Novel approaches for exposure estimation and epidemiological analysis:

High-resolution air quality mapping in Great Britain (2003-2021) with ensemble machine-learning and remote sensing data

Arturo de la Cruz Libardi, LSHTM @ ISEEY Rennes, 6th June 2024

Air pollution and health

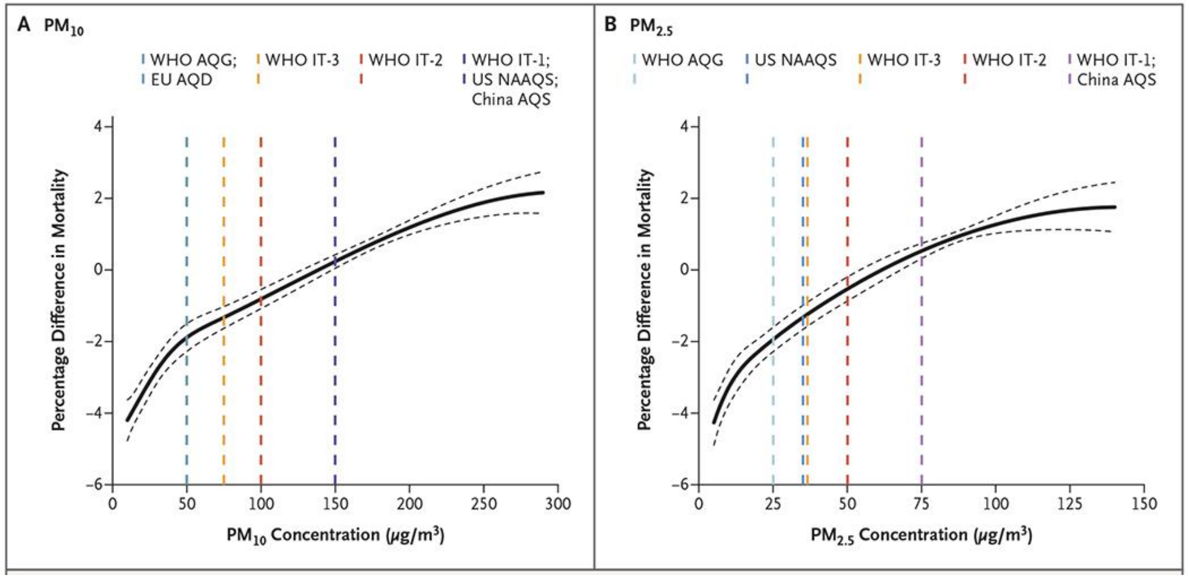
- Air pollution is a global problem
    - WHO air quality guidelines unmet for 99% of people in 2019
    - 4.2 million annual premature deaths (outdoor only)
    - Short and long-term effects

The NEW ENGLAND JOURNAL of MEDICINE

ESTABLISHED IN 1812 AUGUST 22, 2019 VOL. 381 NO. 8

Ambient Particulate Air Pollution and Daily Mortality in 652 Cities

C. Liu, R. Chen, F. Sera, A.M. Vicedo-Cabrera, Y. Guo, S. Tong, M.S.Z.S. Coelho, P.H.N. Saldiva, E. Lavigne, P. Matus, N. Valdes Ortega, S. Osorio Garcia, M. Pascal, M. Stafoggia, M. Scortichini, M. Hashizume, Y. Honda, M. Hurtado-Díaz, J. Cruz, B. Nunes, J.P. Teixeira, H. Kim, A. Tobias, C. Íñiguez, B. Forsberg, C. Åström, M.S. Ragettli, Y.-L. Guo, B.-Y. Chen, M.L. Bell, C.Y. Wright, N. Scovronick, R.M. Garland, A. Milojevic, J. Kyselý, A. Urban, H. Orru, E. Indermitte, J.J.K. Jaakkola, N.R.I. Ryti, K. Katsouyanni, A. Analitis, A. Zanobetti, J. Schwartz, J. Chen, T. Wu, A. Cohen, A. Gasparrini, and H. Kan



