Determinants of residential indoor VOC and $PM_{2.5}$ levels in six European cities

– the EXPOLIS study

Presented by: **Hak Kan Lai**

Email contact: hakkan.lai@imperial.ac.uk

Imperial College, University of London http://www.env.ic.ac.uk/

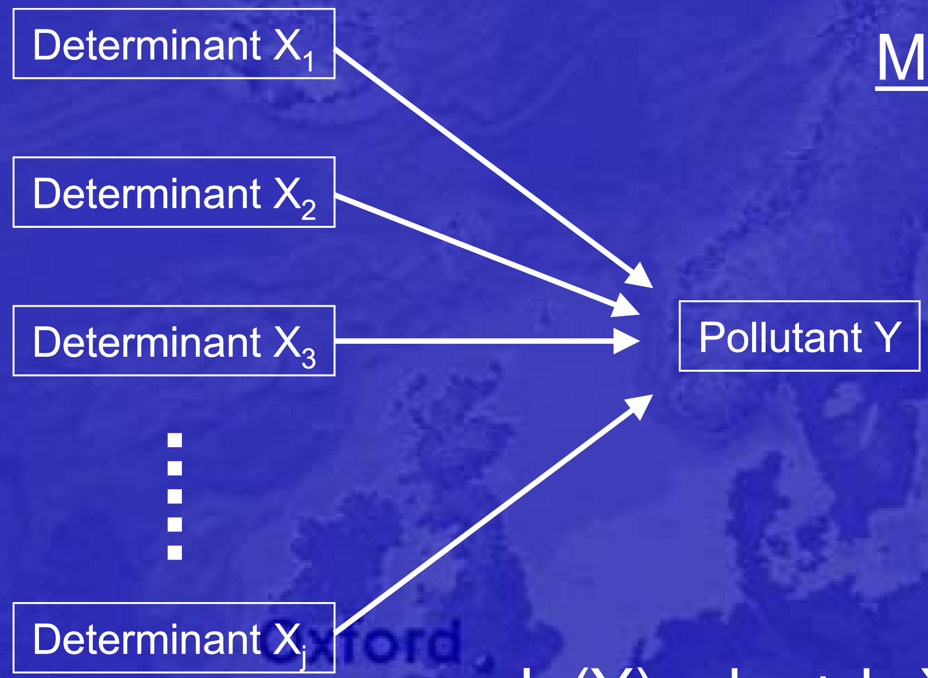
Content

- Objective
- Variables
- Model Results
- Conclusions

Objective

To develop regression models to generalise the determinants of residential indoor levels of $PM_{2.5}$ and VOCs in European cities using the EXPOLIS data from Athens, Basel, Helsinki, Milan, Prague and Oxford.

