UNIVERSITY OF EXETER

From satellites to global burden

(the road to DIMAQ too!)

Gavin Shaddick

Inferential Challenges for
Large Spatio-Temporal Data

Inter-disciplinary collaboration

Matthew Thomas, Amelia Green (Bath), Dan Simpson (Toronto)

WHO Data Integration Task Force

Michael Brauer (UBC)
Rick Burnett (Health Canada)
Howard Chang (Emory)
Aaron Cohen (HEI)
Rita Van Dingenen (JRC)
Aaron van Donkelaar (Dalhousie)
Yang Liu (Emory)
Randall Martin (Dalhousie)
Annette Pruss-Ustun (WHO)
Gavin Shaddick (Exeter)
Lance Waller (Emory)
Jason West (North Carolina)
Jim Zidek (UBC)

Inter-disciplinary collaboration

Matthew Thomas, Amelia Green (Bath) Dan Simpson (Toronto)

WHO Data Integration Task Force

Michael Brauer (UBC)
Rick Burnett (Health Canada)
Howard Chang (Emory)
Aaron Cohen (HEI)
Rita Van Dingenen (JRC)
Aaron van Donkelaar (Dalhousie)
Yang Liu (Emory)

Randall Martin (Dalhousie)
Annette Pruss-Ustun (WHO)
Gavin Shaddick (Exeter)
Lance Waller (Emory)
Jason West (North Carolina)
Jim Zidek (UBC)

Outline

- Air pollution and global health
- Data integration in global burden of disease
- The way to DIMAQ too!
- Black boxes
- So much more to do…

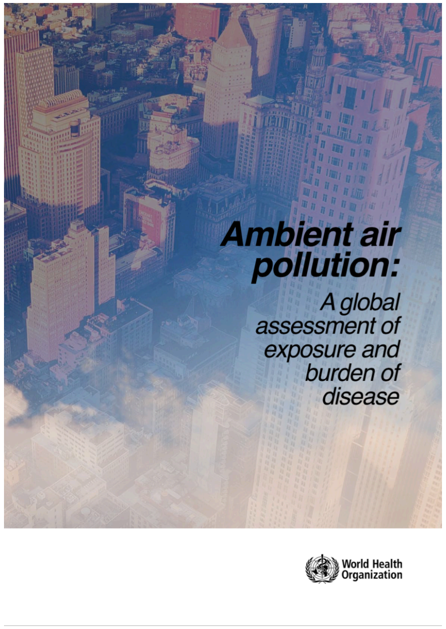
Air pollution and global health

A global health priority

- Air pollution has been identified as a global health priority in the sustainable development agenda.
- Sustainable Development Goals (SDGs):
  - Health (Goal 3);
  - Cities (Goal 11);
  - Energy (Goal 7).
- SDG Indicators:
  - 11.6.2: Annual mean levels of fine particulate matter (PM2.5) (population-weighted);
  - 3.9.1: Mortality rate attributed to household and ambient air pollution.

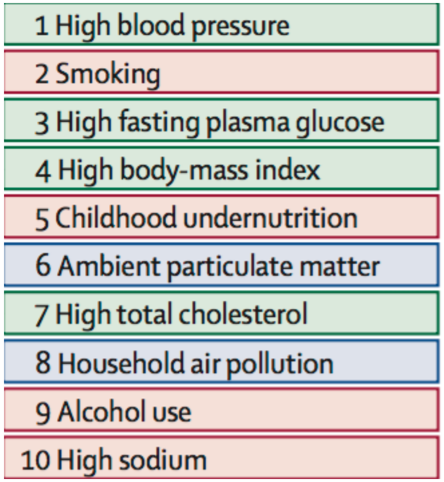
The global burden

- In 2016, the WHO estimated that over 3 million deaths can be attributed to ambient (outdoor) air pollution (AAP).
- The Global Burden of Disease project (Institute of Health Metric Evaluation) estimate that in 2015 AAP was in the top ten leading risks to global health
- Burden of disease calculations require information on population exposures for each country



Ambient air pollution:
A global assessment of exposure and burden of disease

World Health Organization



1 High blood pressure
2 Smoking
3 High fasting plasma glucose
4 High body-mass index
5 Childhood undernutrition
6 Ambient particulate matter
7 High total cholesterol
8 Household air pollution
9 Alcohol use
10 High sodium

Attributable burden

- Population attributable fraction (PAF), for each country

$$PAF = \frac{\sum_{i=1}^{n} P_i(RR - 1)}{\sum_{i=1}^{n} P_i(RR - 1) + 1}$$

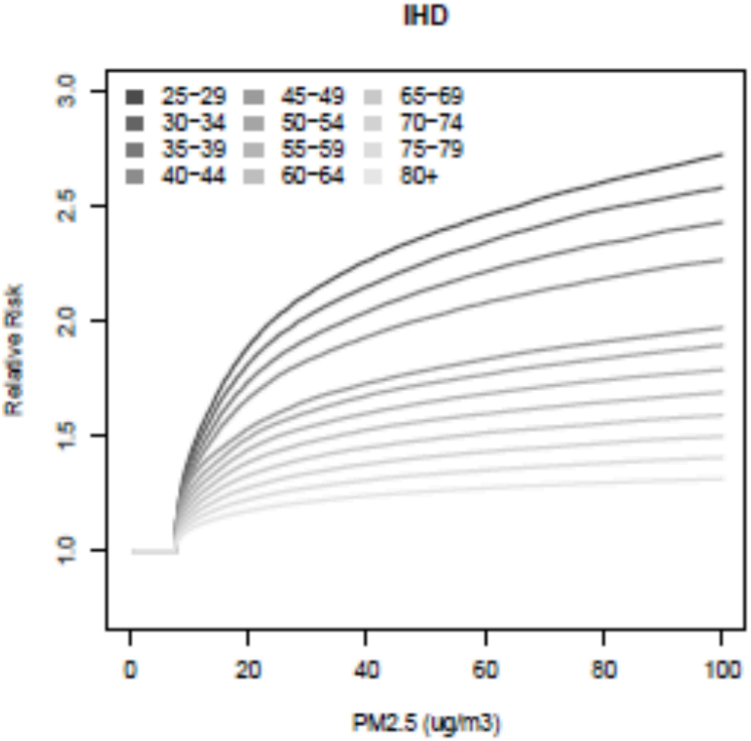
- Attributable burden (AB)

AB = PAF x health outcome

- This requires the percentage of the population, P_i, exposed to PM2.5, by country
  - increments of 1 µg/m^3

The PAF

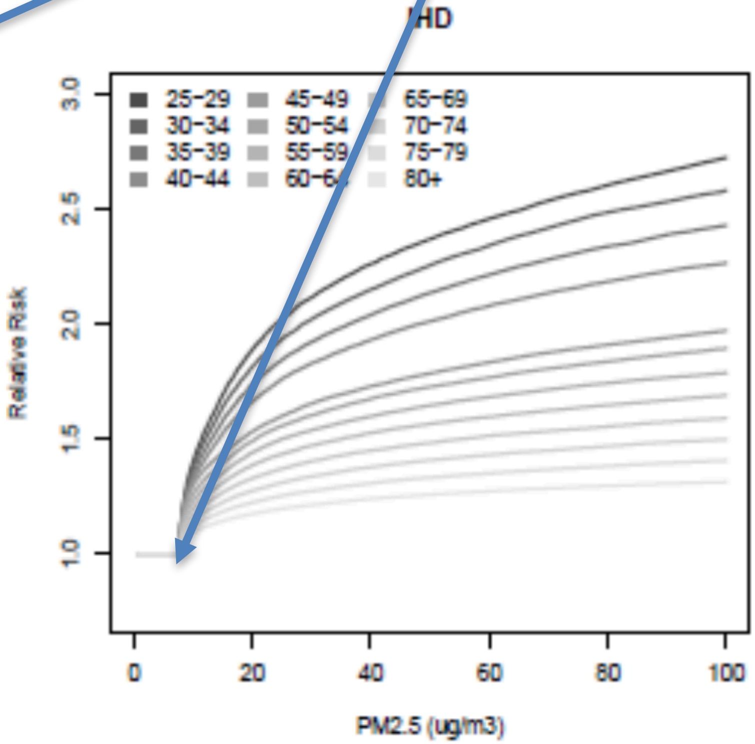
PM2.5	RR	RR-1	P
0	1	0	0
1	1	0	0
2	1	0	0
3	1	0	100
4	1	0	980
5	1	0	54567
6	1.02	0.02	34523
7	1.02	0.02	87645
8	1.03	0.03	99876
9	1.04	0.04	123876
10	1.05	0.05	546987
11	1.06	0.06	846599
12	1.08	0.08	#######
13	1.08	0.08	#######
.	.	.	.
.	.	.	.
.	.	.	.



The PAF

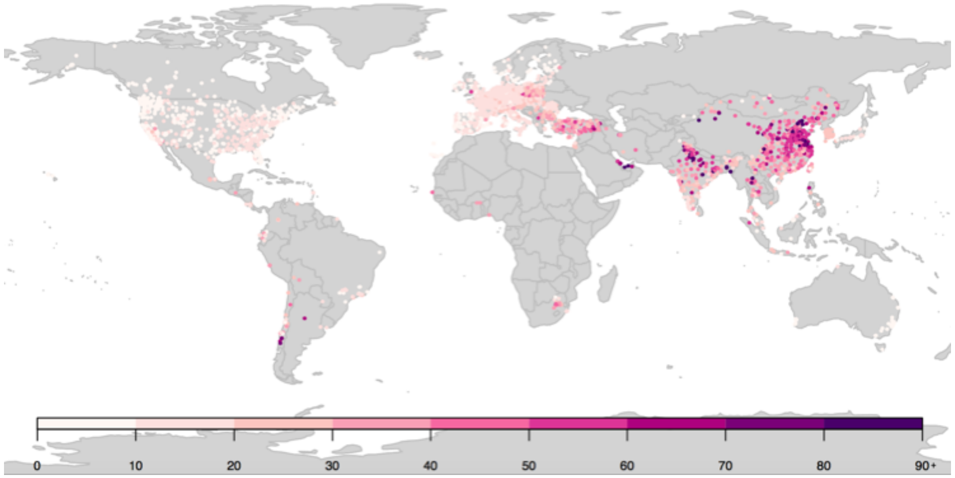
counterfactual

PM2.5	RR	RR-1	P
0	1	0	0
1	1	0	0
2	1	0	0
3	1	0	100
4	1	0	980
5	1	0	54567
6	1.02	0.02	34523
7	1.02	0.02	87645
8	1.03	0.03	99876
9	1.04	0.04	123876
10	1.05	0.05	546987
11	1.06	0.06	846599
12	1.08	0.08	12547895
13	1.08	0.08	54345670
.	.	.	.
.	.	.	.
.	.	.	.