Open access Original research

BMJ Open Diabetes Research & Care

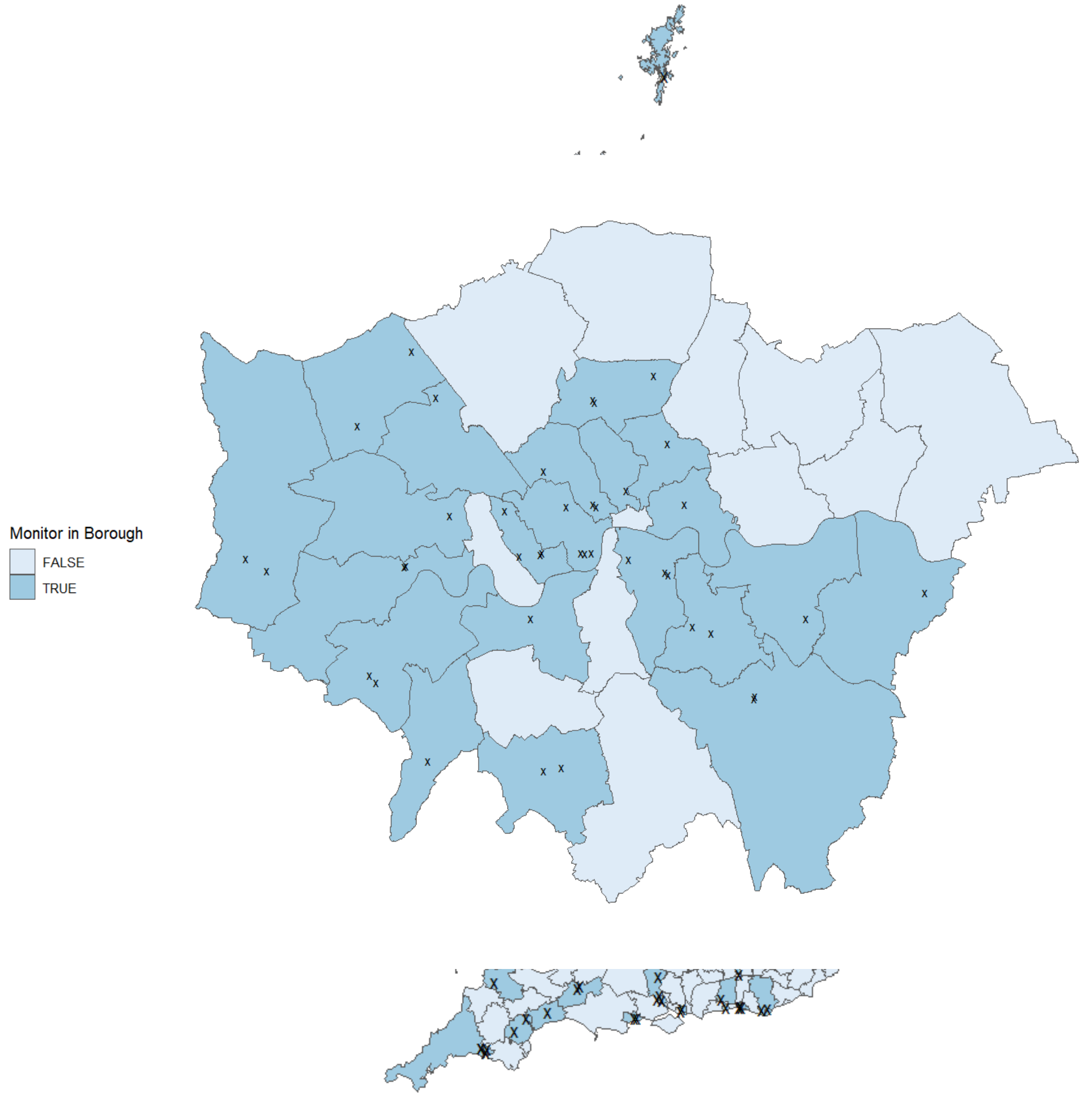
$PM_{2.5}$ exposure, glycemic markers and incidence of type 2 diabetes in two large Indian cities

Siddhartha Mandal,[1] Suganthi Jaganathan,[1] Dimple Kondal,[1,2] Joel D Schwartz,[3] Nikhil Tandon,[4] Viswanathan Mohan,[5] Dorairaj Prabhakaran,[1,2] K M Venkat Narayan[6]

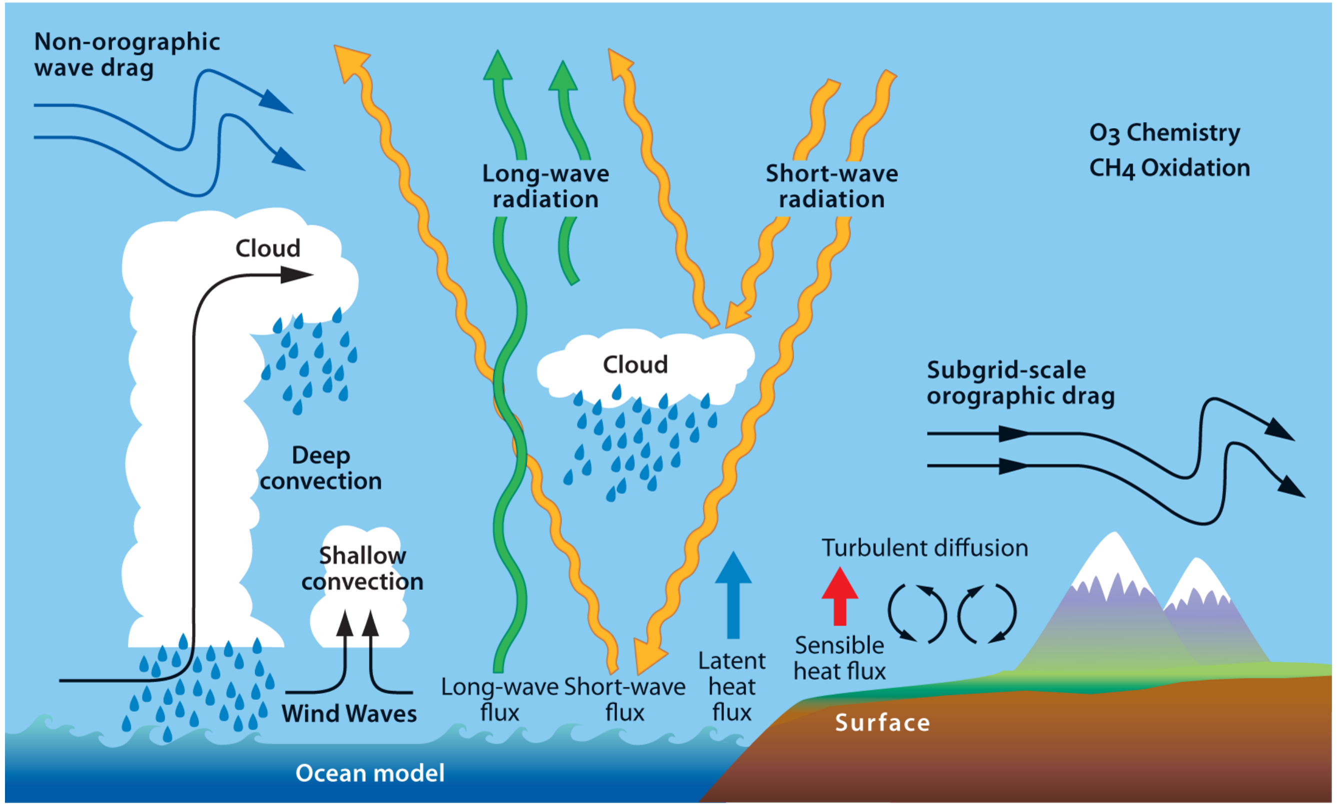
Results We observed that 10 μg/m^3 differences in monthly average exposure to $PM_{2.5}$ was associated with a 0.40 mg/dL increase in FPG (95% CI 0.22 to 0.58) and 0.021 unit increase in HbA1c (95% CI 0.009 to 0.032). Further, 10 μg/m^3 differences in annual average $PM_{2.5}$ was associated with 1.22 (95% CI 1.09 to 1.36) times increased risk of incident T2DM, with non-linear exposure response.