Multiple regression



$$\ln(Y) = b_0 + b_1X_1 + b_2X_2 + \cdots + b_jX_j + e$$

where,

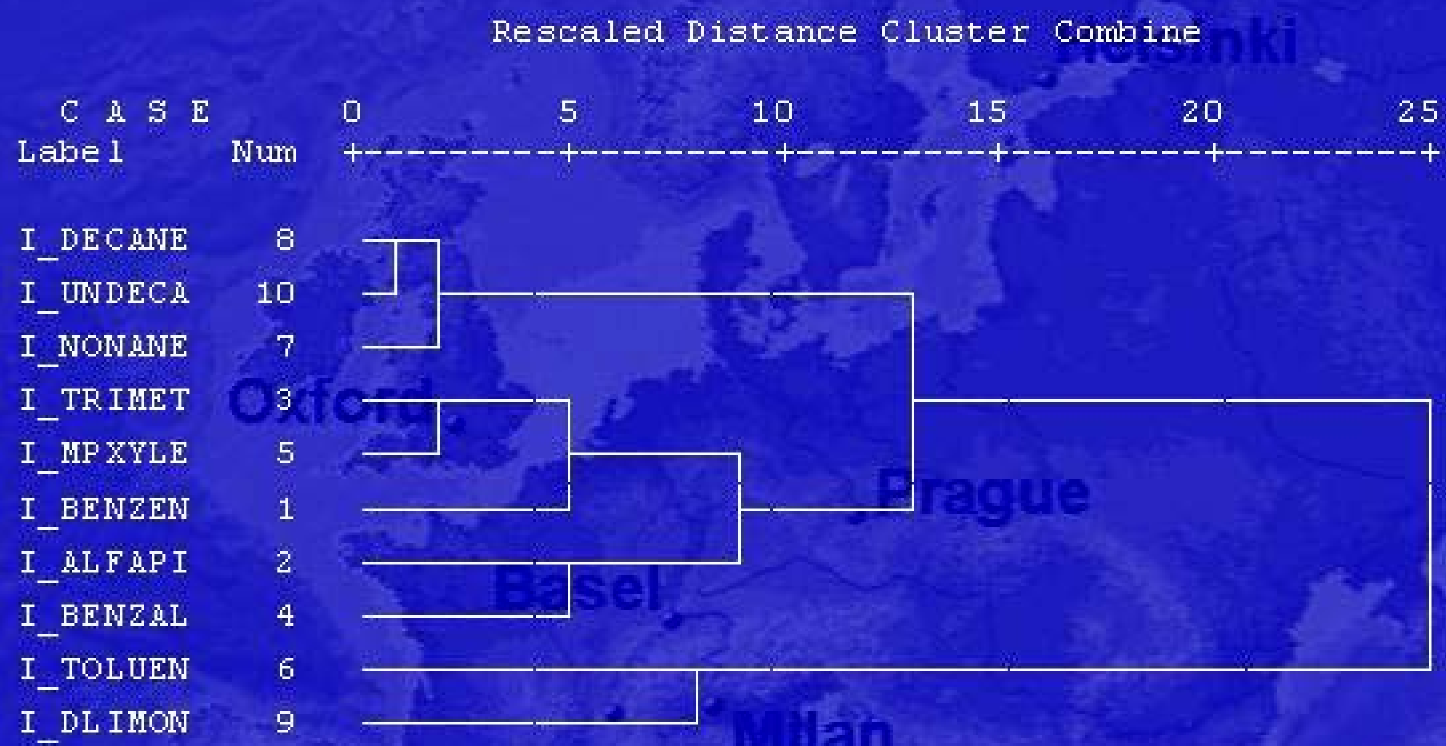
Y = target indoor pollutant
X_j = determinant
b_0 = intercept
b_j = parameter estimates
e = error term

Selected VOCs for modelling - % of n > DL

VOC	Athens	Basel	Helsinki	Milan	Oxford	Prague	
hexane	67%	100%	9%	13%	15%	19%	
cyclohexane	28%	98%	13%	43%	15%	65%	
nonane	95%	100%	73%	93%	70%	92%	OK
decane	91%	98%	73%	85%	68%	67%	OK
undecane	93%	100%	89%	91%	85%	88%	OK
benzene	100%	100%	67%	91%	70%	94%	OK
naphthalene	100%	45%	10%	76%	8%	71%	
trimethylbenzene	100%	100%	55%	96%	43%	92%	OK
ethylbenzene	100%	96%	80%	96%	28%	98%	
styrene	58%	87%	26%	76%	13%	50%	
propylbenzene	86%	91%	36%	83%	38%	88%	
m,p-xylene	100%	100%	95%	93%	63%	98%	OK
o-xylene	98%	100%	47%	93%	13%	83%	
toluene	100%	100%	98%	96%	88%	100%	OK
1-butanol	56%	0%	90%	24%	15%	85%	
2-methyl-1-propanol	28%	0%	68%	61%	3%	62%	
2-ethyl-1-hexanol	30%	66%	47%	17%	15%	60%	
phenol	2%	0%	10%	7%	3%	8%	
1-octanol	0%	0%	2%	0%	3%	0%	
2-butoxy-ethanol	58%	70%	22%	52%	10%	31%	
hexanal	70%	100%	82%	30%	25%	56%	
benzaldehyde	95%	98%	93%	96%	88%	92%	OK
octanal	44%	100%	70%	33%	5%	31%	
1,1,2-trichloroethane	0%	100%	1%	2%	0%	0%	
trichloroethene	35%	72%	2%	70%	5%	15%	
tetrachloroethene	49%	94%	6%	93%	5%	77%	
α-pinene	81%	98%	86%	57%	70%	71%	OK
d-limonene	98%	100%	94%	89%	75%	98%	OK
1-methyl-2-pyrrolidinone	0%	15%	1%	0%	5%	2%	
3-carene	44%	91%	73%	9%	75%	35%	