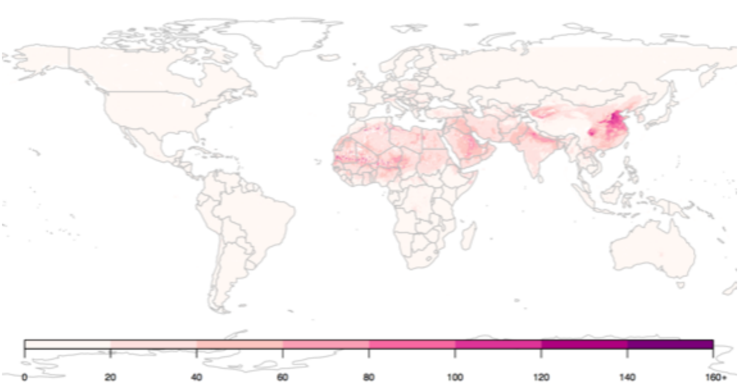
Estimating PM2.5

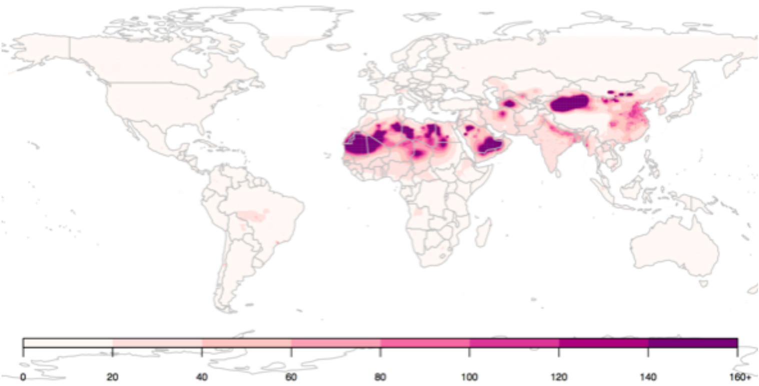
- There is a need for accurate estimates of exposure to air pollution: at global, national and local levels
  - Measures of uncertainty
- Ground monitoring (GM) is limited in many areas of the world



Data from multiple sources

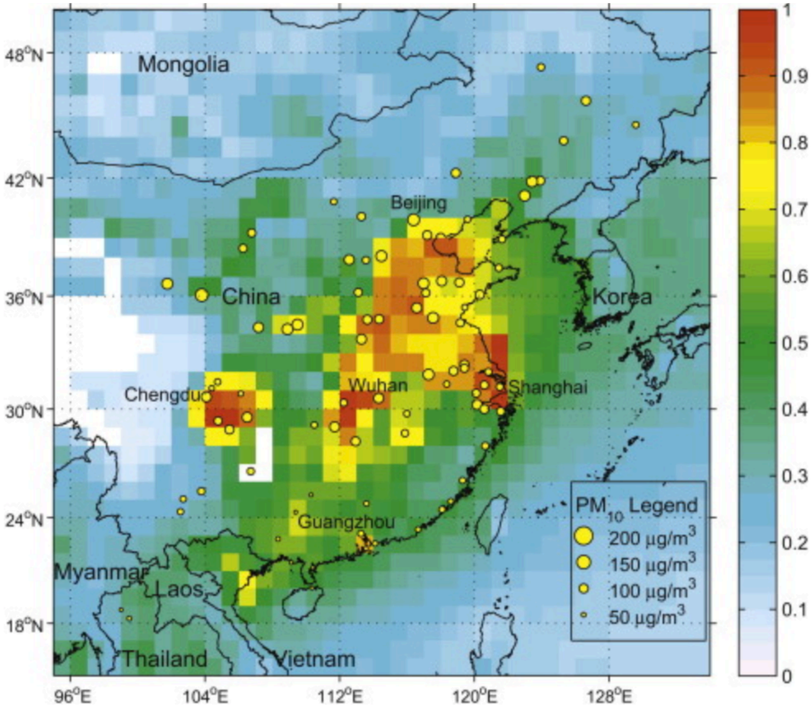
- Can utilise information from other sources
  - satellite remote sensing (SAT)
  - atmospheric/ chemical transport models (CTM)
  - population estimates
  - local network characteristics





What data do we have?

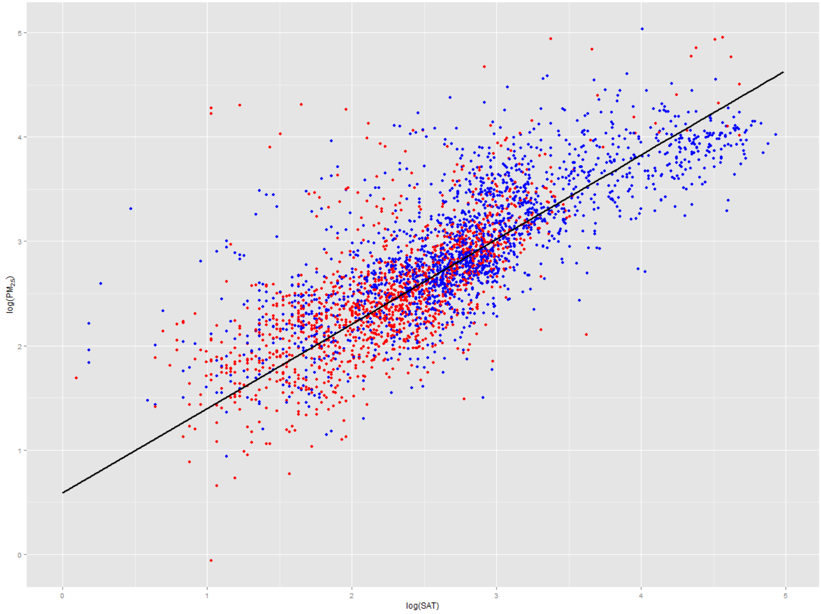
- Multiple sources
  - National, regional, global
- Multiple measures
  - fundamentally different quantities
- Multiple scales
  - point locations, grid cells
  - hourly, daily, annual averages
- Different error structures and uncertainties
  - Vary over space and time



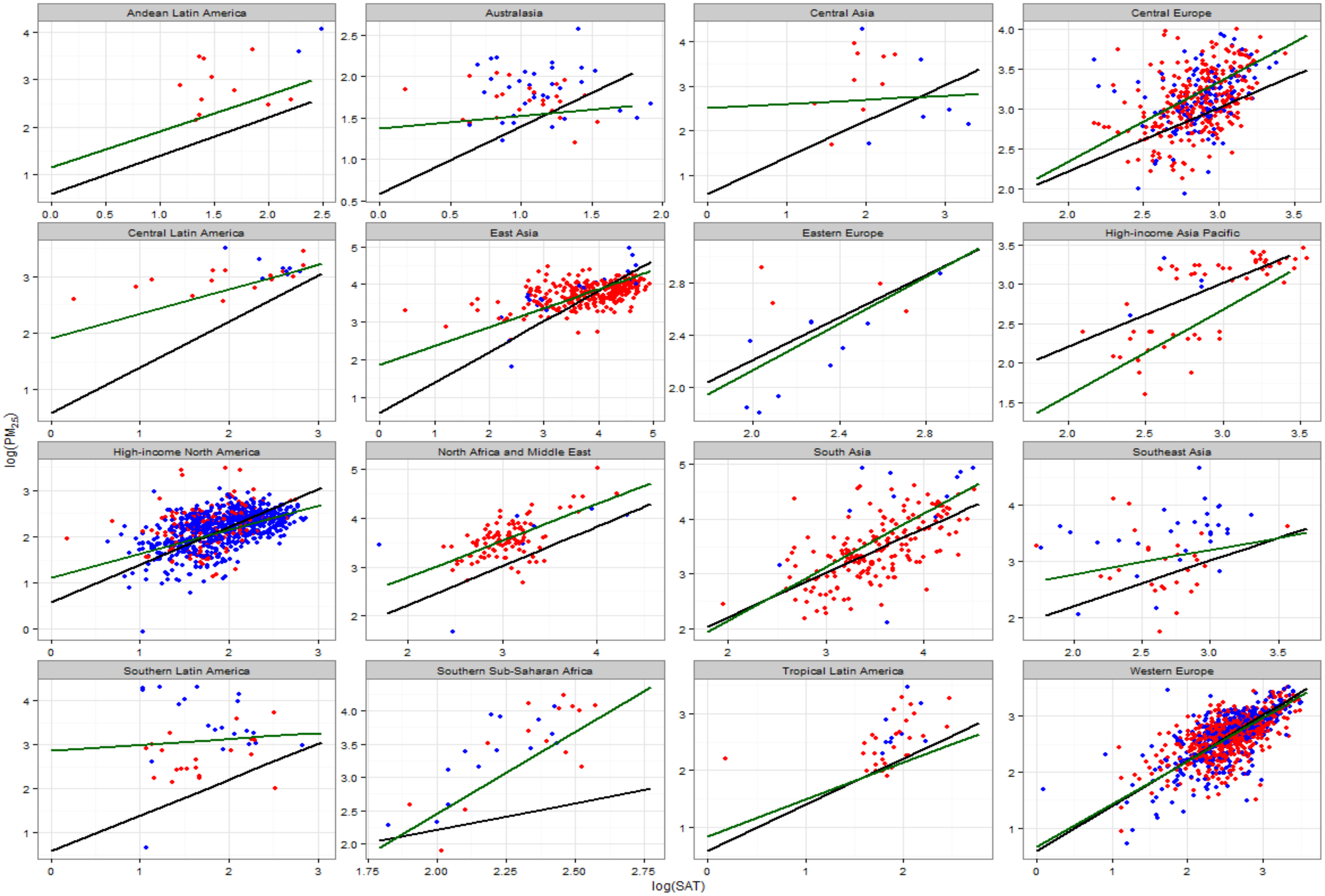
Data integration in global burden of disease

Data integration in GBD 2013

- Combined estimates from remote sensing satellites and a chemical transport model
  - 0.1^0 grid cells
- Single relationship. between ground measurements, SAT and CTM for all areas of the world.



Regional variation



DIMAQ (GBD2015, 2016 and WHO2016)

- Calibration of GMs with SAT, CTM and other factors
  - Relationships allowed to vary by country
  - Where GM information is sparse, information can be ‘borrowed’
  - Country, region, super-region, spatial dependence
- Summaries of predictions and uncertainty can be mapped
  - 0.1^{o} resolution
  - globally, by country, within country
- Accuracy and uncertainty will vary according to local information available from ground monitoring

The DIMAQ model

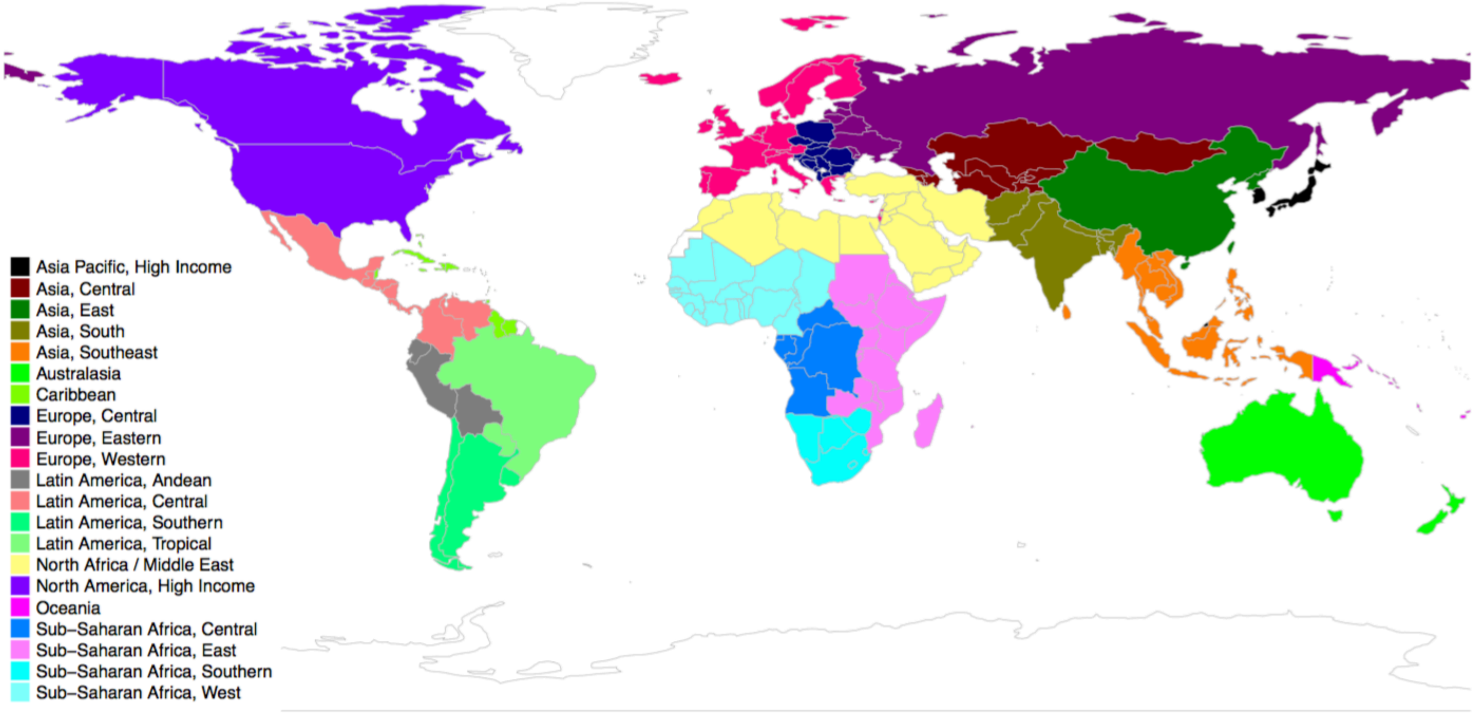
$$\begin{aligned}\log(Y_{slijk}) = \tilde{\beta}_{0,lijk} \quad & + \quad \sum_{q \in Q} \tilde{\beta}_{q,ijk} X_{q,lijk} \\ & + \quad \sum_{p_1 \in P_1} \beta_{p_1} X_{p_1,lijk} + \sum_{p_2 \in P_2} \beta_{p_2} X_{p_2,slijk} \\ & + \quad \epsilon_{slijk} \ ,\end{aligned}$$