Ambient outdoor air pollution exposure assessment

- Health data needs to be linked to environmental exposure
- AP is spatially continuous and heterogeneous*
- Many administrative areas are unmonitored at the national level



- Chemistry Transport Models (CTM)
- Land-use regression
- Machine-learning



Data

- Air pollution observations
- CTM
- Reanalysis, Satellite
- Land features

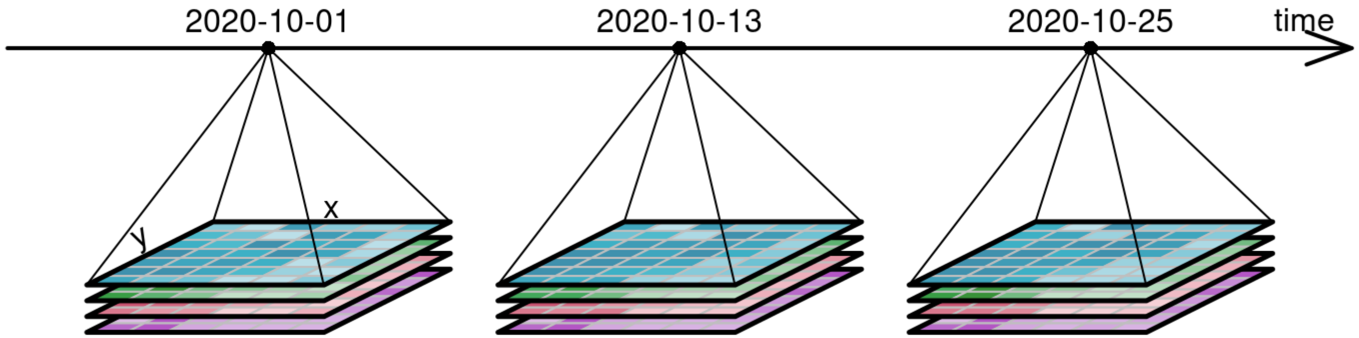
- UK Air quality networks
- EMEP4UK
- ERA's, Terra/Aqua, OMI
- Nightlight, roads, traffic, elevation, vegetation, use type, …

