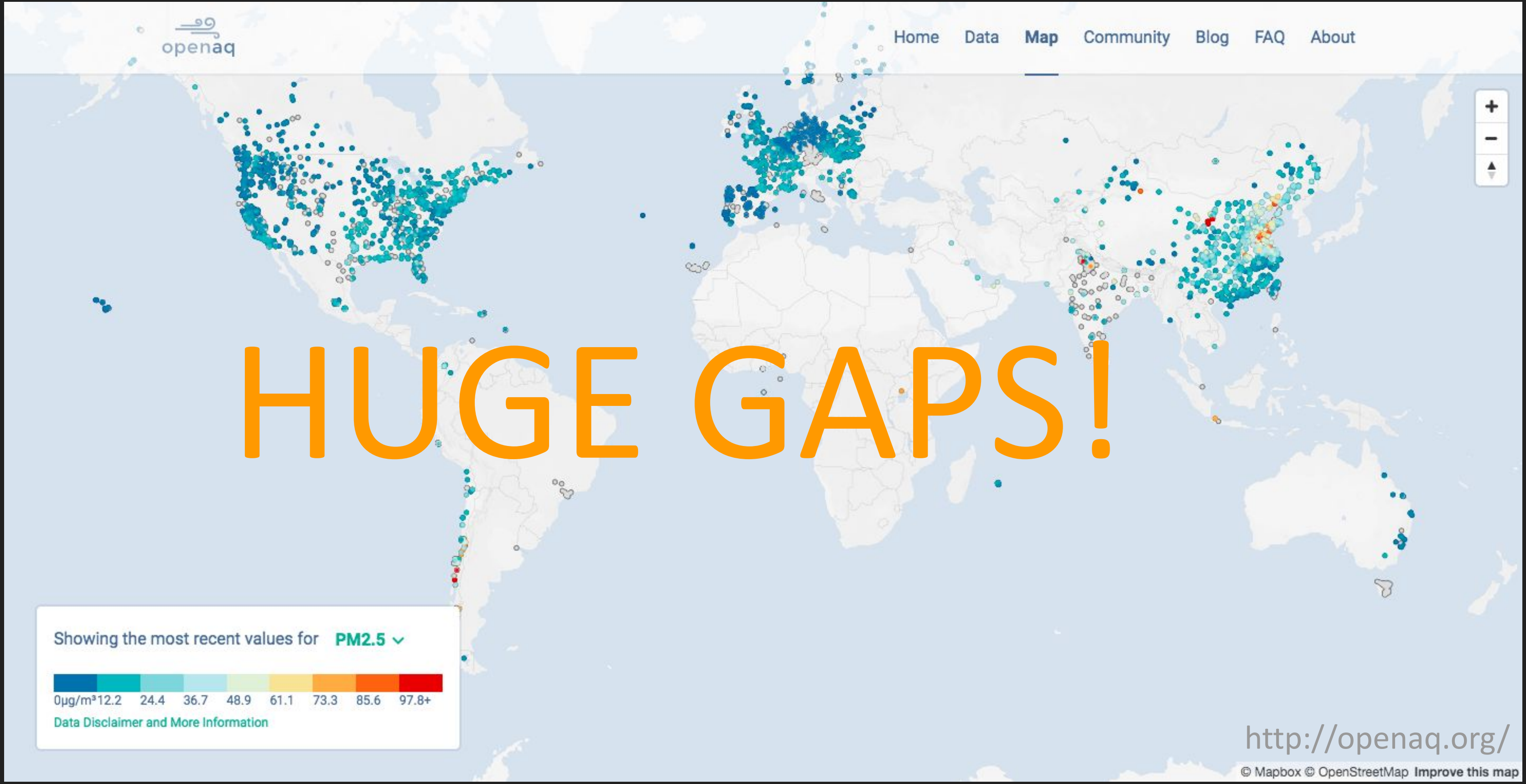
Denver

Calibrating Networks of Low-Cost Sensors

Priyanka deSouza
Assistant Professor
University of Colorado Denver

Air Quality Data reported by countries



Air Quality Monitors: high cost to low cost

https://archive.epa.gov/pesticides/region4/sesd/pm25/web/html/p2.html

http://senseable.mit.edu/cleanair-nairobi/

deSouza, P., Nthusi, V., Klopp, J.M., Shaw, B.E., Ho, W.O., Saffell, J., Jones, R. and Ratti, C., 2017. A Nairobi experiment in using low cost air quality monitors. Clean Air Journal, 27(2), pp.12-42.