- The random effect terms have contributions from the country, the region and the super–region
- The intercept also having a random effect for the cell representing within-cell variation in ground measurements
- R-INLA

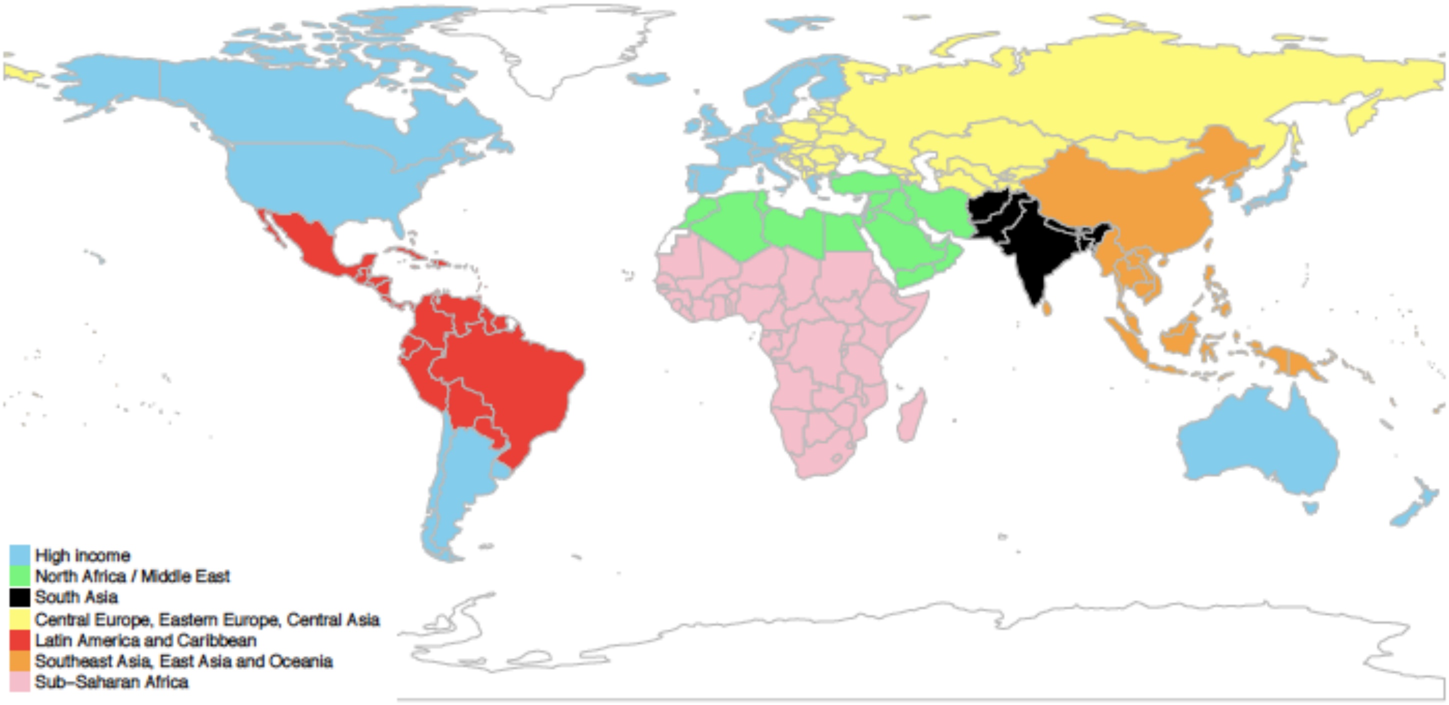
A geographical hierarchy

- The structure of the random effects used here exploits a geographical nested hierarchy
    - each of the 187 countries considered are allocated to one of 21 regions and, further, to one of 7 super-regions.
- Where there are limited monitoring data within a country, information can be borrowed from higher up the hierarchy
    - i.e. from other countries within the region and further, from the wider super-region.

Countries within regions…



… within super-regions


🟦 High income
🟩 North Africa / Middle East
⬛ South Asia
🟨 Central Europe, Eastern Europe, Central Asia
🟥 Latin America and Caribbean
🟧 Southeast Asia, East Asia and Oceania
▨ Sub−Saharan Africa

Random effects structure

- The coefficients for super-regions are distributed with mean equal to the overall mean (β_0, the fixed effect) and variance, σ^2 , representing between super-region

$$\beta_k^{SR} \sim N(\beta_0, \sigma_{SR}^2)$$

- The coefficient for region j (in super–region k) that will be distributed with mean equal to to the coefficient for the super-region and variance representing the between region (within super–region) variability

$$\beta_{jk}^{R} \sim N(\beta_k^{SR}, \sigma_{R,k}^2)$$

- The country level effect will be distributed with mean equal to the coefficient for region j within super-region k with variance representing the between country (within region) variability

$$\beta_{ijk}^{C} \sim N(\beta_{jk}^{R}, \sigma_{C,jk}^2)$$

$$\beta_i^C | \beta_{i'}^C, \; i' \in \partial_i \sim N\left(\overline{\beta}_i^C, \frac{\psi^2}{N_{\partial i}}\right)$$

Evaluation

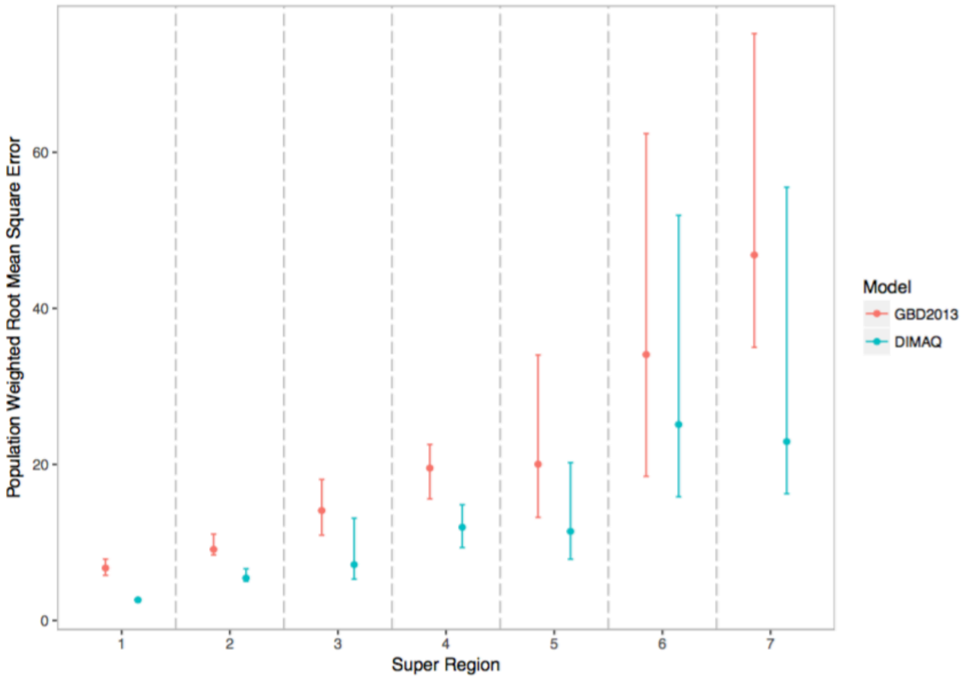


Figure: Summaries of predictive ability of the GBD2013 model and DIMAQ, for each of seven super–regions: 1, High income; 2, Central Europe, Eastern Europe, Central Asia; 3, Latin America and Caribbean; 4, Southeast Asia, East Asia and Oceania; 5, North Africa / Middle East; 6, Sub-Saharan Africa; 7, South Asia. For each model, population weighted root mean squared errors (μgm^{-3}) are given with dots denoting the median of the distribution from 25 training/evaluation sets and the vertical lines the range of values.