Data

- Different data types

- To a gridded support
- To a common support

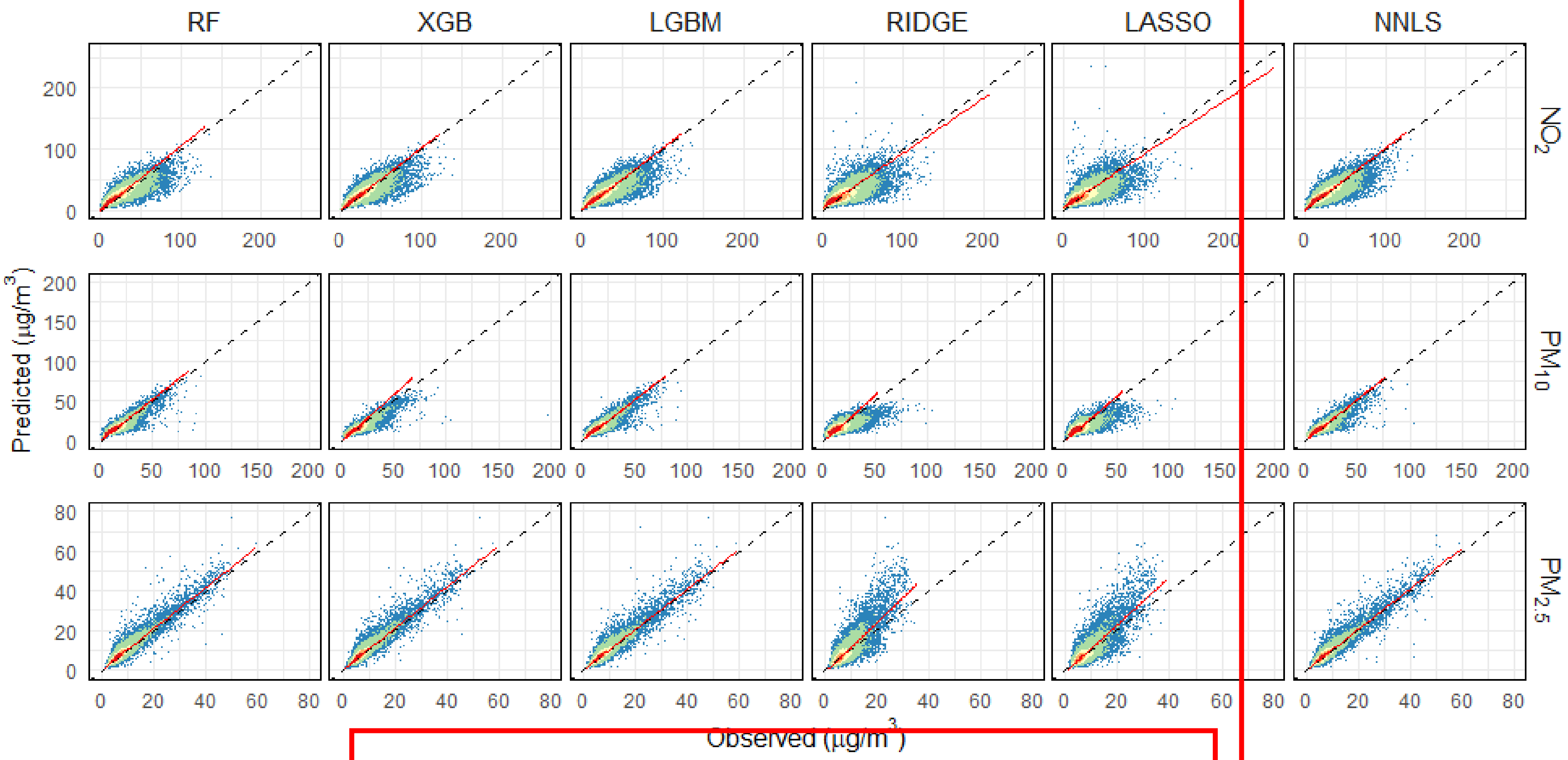
- Training dataset assembly



Predictive modelling

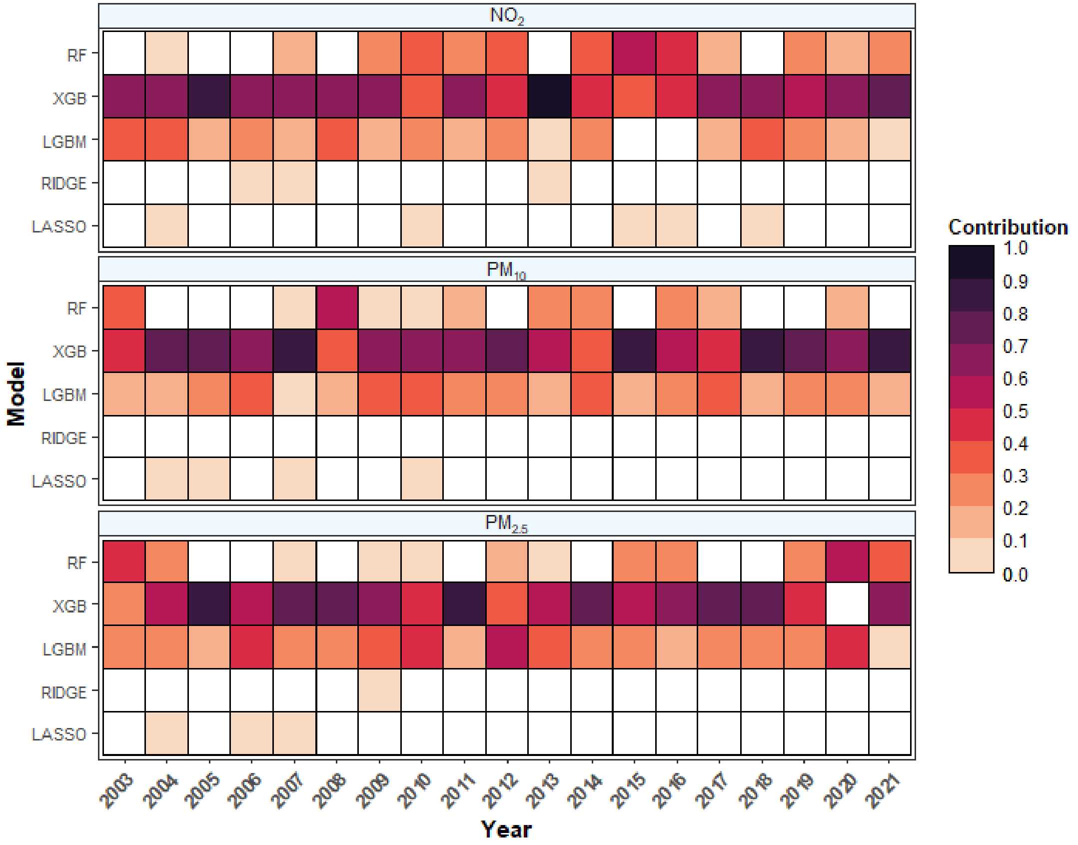
- Ensemble make-up
  - Random Forest, Extreme Gradient Boosting, LightGBM, Lasso, Ridge
  - Non-Negative Least Squares
- Performance evaluation
  - 10x by-monitor cross-validation
  - R2, RMSE