Reference Air Quality Monitoring Station

-High accuracy
-High cost ($150,000-$200,000)

Low Cost Air Quality Monitor

-Low accuracy
-Low(er) cost (~< $2,500 as defined by USEPA Air Toolbox)

Calibrating Low-cost sensor data by co-locating sensors with reference monitors

Calibration has been shown to be impacted by the following decisions (Giordano et al., 2021)

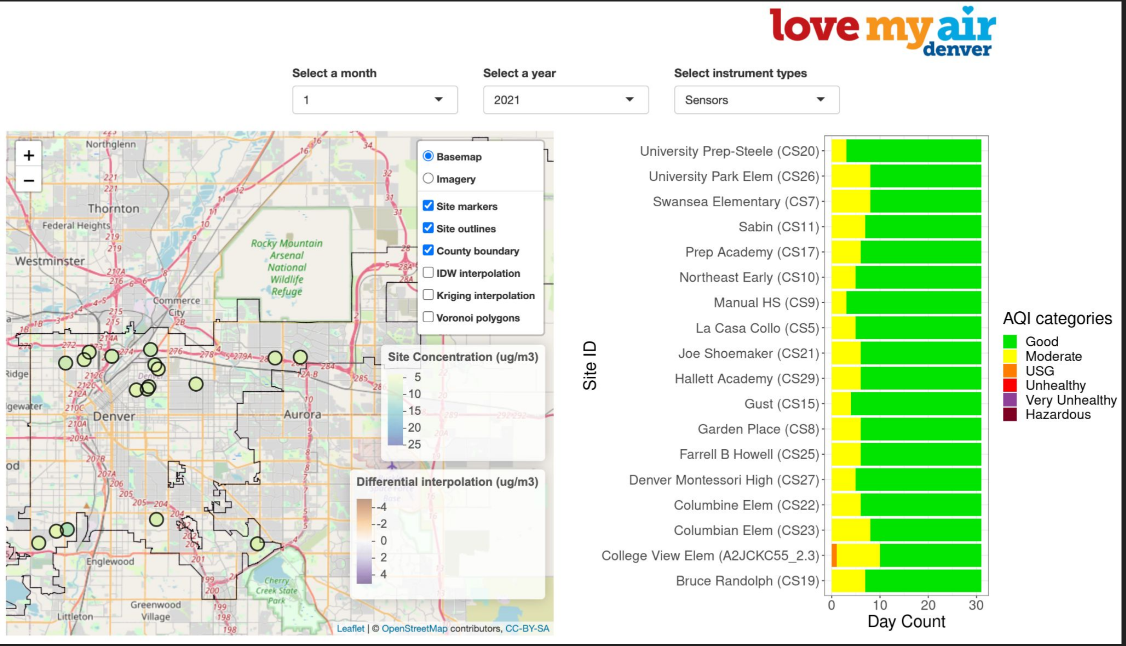
1) Choice of reference monitor
2) Time-averaging interval
3) How long the co-location takes place for
4) Cross-validation technique
5) Correction model used

As we move towards networks of low-cost sensors how ‘transferable’ are calibration algorithms across space and time?

Giordano, M.R., Malings, C., Pandis, S.N., Presto, A.A., McNeill, V.F., Westervelt, D.M., Beekmann, M. and Subramanian, R., 2021. From low-cost sensors to high-quality data: A summary of challenges and best practices for effectively calibrating low-cost particulate matter mass sensors. *Journal of Aerosol Science*, *158*, p.105833.

Calibrating Networks of Sensors

24 sensors
5 co-located sites
Which algorithm is most ‘transferable’ ?



deSouza, P., Kahn, R., Stockman, T., Obermann, W., Crawford, B., Wang, A., Crooks, J., Li, J. and Kinney, P., 2022. Calibrating Networks of Low-Cost Air Quality Sensors. *Atmospheric Measurement Techniques Discussions*, pp.1-34.