Global predictions of PM2.5

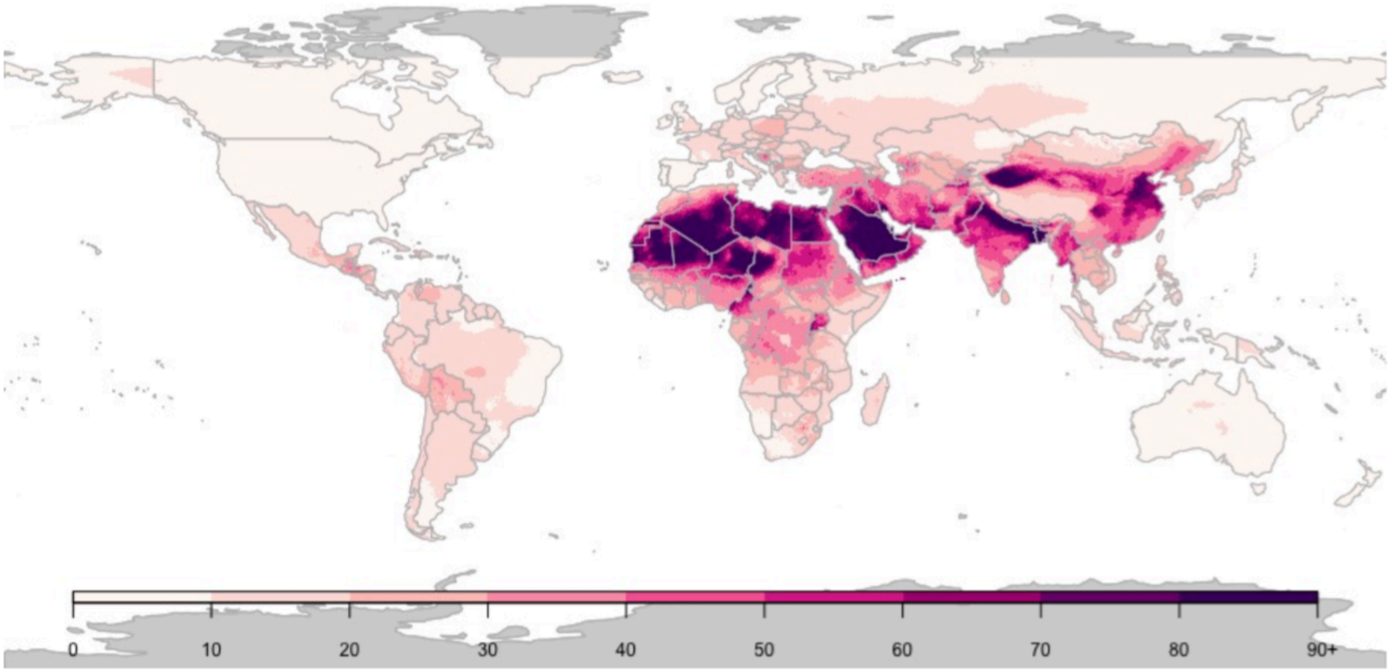
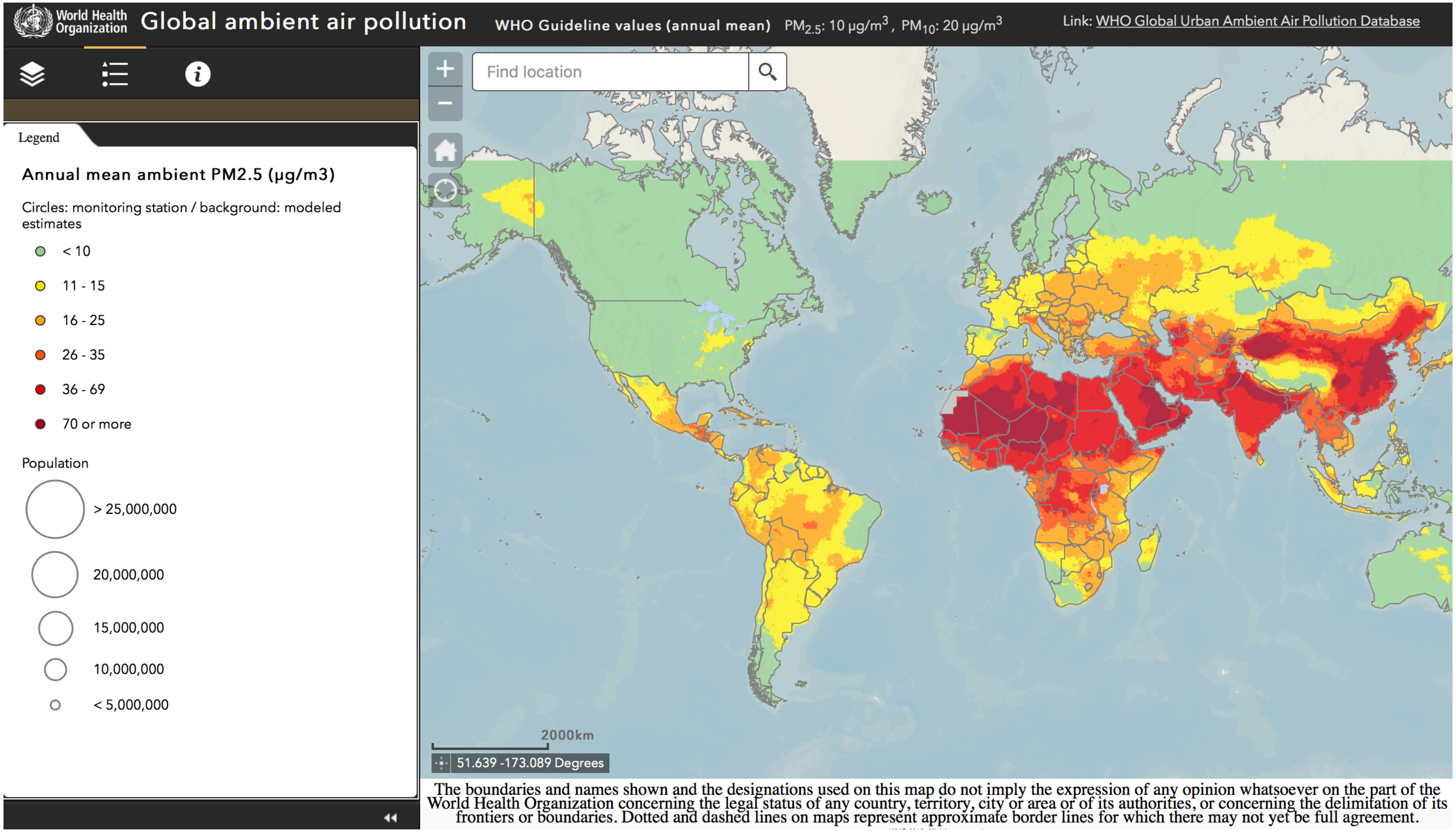


Figure: Median estimates of annual averages of $PM_{2.5}$ (μgm^{-3}) for 2014 for each grid cell ($0.1^o \times 0.1^o$ resolution) using DIMAQ.

Interactive map



World Health Organization

Global ambient air pollution

WHO Guideline values (annual mean) $PM_{2.5}$: 10 µg/m^3 , PM_{10}: 20 µg/m^3

Link: WHO Global Urban Ambient Air Pollution Database

Annual mean ambient PM2.5 (µg/m3)

Circles: monitoring station / background: modeled estimates

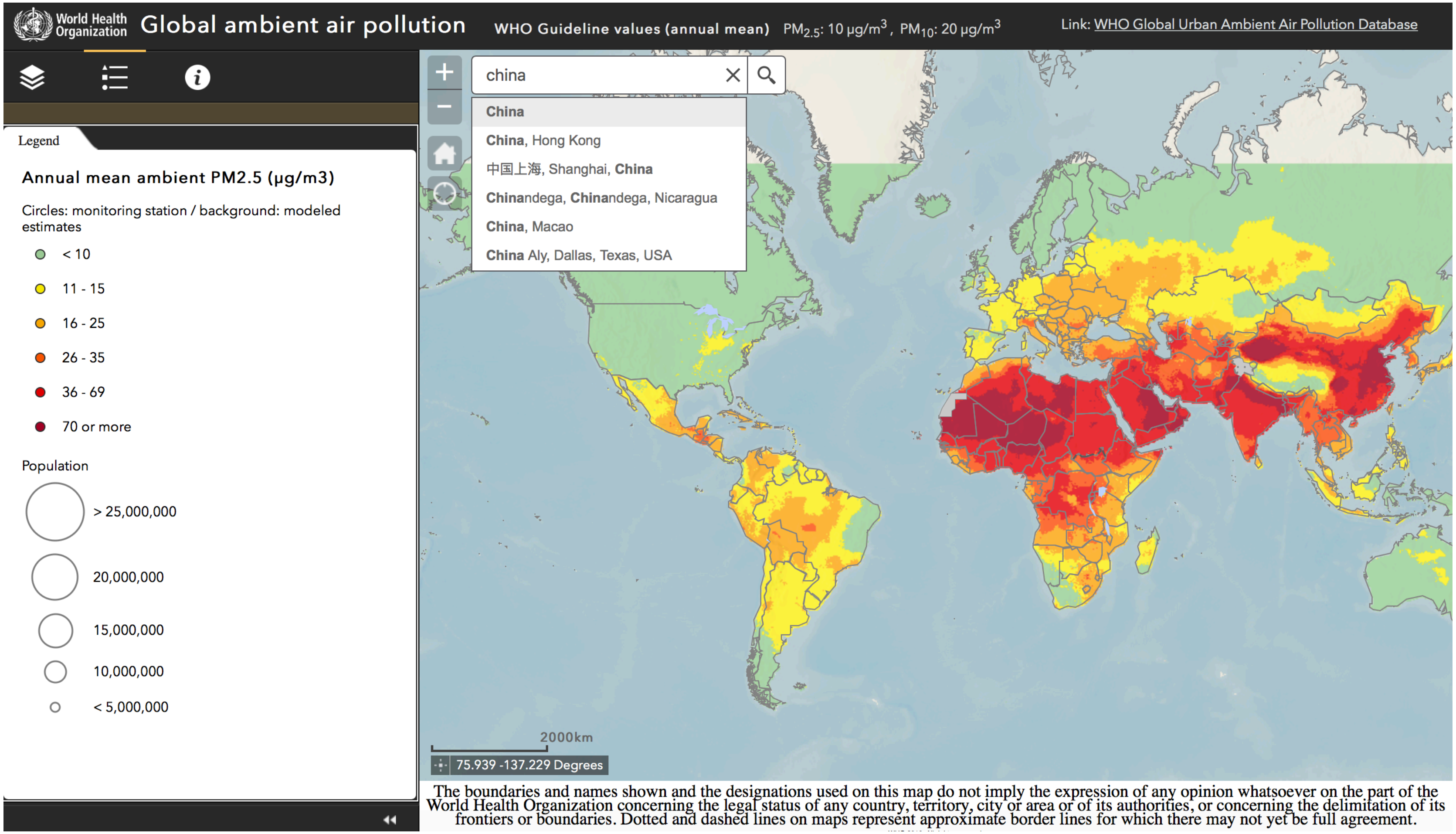
- ● < 10
- ● 11 - 15
- ● 16 - 25
- ● 26 - 35
- ● 36 - 69
- ● 70 or more

Population

- ○ > 25,000,000
- ○ 20,000,000
- ○ 15,000,000
- ○ 10,000,000
- ○ < 5,000,000

The boundaries and names shown and the designations used on this map do not imply the expression of any opinion whatsoever on the part of the World Health Organization concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. Dotted and dashed lines on maps represent approximate border lines for which there may not yet be full agreement.

Interactive map



World Health Organization Global ambient air pollution

WHO Guideline values (annual mean) PM$_{2.5}$: 10 µg/m^3 , PM$_{10}$: 20 µg/m^3

Link: WHO Global Urban Ambient Air Pollution Database

china

- **China**
- **China**, Hong Kong
- 中国上海, Shanghai, **China**
- **Chin**andega, **Chin**andega, Nicaragua
- **China**, Macao
- **China** Aly, Dallas, Texas, USA

Legend

Annual mean ambient PM2.5 (µg/m3)