Results: observed~predicted, 2019

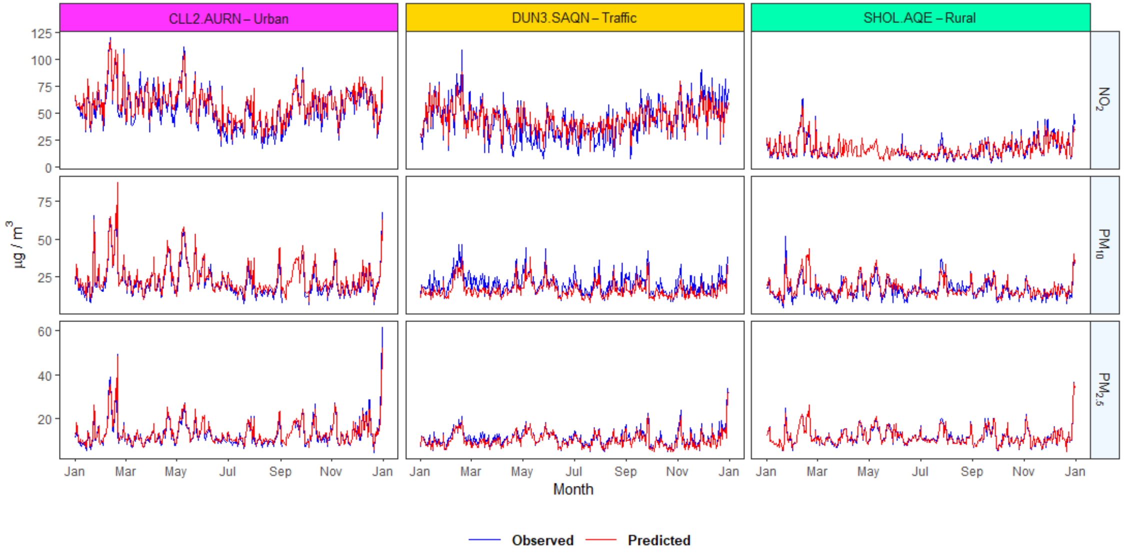


	R2 (min-max)	RMSE (ug/m3)	Inter.	Slope
NO2	0.691 (0.611-0.792)	12.267	1.623	1.01
PM10	0.706 (0.609-0.786)	6.301	0.204	1.029
PM2.5	0.822 (0.746-0.888)	3.115	0.105	1.022