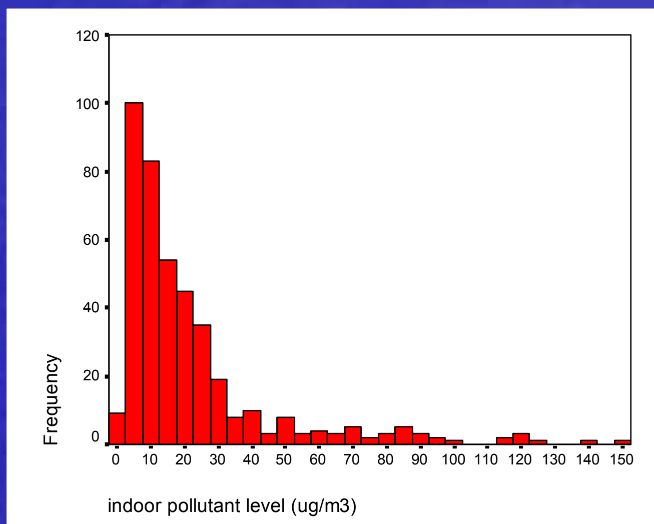
VOCs with >40% of samples above the detection limit are highlighted in red colour. The 10 VOCs selected for modelling are highlighted in all cities

Selected VOCs for today



Cluster analysis was used to select 2 VOCs from all the 10 selected VOCs for this presentation

Outcome Variables (Y)



ln(indoor $PM_{2.5}$)

ln(indoor benzene)

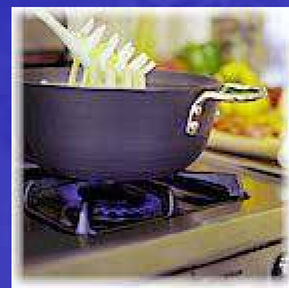
ln(indoor d-limonene)

Outcome variables used in the regression

EXPOLIS

➢Predictor variables - fuel burning

➢Hours using gas-stove for cooking



➢Hours using gas-stove for heating

➢Hours using gas-stove for cooking / heating

EXPOLIS

Different groups of predictor variables used in the regression

➢Predictor variables - traffic factors

➢Nearby heavy vehicle volume (1 never, 2 rare, 3 medium, 4 continuous)

Photo Source: London Transport Museum

➢Nearby traffic volume (1 rare, 2 medium, 3 continuous)

➢Predictor variables - smoking

➢Duration of anybody smoking at home (min)

➢No. of cigarette consumed per day by anybody at home

➢No. of people smoking at home

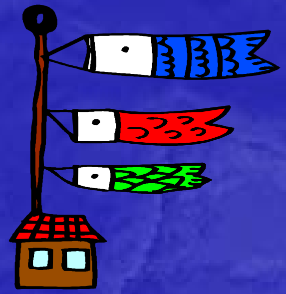
➢Predictor variables - weather

➢indoor temperature (°C)

➢outdoor temperature (°C)

➢cloudiness (oktas 1-8)

➢wind speed (m s^{-1})



Helsinki
Oxford
Prague
Basel
Milan
Athens

EXPOLIS

➢Predictor variables - ventilation

- ➢air-conditioning (hr)

- ➢kitchen extraction fan (hr)
  [type: extract to outside]

- ➢windows opening (hr)

- ➢air-conditioning / kitchen extraction
  fan / windows opening (hr)

➢Predictor variables – household products

➢use of air freshener (y/n)

➢use of chemical cleaning agent (y/n)

➢use of anti-moth product (y/n)

➢Predictor variables – other human activities

➢cooking activity (min)

➢vacuum cleaning (y/n)

➢Predictor variables – demographic info.