Circles: monitoring station / background: modeled estimates

- 🟢 < 10
- 🟡 11 - 15
- 🟠 16 - 25
- 🟠 26 - 35
- 🔴 36 - 69
- 🔴 70 or more

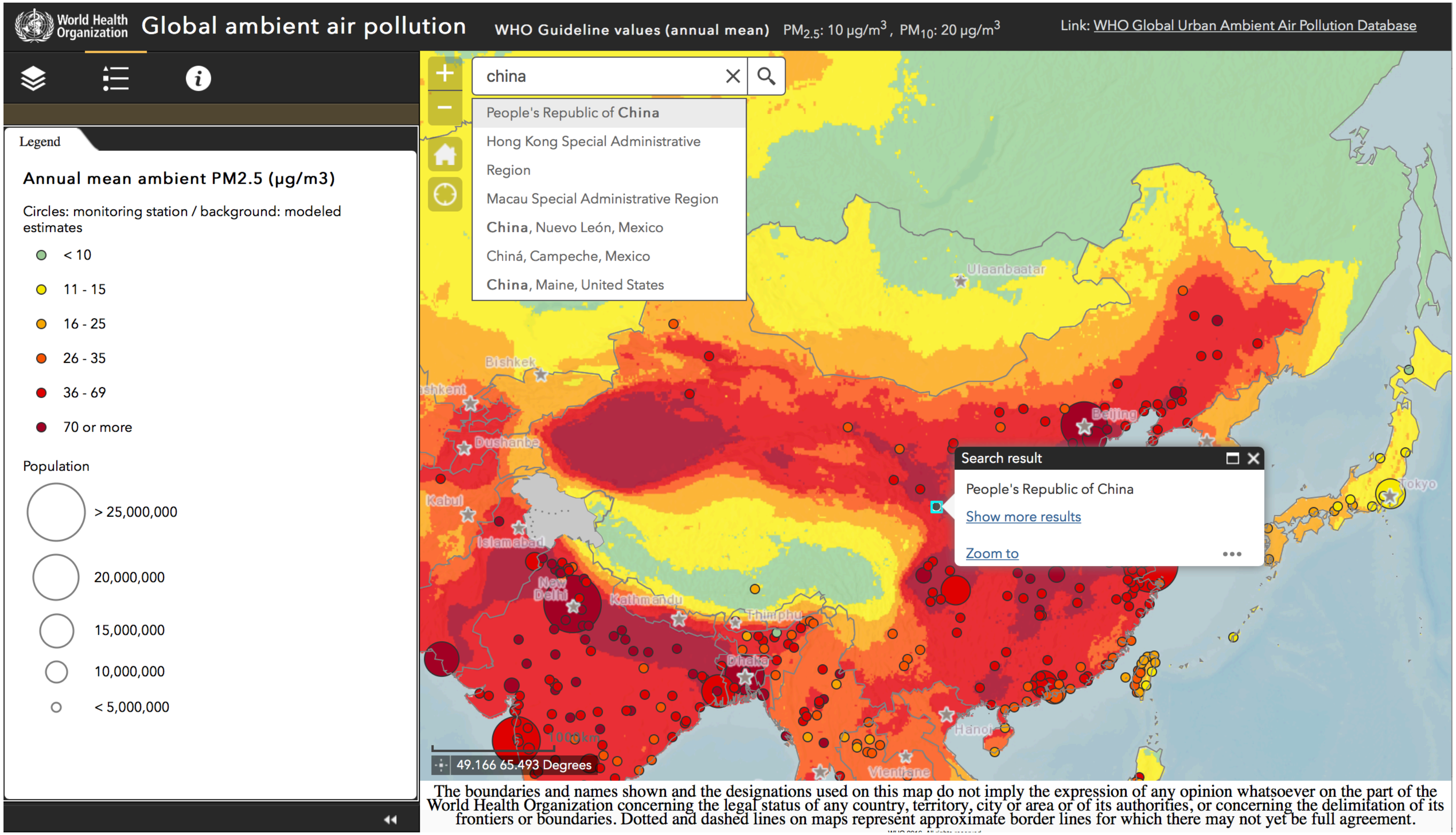
Population

- ○ > 25,000,000
- ○ 20,000,000
- ○ 15,000,000
- ○ 10,000,000
- ○ < 5,000,000

75.939 -137.229 Degrees

The boundaries and names shown and the designations used on this map do not imply the expression of any opinion whatsoever on the part of the World Health Organization concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. Dotted and dashed lines on maps represent approximate border lines for which there may not yet be full agreement.

Interactive map



World Health Organization Global ambient air pollution WHO Guideline values (annual mean) $PM_{2.5}$: 10 µg/m^3, PM_{10}: 20 µg/m^3

Link: WHO Global Urban Ambient Air Pollution Database

Legend

Annual mean ambient PM2.5 (µg/m3)

Circles: monitoring station / background: modeled estimates

- ● < 10
- ● 11 - 15
- ● 16 - 25
- ● 26 - 35
- ● 36 - 69
- ● 70 or more

Population

- ○ > 25,000,000
- ○ 20,000,000
- ○ 15,000,000
- ○ 10,000,000
- ○ < 5,000,000

china

- People's Republic of China
- Hong Kong Special Administrative Region
- Macau Special Administrative Region
- China, Nuevo León, Mexico
- Chiná, Campeche, Mexico
- China, Maine, United States

Search result

People's Republic of China

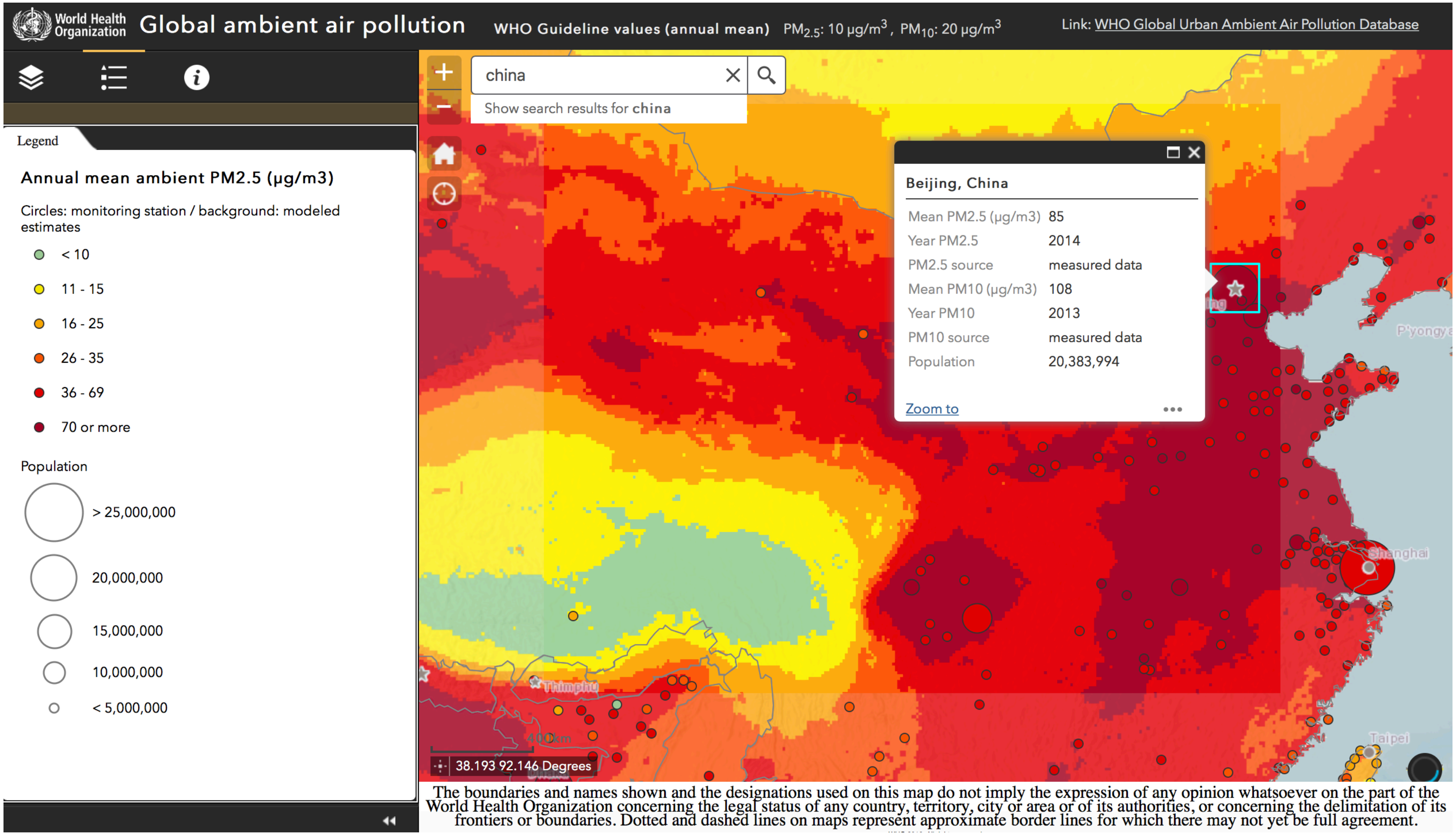
Show more results

Zoom to

49.166 65.493 Degrees

The boundaries and names shown and the designations used on this map do not imply the expression of any opinion whatsoever on the part of the World Health Organization concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. Dotted and dashed lines on maps represent approximate border lines for which there may not yet be full agreement.

Interactive map



World Health Organization **Global ambient air pollution** **WHO Guideline values (annual mean)** PM$_{2.5}$: 10 μg/m^3 , PM$_{10}$: 20 μg/m^3

Link: WHO Global Urban Ambient Air Pollution Database

Legend

Annual mean ambient PM2.5 (μg/m3)

Circles: monitoring station / background: modeled estimates

●	< 10
●	11 - 15
●	16 - 25
●	26 - 35
●	36 - 69
●	70 or more

Population

○	> 25,000,000
○	20,000,000
○	15,000,000
○	10,000,000
○	< 5,000,000

china

Show search results for **china**

Beijing, China

Mean PM2.5 (μg/m3)	85
Year PM2.5	2014
PM2.5 source	measured data
Mean PM10 (μg/m3)	108
Year PM10	2013
PM10 source	measured data
Population	20,383,994

Zoom to

38.193 92.146 Degrees

The boundaries and names shown and the designations used on this map do not imply the expression of any opinion whatsoever on the part of the World Health Organization concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. Dotted and dashed lines on maps represent approximate border lines for which there may not yet be full agreement.

Uncertainty

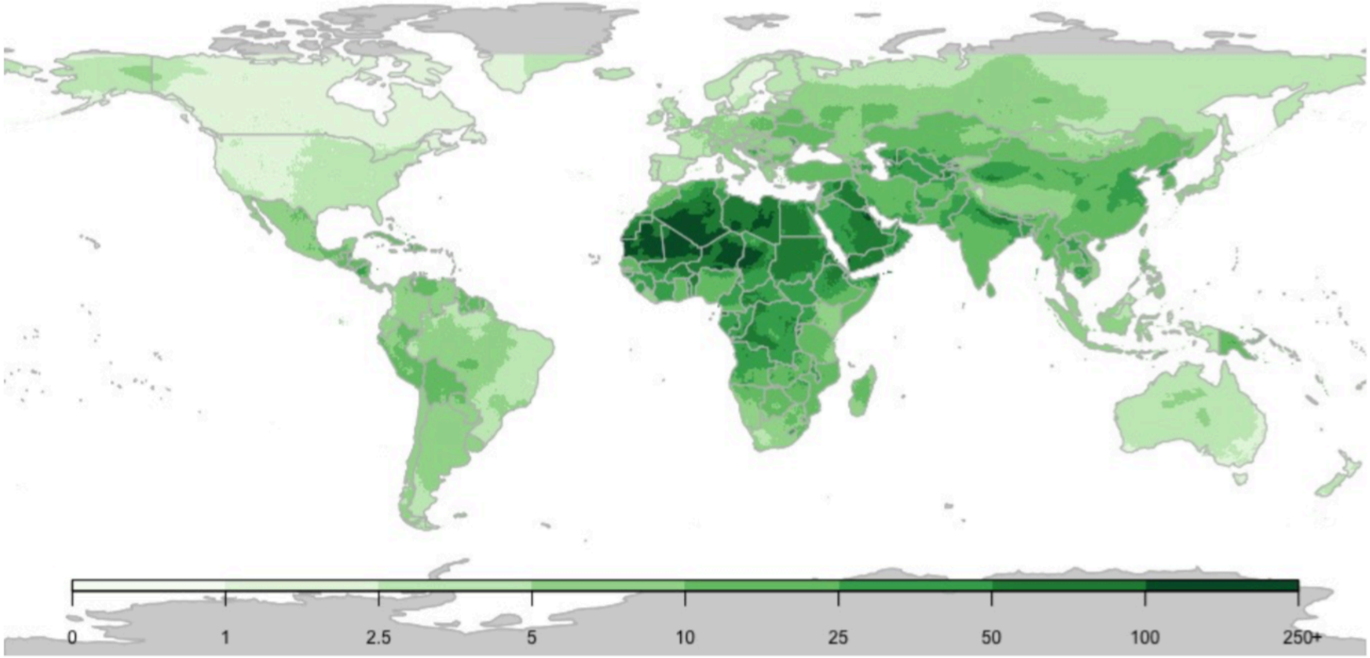


Figure: Half the width of 95% posterior credible intervals for 2014 for each grid cell ($0.1^o \times 0.1^o$ resolution) using DIMAQ.