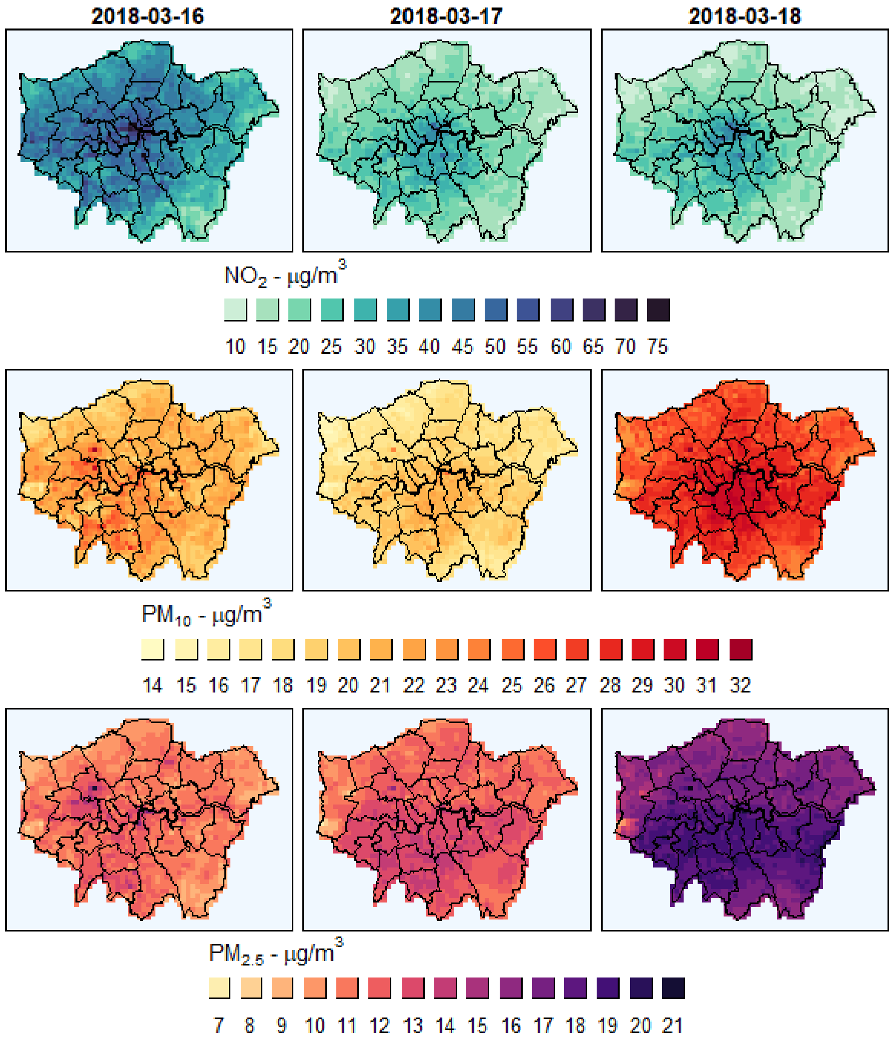
Results: base model contribution



Results

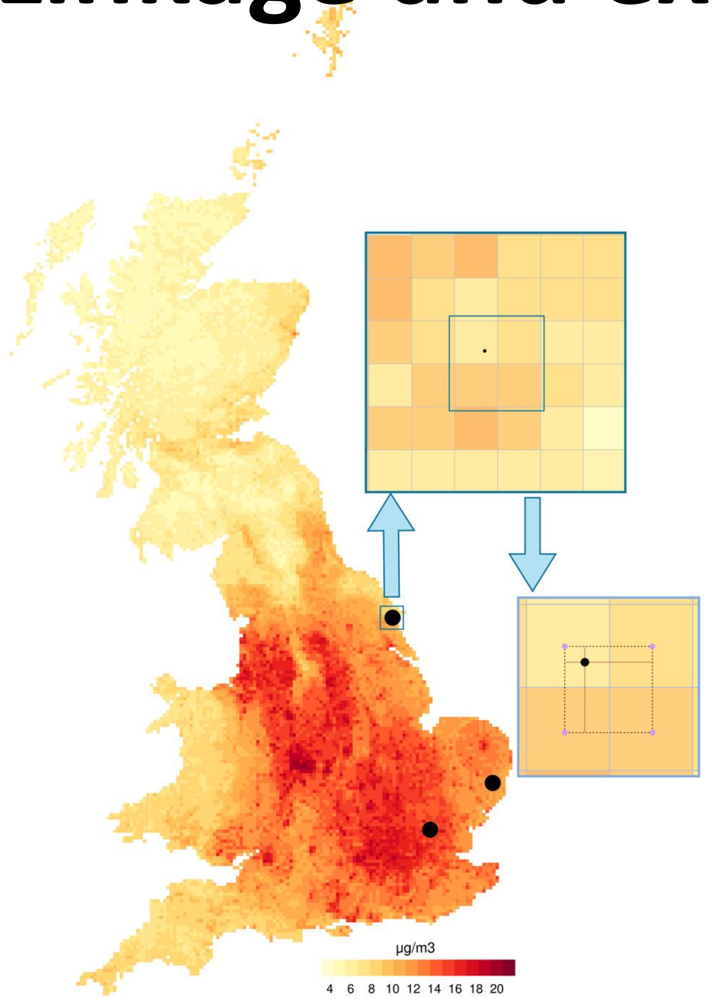


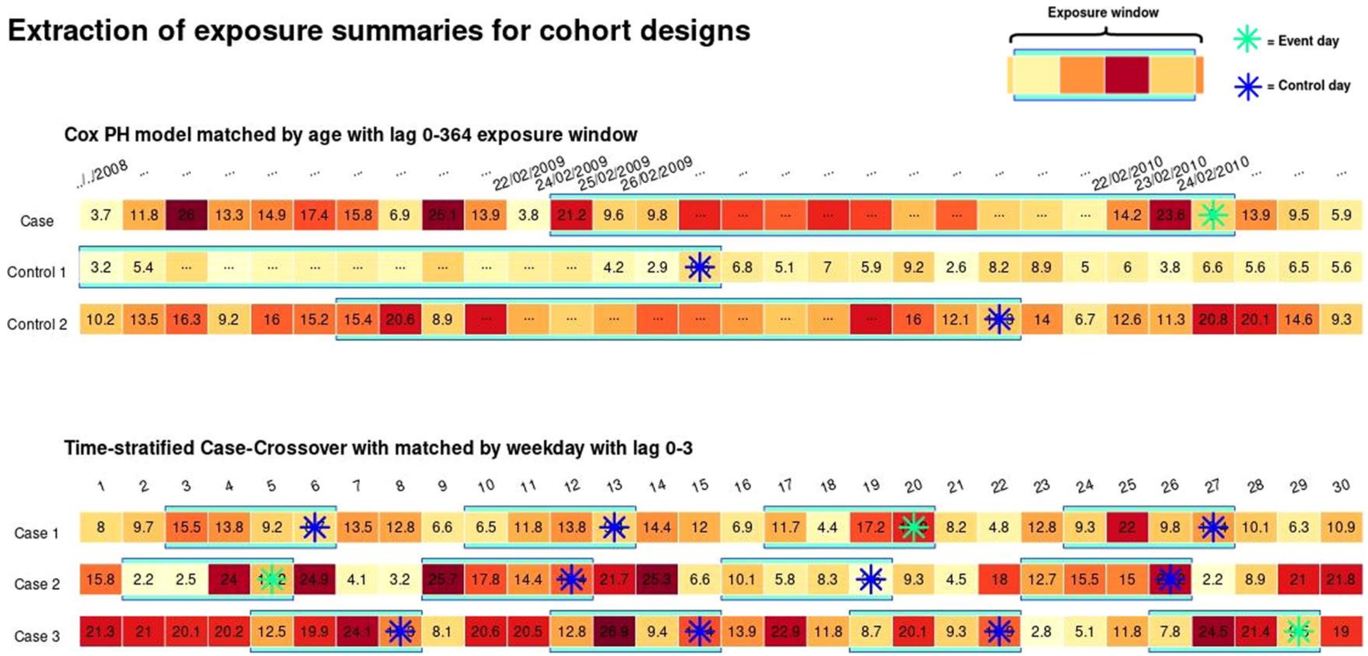
CLL2.AURN – Urban
DUN3.SAQN – Traffic
SHOL.AQE – Rural
NO2
PM10
PM2.5
µg / m3
Jan Mar May Jul Sep Nov Jan
Month
Observed
Predicted