Traditional Evaluation of C1 & 2 at Co-Location Sites

C1: Correction developed on data during the entire period of network operation

C2: On-the-fly correction developed using data for the same week of measurement

ID	Name	Equation	C1		C2	
			R	RMSE ($\mu g/m^3$)	R	RMSE ($\mu g/m^3$)
	Raw Love My Air measurements					
0	Raw		0.927	6.469	-	-
1	Linear	$PM_{2.5, corrected} = PM_{2.5} \times s1 + b$	0.927	3.421	0.944	3.008
2	+RH	$PM_{2.5, corrected} = PM_{2.5} \times s_1 + RH \times s_2 + b$	0.929	3.379	0.948	2.904
5	+RH x T	$PM_{2.5, corrected} = PM_{2.5} \times s_1 + RH \times s_2 + T \times s_3 + RH \times T \times s_4 + b$	0.934	3.260	0.953	2.782
7	+D x T	$PM_{2.5, corrected} = PM_{2.5} \times s_1 + D \times s_2 + T \times s_3 + D \times T \times s_4 + b$	0.928	3.409	0.952	2.798
12	PM x nonlinear RH	$PM_{2.5, corrected} = PM_{2.5} \times s_1 + RH2(1-RH) \times s_2 + RH2(1-RH) \times PM_{2.5} \times s_3 + b$	0.934	3.277	0.948	2.900
17	Random Forest (CV = LOSO)	$PM_{2.5, corrected} = f(PM_{2.5}, T, RH)$	0.983	1.713	0.988	1.450

Traditional Evaluation of C3 & 4 at Co-Location Sites

C3
Correction developed using measurements made in the first two weeks of January

C4
Correction developed using measurements from the first two weeks of January and the first two weeks in May

ID	Name	Equation	C3		C4	
			R	RMSE ($\mu g/m^3$)	R	RMSE ($\mu g/m^3$)
	Raw Love My Air measurements					
0	Raw		0.907	5.008	0.898	3.983
1	Linear	$PM_{2.5,\ corrected} = PM_{2.5} \times s1 + b$	0.907	3.244	0.898	2.591
2	+RH	$PM_{2.5,\ corrected} = PM_{2.5} \times s_1 + RH \times s_2 + b$	0.915	3.110	0.909	2.453
5	+RH x T	$PM_{2.5,\ corrected} = PM_{2.5} \times s_1 + RH \times s_2 + T \times s_3 + RH \times T \times s_4 + b$	0.915	3.103	0.911	2.424
7	+D x T	$PM_{2.5,\ corrected} = PM_{2.5} \times s_1 + D \times s_2 + T \times s_3 + D \times T \times s_4 + b$	0.914	3.118	0.908	2.457
12	PM x nonlinear RH	$PM_{2.5,\ corrected} = PM_{2.5} \times s_1 + RH2(1-RH) \times s_2 + RH2(1-RH) \times PM_{2.5} \times s_3 + b$	0.926	2.898	0.920	2.299
17	Random Forest (CV= LOSO)	$PM_{2.5,\ corrected} = f(PM_{2.5}, T, RH)$	0.982	1.506	0.978	1.234

Evaluating C3 and C4 over the network operation

C3
Correction developed using measurements made in the first two weeks of January

