourne monitor: PM2.5 1-hour mean (μg/m^3) during the period 19/02/18 – 19/04/18.

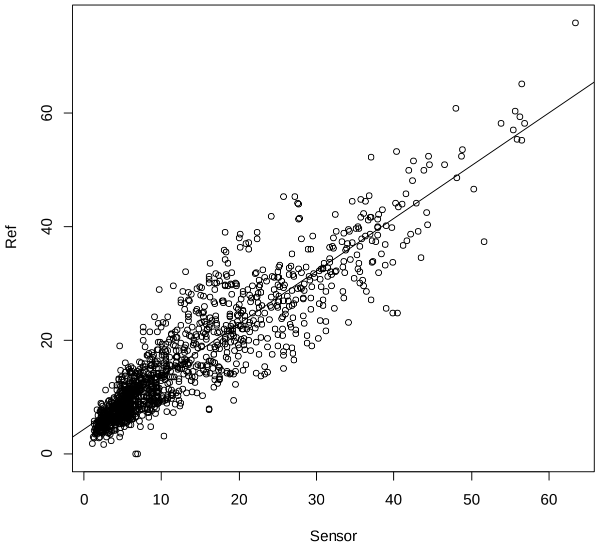
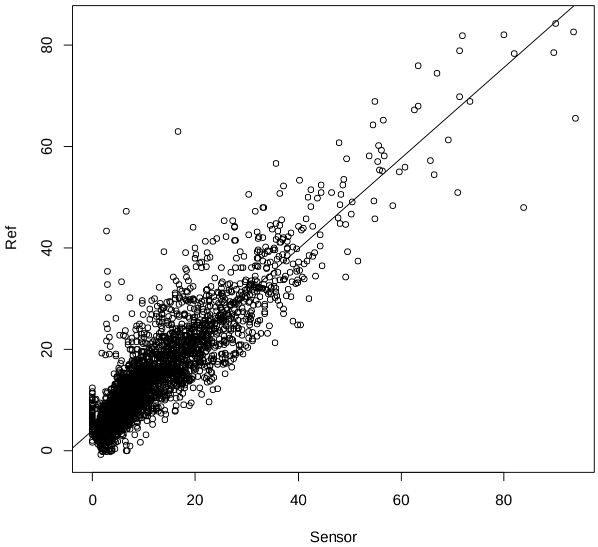
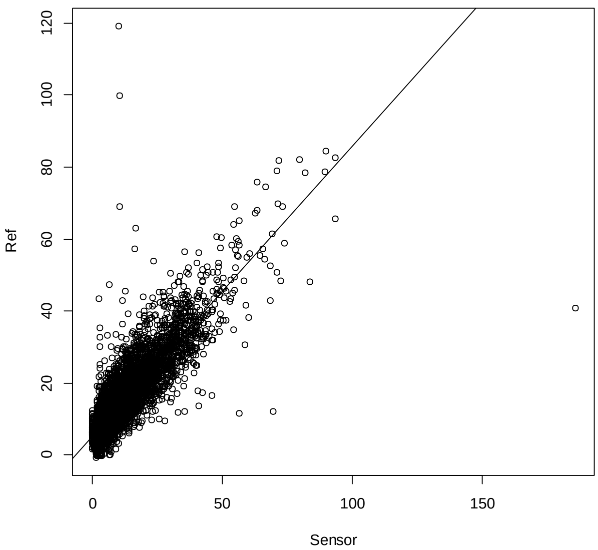
19/02 – 19/04		y = 0.9298x + 4.2576 R^2 = 0.8356
Winter		y = 0.8963x + 3.9109 R^2 = 0.8233
Whole year		y = 0.807x + 5.0839 R^2 = 0.7442

Figure 4. Luftdaten sensor 4696 (“Sensor”) vs. DEFRA Eastbourne monitor (“Ref”): PM2.5 1-hour mean (μg/m^3).

Discussion

1. The Luftdaten sensor seems to correlate well with the government (DEFRA) monitor (Figure 4). Given our assumption (*), R^2 is at least 0.84 for the period 19/02 – 19/04, which compares well with the commercial devices listed in Table 1. A similar good result was also found for the whole winter period (the coldest half of 2018); the commercial devices were not tested for such a long period [1-3].

2. The Luftdaten sensor does not correlate so well with the government (DEFRA) monitor when comparing for the whole year (Figure 4), because of lower agreement during the summer. This could be because of different conditions during the winter (for example, wide-area pollution and synchronised wood burning events) which would tend to increase correlation between separated sensors during winter, but this is unlikely because Table 3 shows good agreement between Luftdaten sensors for the whole year. Therefore, it seems likely that the discrepancy between Luftdaten sensors and DEFRA sensors is inherently larger during the summer than during the winter.

3. The tests on commercial sensors (Table 1) were only done for short periods in the colder months. The tests give no evidence about their performance for the rest of the year.

4. As well as R^2, the slope and intercept of the regression are interesting. The tests on the Purple Air PA-II [2] report a low slope for the correlation with both the GRIMM and BAM: the average is only 0.5745. For the correlation between the Luftdaten sensor and DEFRA monitor for 19/02 – 19/04, we find a slope of 0.9298. Both lines are plotted in Figure 5. It seems that the Purple Air PA-II overestimates the PM2.5 concentration when it is high, while the Luftdaten tracks the actual concentration much more closely (assuming that the GRIMM, BAM and DEFRA instruments are completely accurate) but slightly underestimates low readings.