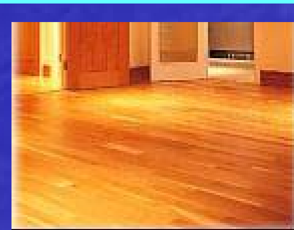
- ➢Total no. of people at home (no.)

- ➢total floor area (m^2)
- ➢No. of people per total floor area (no. / m^2)

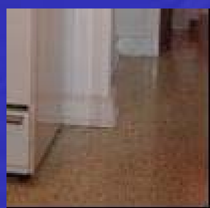
EXPOLIS

➢Predictor variables – interiors (y/n)

➢wooden floor



➢curtain

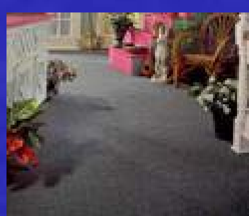


➢linoleum floor

➢chipboard

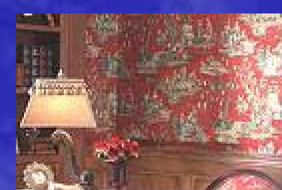
➢PVC floor

➢plaster wall



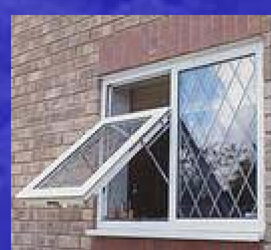
➢wall-to-wall carpet

➢other carpet



➢wall paper

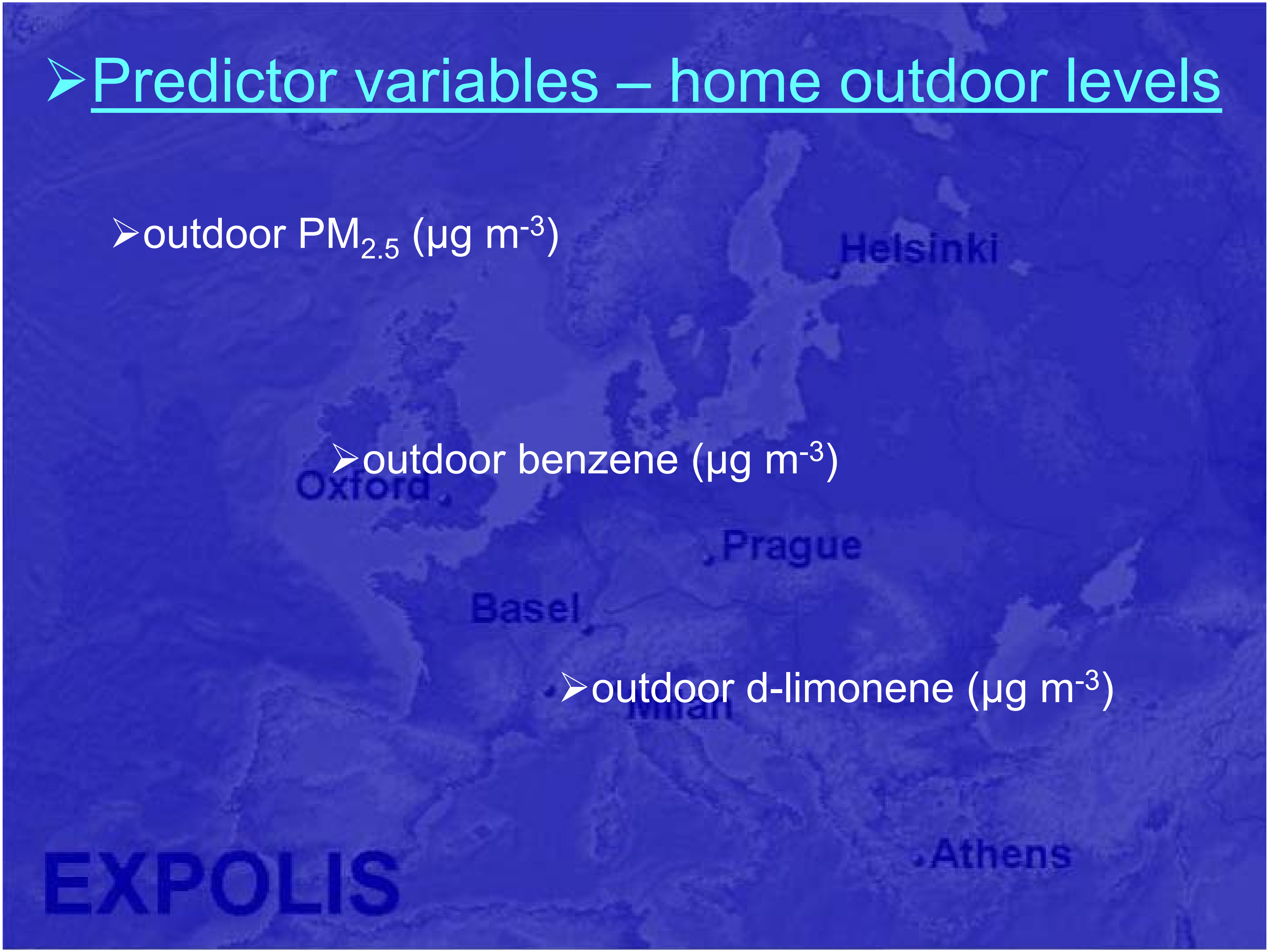
➢Double-glazing



➢wall painting / paper renovation last year

Predictor variables – home outdoor levels

- outdoor $PM_{2.5}$ (µg m^{-3})
- outdoor benzene (µg m^{-3})
- outdoor d-limonene (µg m^{-3})



How?

- >50 predictor variables → multicollinearity
  → how to select?

- Helsinki (n=200) vs other cities (n=50)
  → what can we do about it?

How to manage “trial and error”?

SAS programming

- Analyse 11 target pollutants consistently
- Quick recovery to correcting any mistakes in any stage

Program = Methodology