Posterior distributions

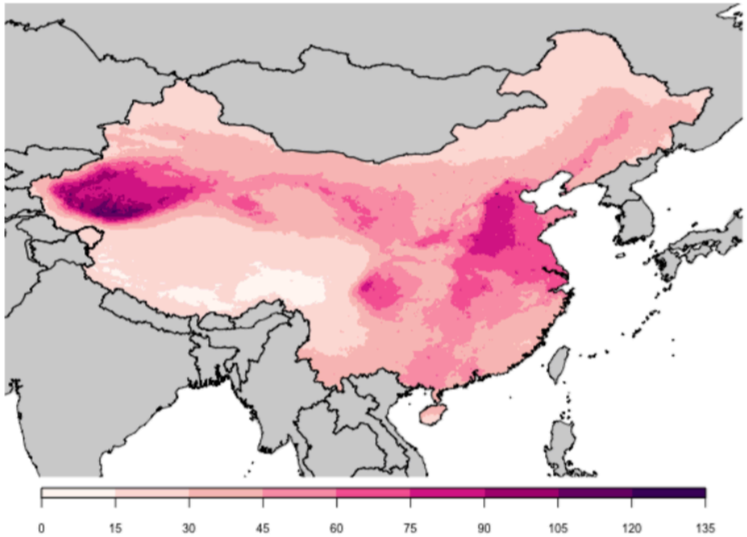


Figure: Medians of posterior distributions for estimates of annual mean $PM_{2.5}$ concentrations (μgm^{-3}) for 2014, in China.

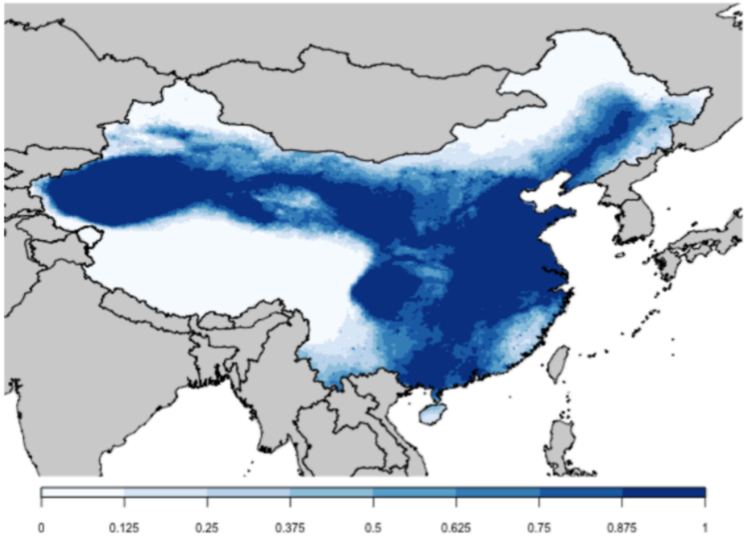


Figure: Probability of exceeding 35 μgm^{-3} using a Bayesian hierarchical model for each grid cell ($0.1^o \times 0.1^o$ resolution) for 2014, in China.

Population exposures

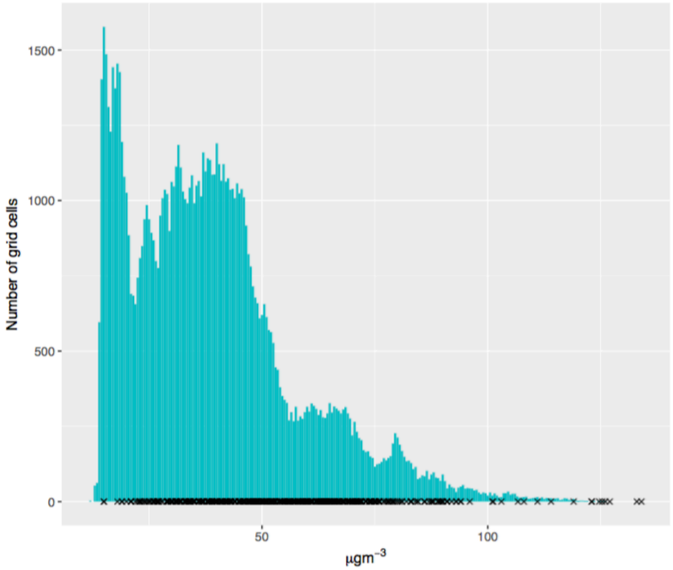


Figure: Estimated annual average concentrations of $PM_{2.5}$ by grid cell ($0.1^o \times 0.1^o$ resolution). Black crosses denote the annual averages recorded at ground monitors.

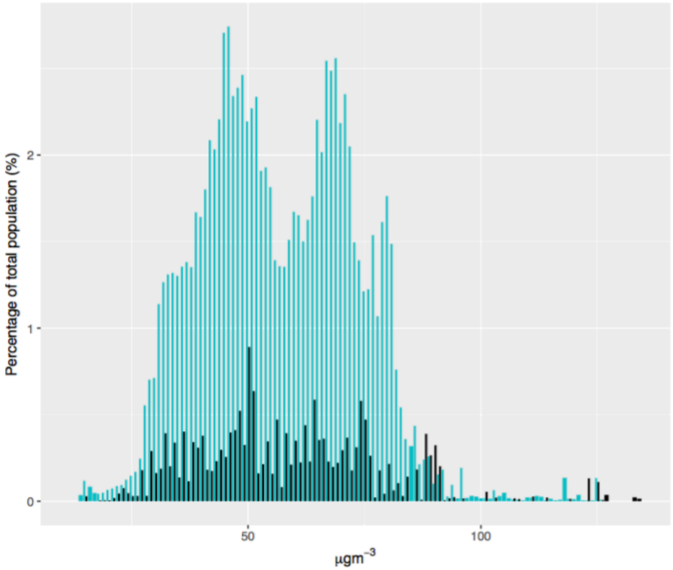
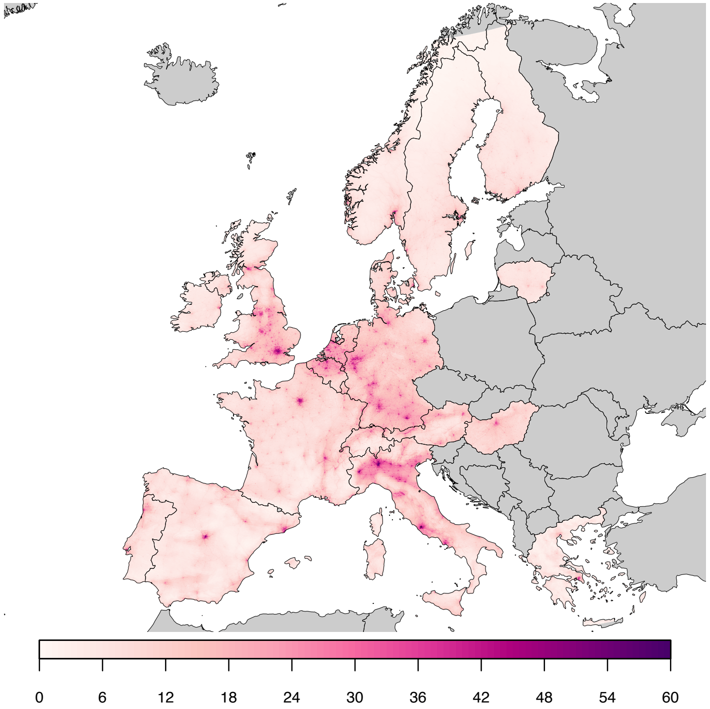


Figure: Estimated population level exposures (blue bars) and population weighted measurements from ground monitors (black bars).

The way to DIMAQ too!

Data, data and more data

- Rapid increase in number, and variety, of data sources
- Within country variation in calibration functions
- Higher resolution
- Time



Points and grids: spatially varying coefficient models

- Stage 1: Data at the lowest level of aggregation (point level) regressed against explanatory variables available at higher aggregation

$$Y_s = \beta_{0s} + \beta_{1s} X_B + \epsilon_s$$

- Stage 2: Regression coefficients are allowed to vary over space and time

$$\beta_{0s} \,\&\, \beta_{1s} \sim GP$$

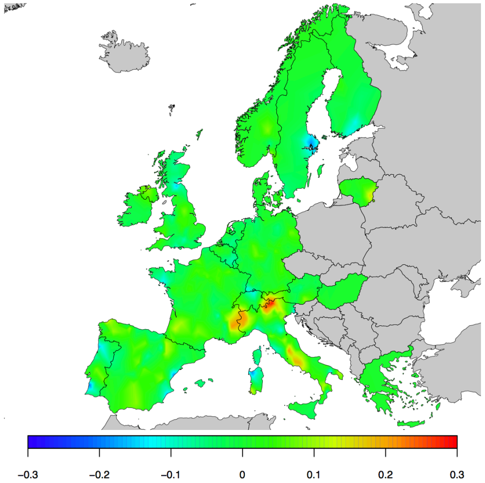
- SPDE models

DIMAQ2

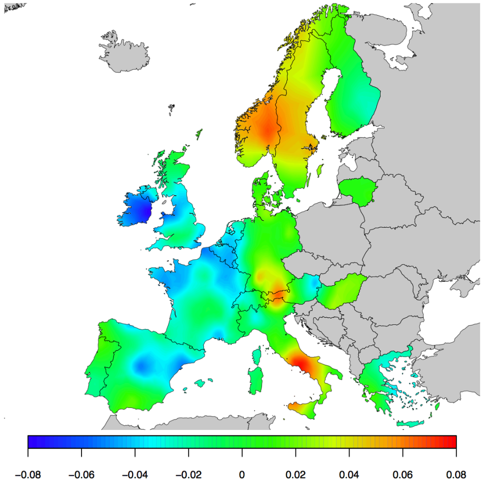
- **Space**: Continuous spatial process for coefficients
    - SPDE
    - PC priors
    - Within-country and within-grid cell variation (downscaling)
- **Time:** Temporal variation in the calibration coefficients
    - Random walks
- Predictions using Monte Carlo simulation
    - Joint samples from the posterior distributions of the parameters

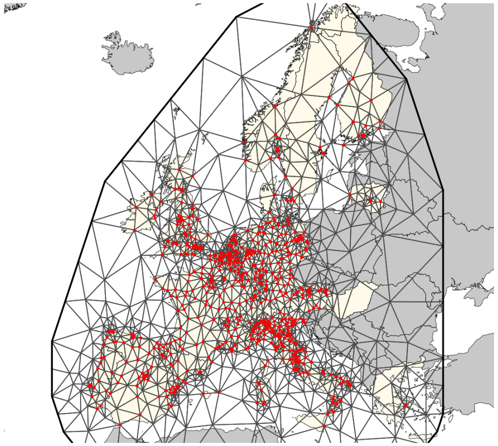
Spatial random effects (Europe)

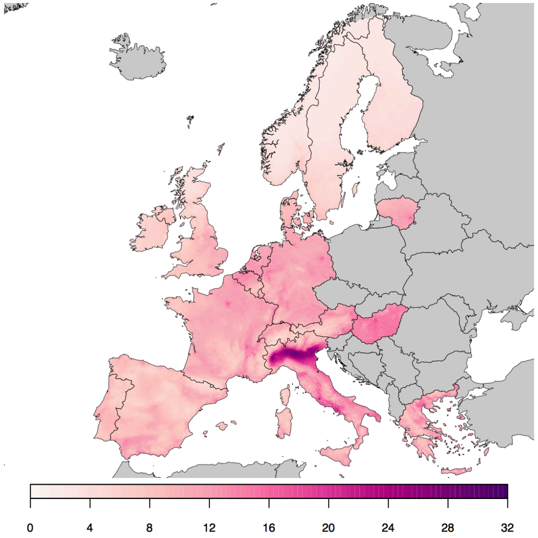
Intercept



Slope (SAT)