2018-03-16
2018-03-17
2018-03-18
NO2 - µg/m3
10 15 20 25 30 35 40 45 50 55 60 65 70 75
PM10 - µg/m3
14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32
PM2.5 - µg/m3
7 8 9 10 11 12 13 14 15 16 17 18 19 20 21

Linkage and exposure series





Final slide for conclusion

Cited

1. World Health Organization. Ambient (outdoor) air pollution. https://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health (2022).
2. Liu, C. *et al.* Ambient Particulate Air Pollution and Daily Mortality in 652 Cities. *N Engl J Med* **381**, 705–715 (2019).
3. Mandal, S. *et al.* PM2.5 exposure, glycemic markers and incidence of type 2 diabetes in two large Indian cities. *BMJ Open Diabetes Res Care* **11**, e003333 (2023).
4. Pebesma, E. & Bivand, R. Spatial Data Science. https://r-spatial.org/book/ (2023)
5. Vanoli, J. *et al.* Reconstructing individual-level exposures in cohort analyses of environmental risks: an example with the UK Biobank. *J Expo Sci Environ Epidemiol* (2024) doi:10.1038/s41370-023-00635-w.

Acknowledgements to co-authors and supervisors:

Pierre Masselot, Jacopo Vanoli, Rochelle Schneider, Malcolm Mistry, Ai Milojevic, Emily Nightingale, Antonio Gasparrini

Thank you

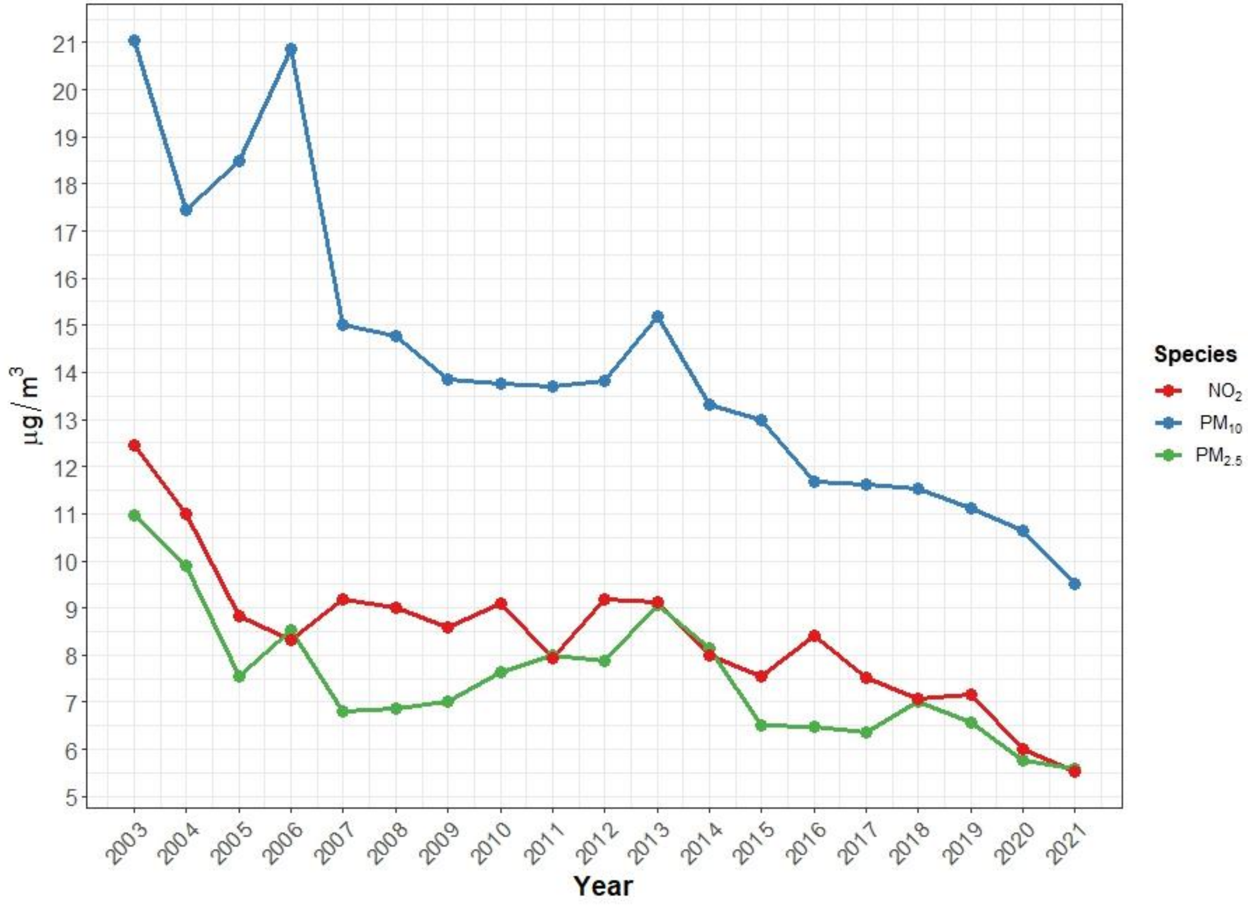
Limitations

- Monitor clustering and only one CV type used
- Sparse observations in early years
- Feature resolution