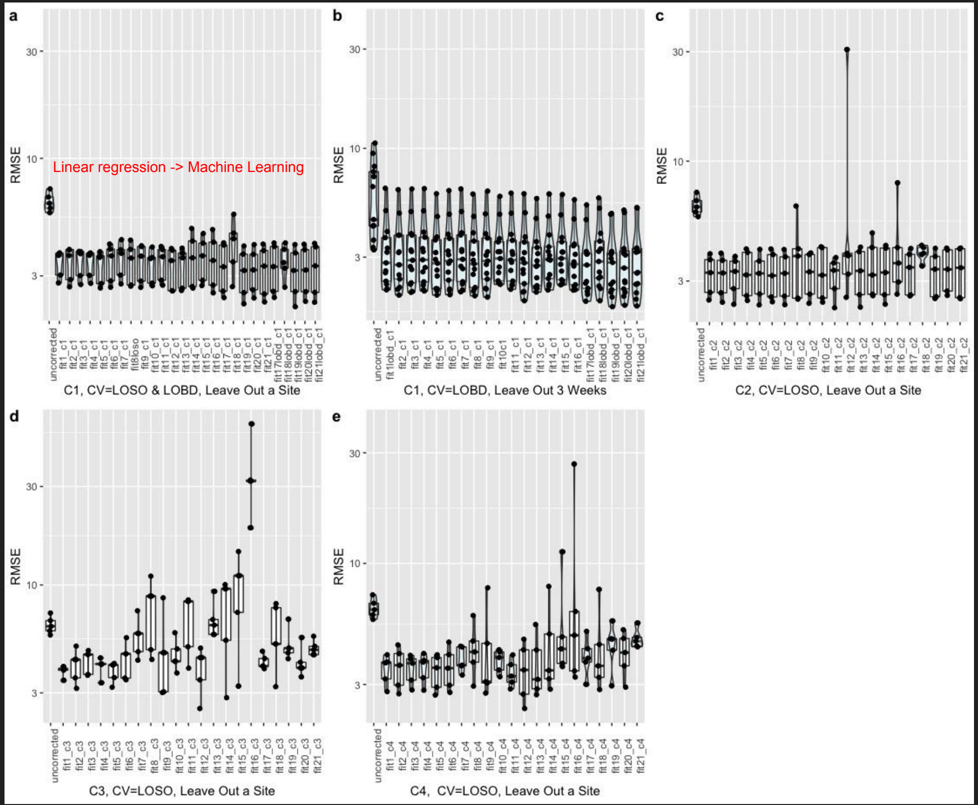
C4
Correction developed using measurements from the first two weeks of January and the first two weeks in May

ID	Name	Equation	C3		C4	
			R	**RMSE** ($\mu g/m^3$)	**R**	**RMSE** ($\mu g/m^3$)
	Raw Love My Air measurements					
0	Raw		0.927	6.469	-	-
1	Linear	$PM_{2.5, corrected} = PM_{2.5} \times s1 + b$	0.927	3.486	0.927	3.424
2	+RH	$PM_{2.5, corrected} = PM_{2.5} \times s_1 + RH \times s_2 + b$	0.928	3.618	0.929	3.462
5	+RH x T	$PM_{2.5, corrected} = PM_{2.5} \times s_1 + RH \times s_2 + T \times s_3 + RH \times T \times s_4 + b$	0.931	3.452	0.933	3.344
7	+D x T	$PM_{2.5, corrected} = PM_{2.5} \times s_1 + D \times s_2 + T \times s_3 + D \times T \times s_4 + b$	0.888	5.698	0.921	3.720
12	PM x nonlinear RH	$PM_{2.5, corrected} = PM_{2.5} \times s_1 + RH2(1-RH) \times s_2 + RH2(1-RH) \times PM_{2.5} \times s_3 + b$	0.931	3.510	0.932	3.403
17	Random Forest (CV = LOSO)	$PM_{2.5, corrected} = f(PM_{2.5}, T, RH)$	0.913	3.926	0.911	3.824

Leave one co-location site out

Leave out one co-location site. How well does calibration transfer to the left-out site/left out dates?

(a) C1 : Developed using training data over the entire period of measurement
(b) C1: Developed using training data over the entire period of measurement
(c) C2: Developed using data for the same week of measurement
(d) C3 : Developed using co-located data collected for 2 weeks at the beginning of the study
(e) C4: Developed using co-located data for two 2 week periods in Jan and May



deSouza, P., Kahn, R., Stockman, T., Obermann, W., Crawford, B., Wang, A., Crooks, J., Li, J. and Kinney, P., 2022. Calibrating Networks of Low-Cost Air Quality Sensors. *Atmospheric Measurement Techniques Discussions*, pp.1-34.

Transferability across time-averaging intervals

Applying corrections derived from hourly averaged data to minute-level measurements