'Trial and error' was managed by SAS programming. My program can now handle the multicollinearity and the unequal sample size problem, and it is actually the analytical methodology. An overnight run is enough to get the model result for one VOC!

Results: $PM_{2.5}$

Y: ln(indoor $PM_{2.5}$) Adjusted Model R^2: 0.51

X	B	Sig.	95% C.I. for B		Partial	Collinearity Stat	
			Lower	Upper	Corr.	Tolerance	VIF
(Constant)	0.832	0.000	0.375	1.289			
ln(outdoor $PM_{2.5}$) (µg m^{-3})	0.595	0.000	0.467	0.723	0.515	0.761	1.314
no. of smokers at home	0.333	0.000	0.207	0.460	0.321	0.965	1.036
gas cooking /heating (hr)	0.016	0.000	0.007	0.025	0.226	0.920	1.087
outdoor temperature (°C)	0.016	0.016	0.003	0.030	0.157	0.694	1.440
wind speed (m s^{-1})	-0.060	0.014	-0.107	-0.012	-0.160	0.726	1.378
nearby traffic volume (1rare/2med./3cont.)	0.117	0.033	0.010	0.225	0.139	0.935	1.069

ln(indoor $PM_{2.5}$) = 0.832 + 0.595*[ln(outdoor $PM_{2.5}$)] + 0.333*(no. of smoker)
+ 0.016*(gas cooking & heating) + 0.016*(outdoor temp.)
– 0.060*(wind speed) + 0.117*(nearby traffic vol.)