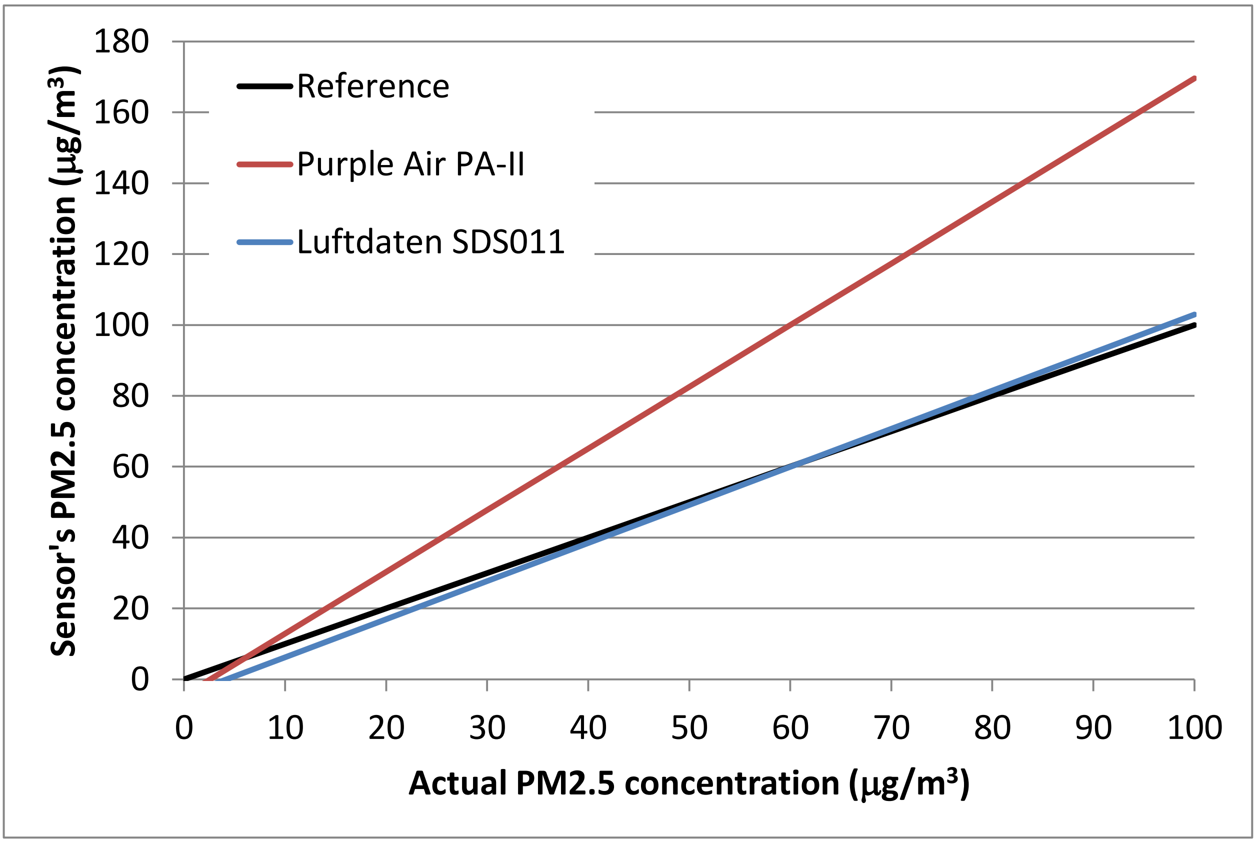


Figure 5. Regression lines for Luftdaten sensor and Purple Air PA-II sensor readings vs. actual (reference instrument) concentrations.

Points 2 and 4 are backed up by a simple comparison (Table 5) of the mean concentration in summer and winter from each city's DEFRA monitor with the mean of all Luftdaten sensors in the same city. The Luftdaten sensor means are slightly too low in winter, as suggested by Figure 5, but more seriously low in summer.

Table 5. Comparison between Luftdaten and DEFRA in summer and winter.

Town	Type	Winter		Summer	
		Mean	Std Dev	Mean	Std Dev
Bristol	Luftdaten	9.8	10.4	6.2	7.0
	DEFRA	12.0	9.4	12.2	7.6
Eastbourne	Luftdaten	10.4	11.1	7.1	8.2
	DEFRA	13.2	10.4	12.3	8.3
Aberdeen	Luftdaten	7.0	8.9	5.1	5.5
	DEFRA	7.5	5.9	6.2	4.5

Future work

It would be worthwhile to do the same analysis for Luftdaten sensors elsewhere, especially in Germany where the density of Luftdaten sensors is highest. It would also be useful to do the same for (e.g.) Purple Air sensors in an area with good coverage of them, such as California. Finally, if there are locations with both types of sensor installed, the two types could be directly compared in the field.

References

1. Field Evaluation Purple Air PM Sensor. Air Quality Sensor Performance Evaluation Center. 2016. http://www.aqmd.gov/docs/default-source/aq-spec/field-evaluations/purpleair---field-evaluation.pdf

2. Field Evaluation Purple Air (PA-II) PM Sensor. Air Quality Sensor Performance Evaluation Center. 2017. www.aqmd.gov/docs/default-source/aq-spec/field-evaluations/purple-air-pa-ii---field-evaluation.pdf

3. Field Evaluation Aeroqual AQY (v0.5). Air Quality Sensor Performance Evaluation Center. 2018. www.aqmd.gov/docs/default-source/aq-spec/field-evaluations/aeroqual-aqy-v0-5---field-evaluation.pdf