ID	Name	Equation	C1		C2		C3		C4	
			R	RMSE ($\mu g/m^3$)	R	RMSE ($\mu g/m^3$)	R	RMSE ($\mu g/m^3$)	R	RMSE ($\mu g/m^3$)
	Raw Love My Air measurements									
0	Raw		0.497	16.409	-	-	-	-	-	-
1	Linear	$PM_{2.5,\ corrected} = PM_{2.5} \times s1 + b$	0.497	15.667	0.498	15.646	0.497	15.657	0.497	15.663
2	+RH	$PM_{2.5,\ corrected} = PM_{2.5} \times s_1 + RH \times s_2 + b$	0.495	15.678	0.500	15.618	0.492	15.721	0.494	15.686
5	+RH x T	$PM_{2.5,\ corrected} = PM_{2.5} \times s_1 + RH \times s_2 + T \times s_3 + RH \times T \times s_4 + b$	0.499	15.634	0.500	15.621	0.495	15.669	0.498	15.640
7	+D x T	$PM_{2.5,\ corrected} = PM_{2.5} \times s_1 + D \times s_2 + T \times s_3 + D \times T \times s_4 + b$	0.470	15.928	0.014	323.684	0.018	257.153	0.032	135.647
12	PM x nonlinear RH	$PM_{2.5,\ corrected} = PM_{2.5} \times s_1 + RH2(1-RH) \times s_2 + RH2(1-RH) \times PM_{2.5} \times s_3 + b$	0.496	15.659	0.497	15.650	0.494	15.705	0.495	15.681
17	Random Forest	$PM_{2.5,\ corrected} = f(PM_{2.5}, T, RH)$	0.505	15.565	0.510	15.527	0.489	15.863	0.488	15.821