EXPOLIS

Results: benzene

Y: ln(indoor benzene) Adjusted Model R^2: 0.39

X	B	Sig.	95% C.I. for B		Partial	Collinearity Stat	
			Lower	Upper	Corr.	Tolerance	VIF
(Constant)	0.811	0.000	0.395	1.228			
ln(outdoor benzene) ($\mu g\ m^{-3}$)	0.387	0.000	0.279	0.495	0.420	0.808	1.237
no. of smokers at home	0.201	0.014	0.040	0.361	0.159	0.948	1.055
wind speed ($m\ s^{-1}$)	-0.102	0.001	-0.164	-0.041	-0.211	0.895	1.118
ventilation by air-con. /ext. fan /windows op (hr)	-0.010	0.006	-0.017	-0.003	-0.179	0.935	1.070
use of anti-moth product (1 yes / 0 no)	0.389	0.005	0.118	0.660	0.182	0.846	1.183
nearby traffic volume (1rare/2med./3cont.)	0.210	0.006	0.061	0.358	0.179	0.950	1.053

ln(indoor benzene) = 0.811 + 0.387*[ln(outdoor benzene)] + 0.201*(no. of smoker)
– 0.102*(wind speed) – 0.010*(ventilation by ac & ef & wo)
+ 0.389*(use of anti-moth) + 0.117*(nearby traffic vol.)

Homes which needed to use anti-moth products might also have used insecticides or other spray products. The use of anti-moth product is like a surrogate.

Results: d-limonene

Y: ln(indoor d-limonene)   Adjusted Model R^2: 0.18

X	B	Sig.	95% C.I. for B		Partial Corr.	Collinearity Stat	
			Lower	Upper		Tolerance	VIF
(Constant)	2.514	0.000	2.218	2.809			
ventilation by air-con. /ext. fan /windows op (hr)	-0.028	0.000	-0.038	-0.018	-0.350	0.972	1.029
cooking activity (min)	0.003	0.008	0.001	0.005	0.171	0.985	1.015
use of air freshener (1 yes / 0 no)	0.455	0.029	0.048	0.862	0.142	0.956	1.046
use of anti-moth product (1 yes / 0 no)	0.389	0.036	0.026	0.752	0.136	0.952	1.050

ln(indoor d-limonene) = 2.514 – 0.028*(ventilation by ac & ef & wo) + 0.003*(cooking) + 0.455*(use of air freshener) + 0.389*(use of anti-moth)

Duration of cooking activity is also associated because the longer the cooking hour, the more dishes will be needed to clean after the meal, thus associated with an increased amount of detergent or kitchen cleaning agent being used.

Adj Model R^2 = 0.51

$\ln(\text{indoor } PM_{2.5}) = 0.832 + 0.595*[\ln(\text{outdoor } PM_{2.5})] + 0.333*(\text{no. of smoker}) + 0.016*(\text{gas cooking \& heating}) + 0.016*(\text{outdoor temp.}) – 0.060*(\text{wind speed}) + 0.117*(\text{nearby traffic vol.})$

		Example		City average					
		home 1	home 2	Athens	Basel	Helsinki	Milan	Oxford	Prague
X_j	outdoor $PM_{2.5}$ (µg m^{-3})	10	10	36.63	19.39	9.69	41.31	9.42	26.94
X_j	no. of smoker at home (no.)	0	2	0.52	0.37	0.13	0.78	0.37	0.38
X_j	gas cooking / heating (hr) [for 2 days]	0	0	0.29	1.79	0.05	1.74	1.74	11.97
X_j	outdoor Temp (ºC)	15	15	14.97	12.32	6.03	14.17	8.83	9.56
X_j	wind speed (m s^{-1})	2	2	3.29	1.77	5.16	0.93	3.13	3.73
X_j	nearby traffic vol (1rare, 2med., 3 cont.)	1	1	1.66	1.66	1.71	2.02	1.65	2.70
Y →	**Estimated indoor $PM_{2.5}$ (µg m^{-3})**	**11.46**	**22.31**	**29.63**	**20.76**	**9.15**	**42.14**	**11.76**	**28.64**
	Calculated GM of indoor $PM_{2.5}$ (µg m^{-3})	n/a	n/a	27.53	18.54	8.54	31.50	11.50	28.14

This slide shows 2 examples of fitting the predictors to estimate the indoor levels. The average values for each predictor are also fitted for a simple validation with the measured levels (geomeric mean) for each city. Model predicts quite well !