Air pollution summary data from monitors

	NO_2			PM_{10}			$PM_{2.5}$		
Year	Obs.	Mon.	Mean (ug/m^3)	Obs.	Mon.	Mean (ug/m^3)	Obs.	Mon.	Mean (ug/m^3)
2003	101,697	322	42.04	60,996	179	29.03	3,435	10	16.65
2004	103,968	322	38.14	71,667	209	24.91	3,426	10	14.93
2005	112,733	347	37.69	72,504	217	25.66	4,714	14	14.25
2006	119,377	370	37.86	79,783	238	26.17	5,297	17	14.65
2007	125,175	397	37.21	79,476	238	24.3	5,088	16	13.01
2008	139,012	426	37.29	100,862	301	22.38	9,091	52	12.09
2009	142,418	436	36.78	98,502	302	21.62	23,329	77	12.53
2010	138,769	435	38.18	92,824	284	21.61	26,486	83	13.3
2011	138,375	430	35.2	90,189	272	22.78	25,596	82	14.18
2012	137,768	430	36.54	87,882	263	20.55	27,897	84	12.93
2013	139,315	427	34.87	83,598	252	21.33	26,780	82	13.04
2014	134,480	425	34.37	84,546	258	19.76	28,427	91	12.27
2015	154,133	470	32.01	88,808	268	18.25	34,854	109	10.1
2016	172,965	528	33.6	103,521	309	18.1	50,101	157	10.28
2017	182,779	538	30.97	106,485	316	17.12	55,945	175	9.48
2018	179,341	530	29.09	100,771	296	17.75	57,711	172	9.33
2019	190,214	568	28.33	112,599	337	17.15	66,175	198	9.14
2020	232,436	700	20.33	110,651	323	15.4	68,243	204	7.49
2021	230,039	697	21.87	111,654	326	14.84	74,954	226	7.62
Mean	**151,315**	**463.05**	**33.81**	**91,438**	**273.05**	**20.98**	**31,450**	**97.84**	**11.96**
Total	**2,874,994**			**1,737,318**			**597,549**		

Air pollution trends from full grid predictions



Feature importance

Rank	NO2			PM10			PM2.5		
	RF	**XGB**	**LGBM**	**RF**	**XGB**	**LGBM**	**RF**	**XGB**	**LGBM**
1	EMEP_NO2	EMEP_NO2	EMEP_NO2	EMEP_pm25rh50	EMEP_pm25rh50	EMEP_pm25rh50	EMEP_pm25rh50	EMEP_pm25rh50	EMEP_pm25rh50
2	EMEP_NO	EMEP_NO	d2hs	EMEP_pm10rh50	EMEP_pm10rh50	EMEP_pm10rh50	EMEP_pm10rh50	EMEP_pm10rh50	EMEP_pm10rh50
3	d2hs	d2hs	EMEP_NO	EMEP_SO4	EMEP_SO4	doy	EMEP_SO4	EMEP_SO4	doy
4	DEFRA_no2	DEFRA_no2	respop	precipitation	winddirection	precipitation	winddirection	doy	EMEP_SO4
5	d2bkg	respop	DEFRA_no2	winddirection	precipitation	winddirection	blh00	winddirection	idw1background
6	workpop	d2bkg	temp2m	mslp	doy	mslp	precipitation	blh00	winddirection
7	idw2background	workpop	doy	blh00	mslp	idw1background	idw1background	precipitation	idw2hotspot
8	idw1background	temp2m	windspeed	doy	blh00	EMEP_SO4	idw2background	mslp	precipitation
9	respop	idw1background	idw2hotspot	EMEP_dust	EMEP_dust	EMEP_dust	idw2hotspot	idw1background	blh00
10	idw1hotspot	windspeed	d2bkg	EMEP_SO2	temp2m	blh00	doy	idw2background	idw2background
11	idw2hotspot	idw1hotspot	workpop	idw1background	idw1background	d2hs	EMEP_SO2	idw2hotspot	mslp
12	temp2m	idw2background	idw1hotspot	idw1hotspot	d2hs	idw1hotspot	mslp	EMEP_seasalt	EMEP_seasalt
13	windspeed	doy	idw2background	idw2hotspot	EMEP_seasalt	idw2hotspot	EMEP_seasalt	EMEP_dust	EMEP_dust
14	nightlight	idw2hotspot	idw1background	temp2m	idw1hotspot	EMEP_seasalt	EMEP_dust	idw1hotspot	temp2m
15	imperviousness	blh00	blh00	EMEP_seasalt	idw2hotspot	temp2m	idw1hotspot	blh12	idw1hotspot

Future directions

- More effective use of EMEP4UK CTM and vector type data
- Test robustness with different cv-strategies
- Quantify uncertainty
- Deep-learning