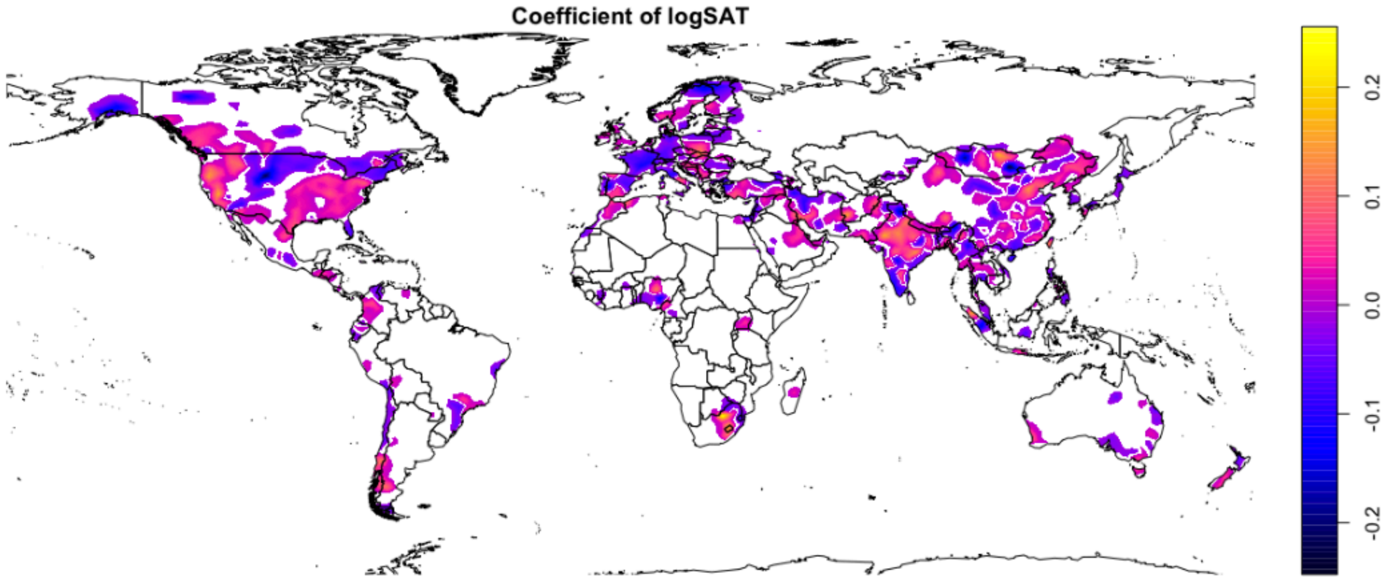






Annual average PM2.5, 1km x 1km, 2015

Spatial random effects (global)

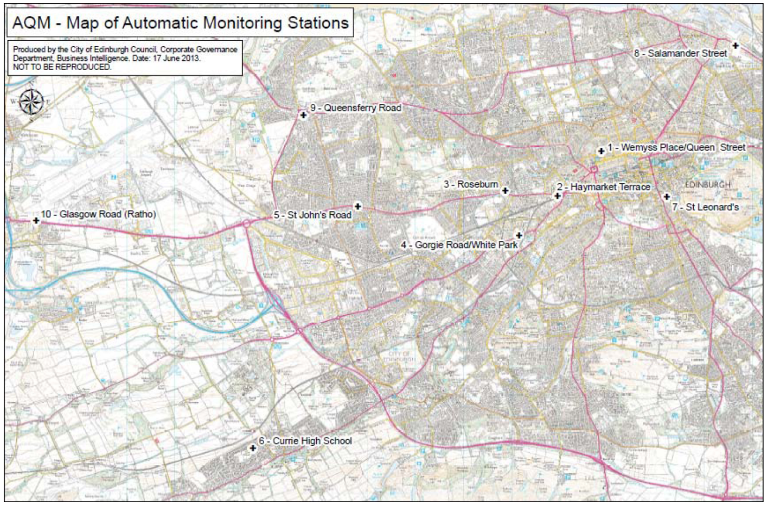


Annual average PM2.5, 10km x 10km, 2014

Black boxes

Where does it come from?

- What is it?
  - Measurements, model outputs….
- Using it for reasons other that those for which it was intended
  - non-standard sampling designs
  - preferential sampling
  - models for the data collection mechanisms

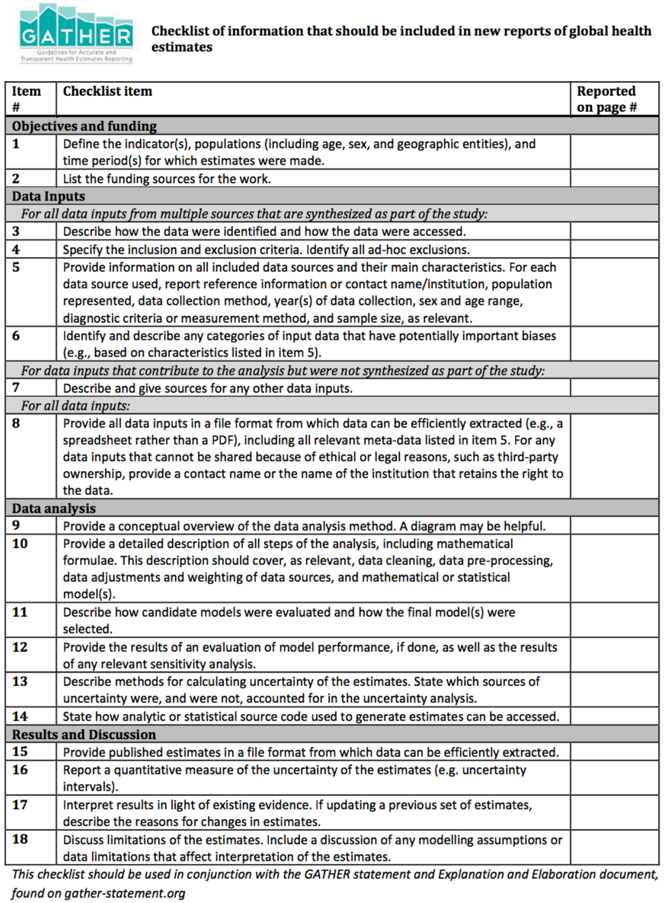


AQM - Map of Automatic Monitoring Stations

Produced by the City of Edinburgh Council, Corporate Governance Department, Business Intelligence. Date: 17 June 2013.
NOT TO BE REPRODUCED.

Where does it go?

- How to pass on complex information?
  - Propagating uncertainty from exposure models
- Black box
- Guidelines for Accurate and Transparent Health Estimates Reporting (GATHER)
  - http://gather-statement.org



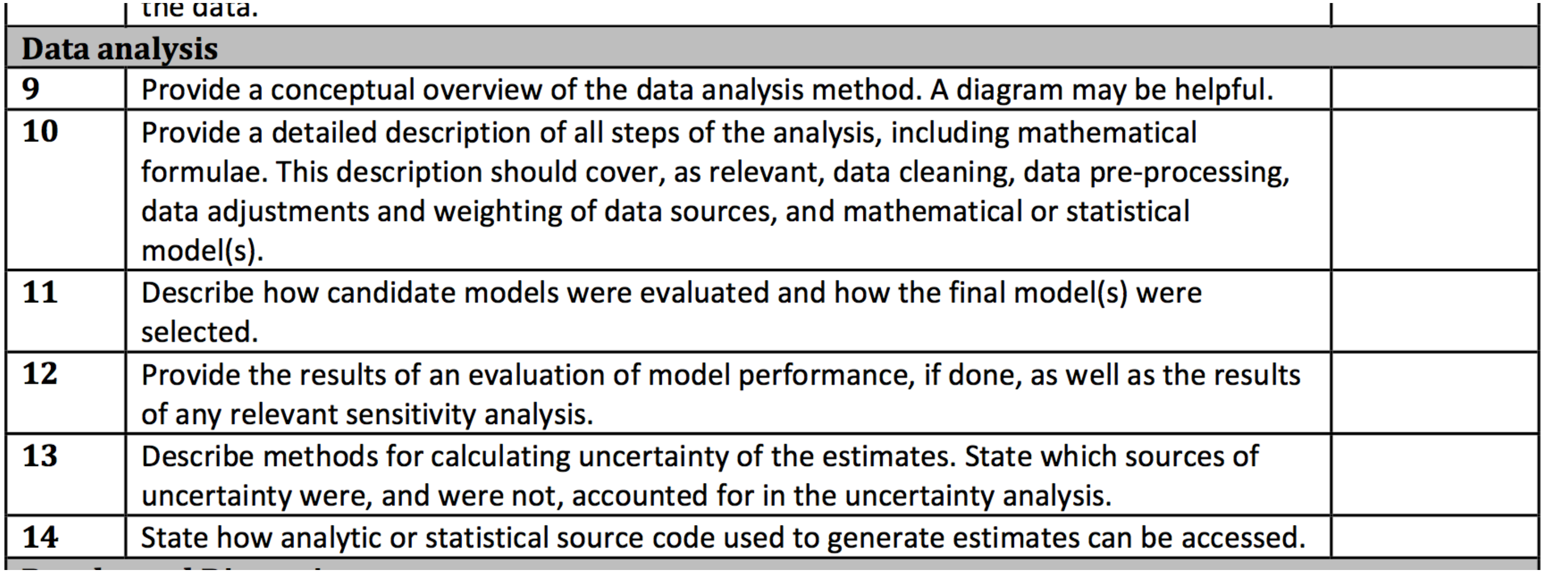
GATHER Guidelines for Accurate and Transparent Health Estimates Reporting

Checklist of information that should be included in new reports of global health estimates

Item #	Checklist item	Reported on page #
Objectives and funding		
1	Define the indicator(s), populations (including age, sex, and geographic entities), and time period(s) for which estimates were made.	
2	List the funding sources for the work.	
Data Inputs		
For all data inputs from multiple sources that are synthesized as part of the study:		
3	Describe how the data were identified and how the data were accessed.	
4	Specify the inclusion and exclusion criteria. Identify all ad-hoc exclusions.	
5	Provide information on all included data sources and their main characteristics. For each data source used, report reference information or contact name/institution, population represented, data collection method, year(s) of data collection, sex and age range, diagnostic criteria or measurement method, and sample size, as relevant.	
6	Identify and describe any categories of input data that have potentially important biases (e.g., based on characteristics listed in item 5).	
For data inputs that contribute to the analysis but were not synthesized as part of the study:		
7	Describe and give sources for any other data inputs.	
For all data inputs:		
8	Provide all data inputs in a file format from which data can be efficiently extracted (e.g., a spreadsheet rather than a PDF), including all relevant meta-data listed in item 5. For any data inputs that cannot be shared because of ethical or legal reasons, such as third-party ownership, provide a contact name or the name of the institution that retains the right to the data.	
Data analysis		
9	Provide a conceptual overview of the data analysis method. A diagram may be helpful.	
10	Provide a detailed description of all steps of the analysis, including mathematical formulae. This description should cover, as relevant, data cleaning, data pre-processing, data adjustments and weighting of data sources, and mathematical or statistical model(s).	
11	Describe how candidate models were evaluated and how the final model(s) were selected.	
12	Provide the results of an evaluation of model performance, if done, as well as the results of any relevant sensitivity analysis.	
13	Describe methods for calculating uncertainty of the estimates. State which sources of uncertainty were, and were not, accounted for in the uncertainty analysis.	
14	State how analytic or statistical source code used to generate estimates can be accessed.	
Results and Discussion		
15	Provide published estimates in a file format from which data can be efficiently extracted.	
16	Report a quantitative measure of the uncertainty of the estimates (e.g. uncertainty intervals).	
17	Interpret results in light of existing evidence. If updating a previous set of estimates, describe the reasons for changes in estimates.	
18	Discuss limitations of the estimates. Include a discussion of any modelling assumptions or data limitations that affect interpretation of the estimates.	