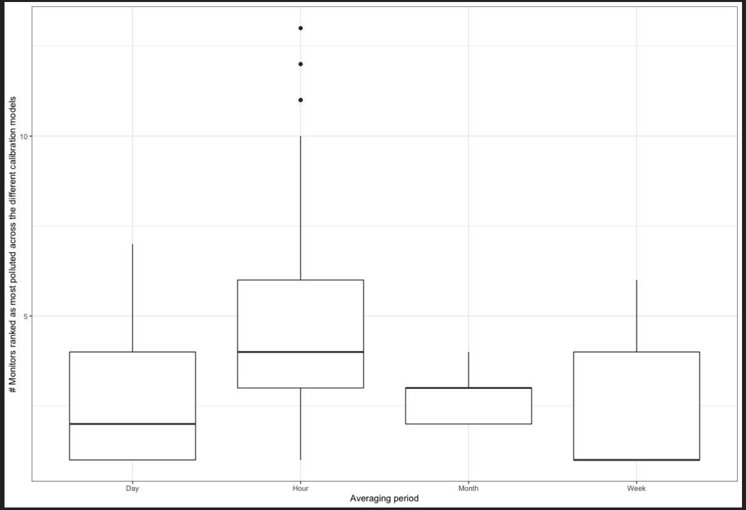
Transferability in pollution regimes

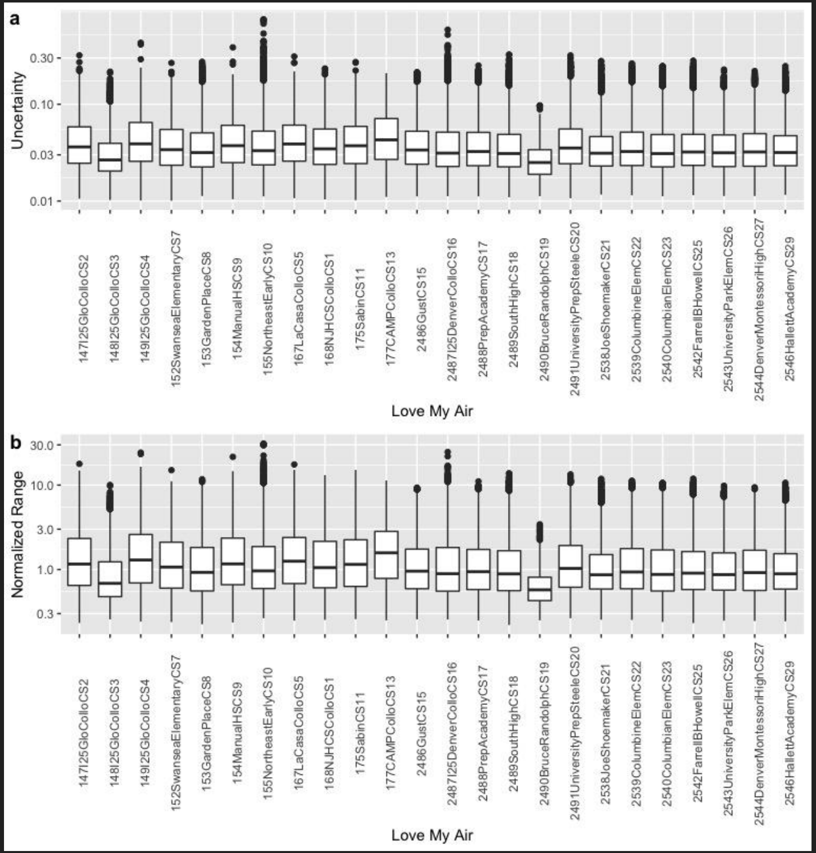
ID	Name	Equation	$PM_{2.5} > 30\ \mu g/m^3$ (n = 1038 measurements)								$PM_{2.5} \leq 30\ \mu g/m^3$ (n=26300 measurements)							
			C1		C2		C3		C4		C1		C2		C3		C4	
			R	RMSE	R	RMSE	R	RMSE	R	RMSE	R	RMSE	R	RMSE	R	RMSE	R	RMSE
0	Raw		0.797	14.928 (0.350)	-	-	-	-	-	-	0.915	5.891 (0.646)	-	-	-	-	-	-
1	Linear	$PM_{2.5,\ corrected} = PM_{2.5} \times s1 + b$	0.797	11.263 (0.264)	0.834	9.522 (0.223)	0.797	10.556 (0.248)	0.797	11.105 (0.260)	0.915	2.676 (0.294)	0.921	2.414 (0.265)	0.915	2.869 (0.315)	0.915	2.705 (0.297)
2	+RH	$PM_{2.5,\ corrected} = PM_{2.5} \times s_1 + RH \times s_2 + i$	0.802	11.083 (0.260)	0.838	9.316 (0.218)	0.806	9.379 (0.220)	0.804	9.979 (0.234)	0.917	2.650 (0.291)	0.927	2.311 (0.254)	0.913	3.184 (0.349)	0.915	2.921 (0.320)
5	+RH x T	$PM_{2.5,\ corrected} = PM_{2.5} \times s_1 + RH \times s_2 + T \times s_3 + RH \times T \times s_4 + i$	0.806	10.772 (0.253)	0.852	8.866 (0.208)	0.804	9.636 (0.226)	0.806	9.868 (0.231)	0.923	2.543 (0.279)	0.933	2.224 (0.244)	0.917	2.954 (0.324)	0.922	2.790 (0.306)
7	+D x T	$PM_{2.5,\ corrected} = PM_{2.5} \times s_1 + D \times s_2 + T \times s_3 + D \times T \times s_4 + i$	0.799	11.211 (0.263)	0.847	8.946 (0.210)	0.789	8.981 (0.211)	0.798	10.033 (0.235)	0.916	2.668 (0.293)	0.933	2.231 (0.245)	0.863	5.529 (0.607)	0.908	3.226 (0.354)
12	PM x nonlinear RH	$PM_{2.5,\ corrected} = PM_{2.5} \times s_1 + RH2(1-RH) \times s_2 + RH2(1-RH) \times PM_{2.5} \times s_3 + i$	0.821	10.695 (0.251)	0.844	9.157 (0.215)	0.815	9.322 (0.219)	0.814	9.712 (0.228)	0.923	2.579 (0.283)	0.927	2.331 (0.256)	0.920	3.063 (0.336)	0.920	2.884 (0.316)
17	Random Forest	$PM_{2.5,\ corrected} = f(PM_{2.5}, T, RH)$	0.940	5.380 (0.126)	0.953	4.670 (0.109)	0.651	13.773 (0.323)	0.610	15.006 (0.352)	0.973	1.382 (0.152)	0.982	1.151 (0.126)	0.903	2.922 (0.321)	0.917	2.513 (0.276)

Calibrating Networks of Sensors

What uncertainties in measurement does the calibration algorithm choice induce?

How sensitive are hotspots in the network are to the calibration algorithm used to averaging by different time periods?





Conclusions

We need to:

1) Evaluate models under different conditions (e.g. pollution concentrations) to evaluate the circumstances under which different calibration algorithms do well to determine which model to use for which use-case.

2) Determine how well calibration adjustments can be transferred to other locations.

3) Examine how well calibration adjustments can be transferred to other time periods.

4) Evaluate how well calibration algorithms developed for a specific time-scale transfer to measurements at other time intervals.

5) Use a variety of approaches to quantify transferability

6) Investigate how adopting a certain time-scale for averaging measurements could mitigate the uncertainty induced by the calibration process.