Adj Model R^2 = 0.39

$$\ln(\text{indoor benzene}) = 0.811 + 0.387 \cdot [\ln(\text{outdoor benzene})] + 0.201 \cdot (\text{no. of smoker}) - 0.102 \cdot (\text{wind speed}) - 0.010 \cdot (\text{ventilation by ac \& ef \& wo}) + 0.389 \cdot (\text{use of anti-moth}) + 0.117 \cdot (\text{nearby traffic vol.})$$

	Example		City average					
	home 1	home 2	Athens	Basel	Helsinki	Milan	Oxford	Prague
outdoor benzene ($\mu g\ m^{-3}$)	4	4	10.08	1.40	1.73	28.81	4.67	4.57
no. of smoker at home (no.)	0	1	0.52	0.37	0.13	0.78	0.37	0.38
wind speed ($m\ s^{-1}$)	2	2	3.29	1.77	5.16	0.93	3.13	3.73
ventilation by air-con. /ext. fan /windows op (hr) [for 2 days]	15	15	12.57	20.49	17.97	10.37	13.90	11.52
use of anti-moth product (1 yes / 2 no)	0	1	0.80	0.26	0.04	0.53	0.10	0.18
nearby traffic vol (1rare, 2med., 3 cont.)	1	1	1.66	1.66	1.71	2.02	1.65	2.70
Estimated indoor benzene ($\mu g\ m^{-3}$)	**3.33**	**6.01**	**7.45**	**2.94**	**2.05**	**14.89**	**4.09**	**5.04**
Calculated GM of indoor benzene ($\mu g\ m^{-3}$)	n/a	n/a	8.02	2.18	1.59	10.71	2.64	6.37

EXPOLIS

Model diagnostics

No heteroscedasticity	White's test: p-value>0.05 Residual plot appears 'rectangular' shape
No perfect multicollinearity	VIF close to 1 Tolerance values high above 0.2
Independent errors	Durbin-Watson statistics close to 2
Normally distributed errors	Kolmogorov-Smirnov's test
Linearity	Plots of y versus x_i all lie along straight lines
No outliers	Maximum Cook's distance < 1 Maximum Leverage value < 1 Minimum Leverage value > 0

All models in this presentation fulfil all important assumptions for multiple regression

Conclusion

- 6 determinants of residential indoor $PM_{2.5}$ level:
  outdoor $PM_{2.5}$ level, no. of smokers at home, gas cooking/heating, outdoor temp., wind speed, nearby traffic volume

- 6 determinants of residential indoor benzene:
  outdoor benzene level, no. of smokers at home, wind speed, ventilation duration, use of anti-moth product, nearby traffic volume

- 4 determinants of residential indoor d-limonene:
  ventilation duration, cooking duration, use of air freshener, use of anti-moth product

Hak Kan Lai[1]; Mark J. Nieuwenhuijsen[1];
Christian Schweizer[2,3]; Rufus Edwards[4]; Lucy Bayer-Oglesby[2,3];
Nino Künzli[2,3]; Klea Katsouyanni[6]; Macro Maroni[6]; Radim Srám[6],
Matti Jantunen[5]

(1) Dept of Environmental Science and Technology, Imperial College, London, UK
(2) University of Southern California Los Angeles, US
(3) Institute of Social and Preventive Medicine, University of Basel, Switzerland
(4) Environmental Health, Science and Policy, University of California at Irvine, US
(5) KTL Finland
(6) EXPOLIS INDEX group

Special thanks to: Dr. Mark Nieuwenhuijsen[1] and Dr. Roy Colvile[1]

Imperial College, University of London http://www.env.ic.ac.uk/