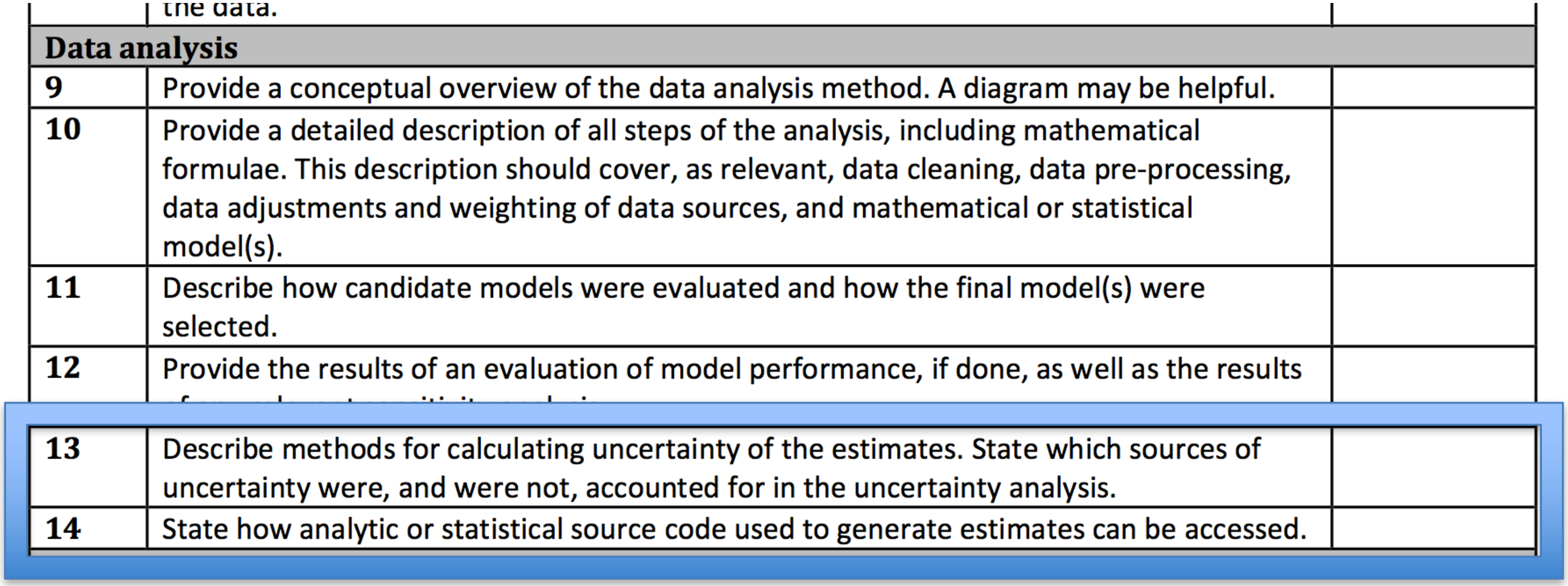
This checklist should be used in conjunction with the GATHER statement and Explanation and Elaboration document, found on gather-statement.org

Where does it go?



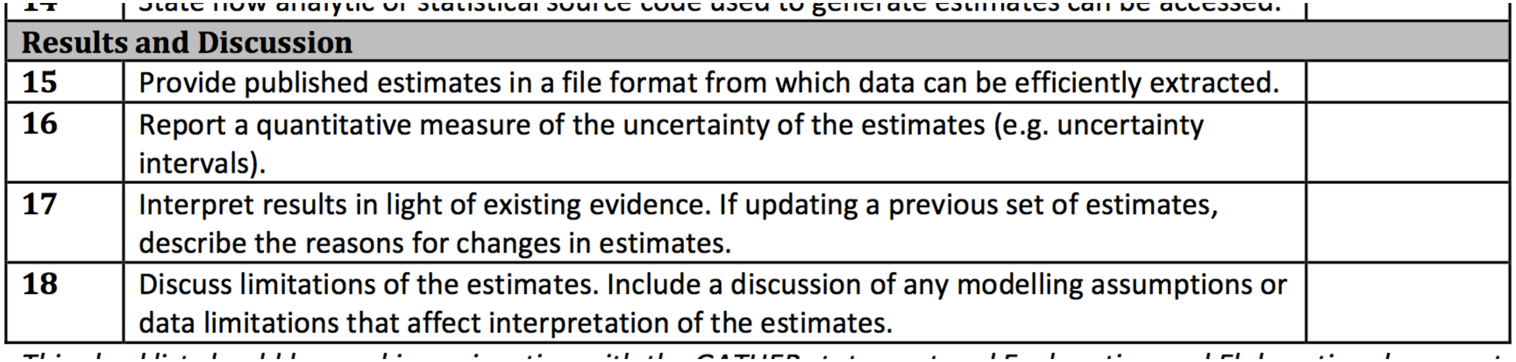
	the data.	
Data analysis		
9	Provide a conceptual overview of the data analysis method. A diagram may be helpful.	
10	Provide a detailed description of all steps of the analysis, including mathematical formulae. This description should cover, as relevant, data cleaning, data pre-processing, data adjustments and weighting of data sources, and mathematical or statistical model(s).	
11	Describe how candidate models were evaluated and how the final model(s) were selected.	
12	Provide the results of an evaluation of model performance, if done, as well as the results of any relevant sensitivity analysis.	
13	Describe methods for calculating uncertainty of the estimates. State which sources of uncertainty were, and were not, accounted for in the uncertainty analysis.	
14	State how analytic or statistical source code used to generate estimates can be accessed.	

Where do the estimates go?



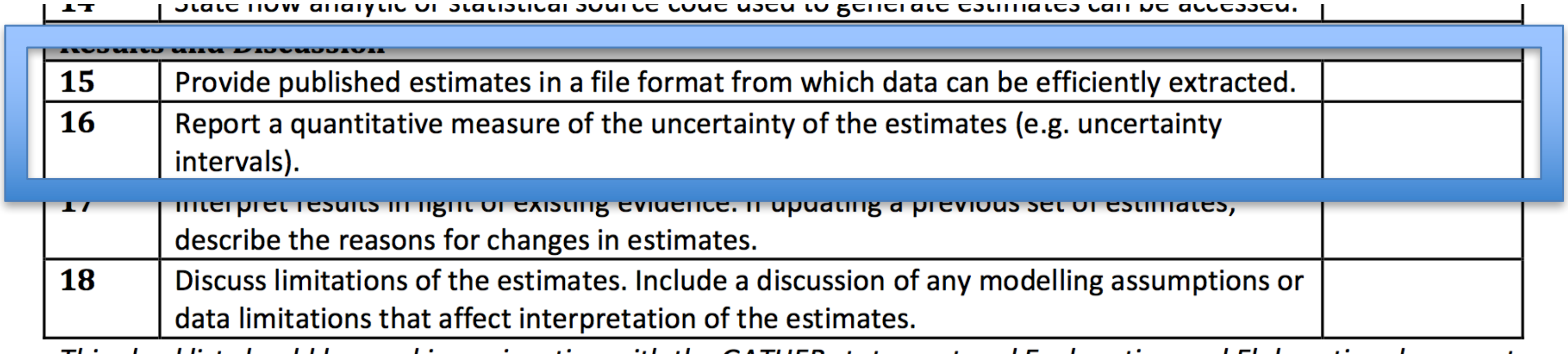
	the data.	
Data analysis		
9	Provide a conceptual overview of the data analysis method. A diagram may be helpful.	
10	Provide a detailed description of all steps of the analysis, including mathematical formulae. This description should cover, as relevant, data cleaning, data pre-processing, data adjustments and weighting of data sources, and mathematical or statistical model(s).	
11	Describe how candidate models were evaluated and how the final model(s) were selected.	
12	Provide the results of an evaluation of model performance, if done, as well as the results of any relevant sensitivity analysis.	
13	Describe methods for calculating uncertainty of the estimates. State which sources of uncertainty were, and were not, accounted for in the uncertainty analysis.	
14	State how analytic or statistical source code used to generate estimates can be accessed.	

Where does it go?



14	State how analytic or statistical source code used to generate estimates can be accessed.	
Results and Discussion		
15	Provide published estimates in a file format from which data can be efficiently extracted.	
16	Report a quantitative measure of the uncertainty of the estimates (e.g. uncertainty intervals).	
17	Interpret results in light of existing evidence. If updating a previous set of estimates, describe the reasons for changes in estimates.	
18	Discuss limitations of the estimates. Include a discussion of any modelling assumptions or data limitations that affect interpretation of the estimates.	

Where do the estimates go?



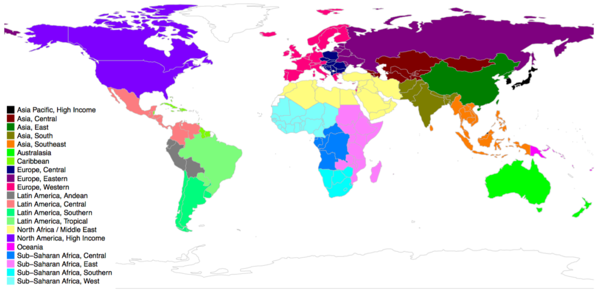
14	State how analytic or statistical source code used to generate estimates can be accessed.	
Results and Discussion		
15	Provide published estimates in a file format from which data can be efficiently extracted.	
16	Report a quantitative measure of the uncertainty of the estimates (e.g. uncertainty intervals).	
17	Interpret results in light of existing evidence. If updating a previous set of estimates, describe the reasons for changes in estimates.	
18	Discuss limitations of the estimates. Include a discussion of any modelling assumptions or data limitations that affect interpretation of the estimates.	

This checklist should be used in conjunction with the GATHER statement and Explanation and Elaboration document

So much more to do…

Future work **(want to join in the fun?)**

- Yearly updates
- Uncertainty
  - Distributions of population exposures
  - Incorporate uncertainty from relative risks
- Higher temporal resolution
  - Daily estimates
- Get involved earlier = AOD
- Preferential sampling



Bayesian melding

- Assumes an unobserved latent process, *Z*$_{st}$, which represents the underlying exposure, e.g. air pollution
- This process drives the different measurements
  - Monitoring data, $Y_{st}^{GM} = f(Z_{st})$
  - Remote sensing, $Y_{Bt}^{SAT} = f\left(\frac{1}{|B|}\int_B Z_{st} ds\right)$
  - Chemical transport models
- The responses are therefore 'linked'
  - Intrinsically correlated
  - Differences in scales are respected

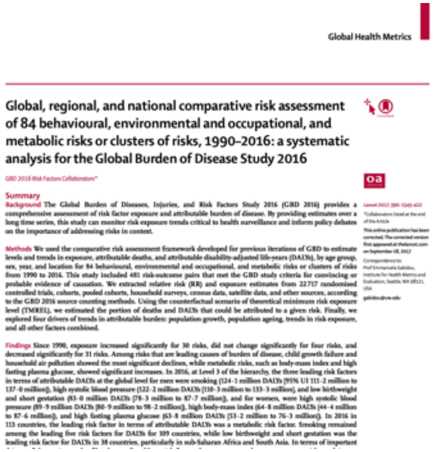
Further information

- WHO `Ambient air pollution: A global assessment of exposure and burden of disease’
- GBD2016 ‘Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016
- Data Integration Model for Air Quality: A Hierarchical Approach to the Global Estimation of Exposures to Ambient Air Pollution. JRSSC 2017

Ambient air pollution:

A global assessment of exposure and burden of disease

World Health Organization



Global Health Metrics

Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016

GBD 2016 Risk Factors Collaborators*

Summary

Background The Global Burden of Diseases, Injuries, and Risk Factors Study 2016 (GBD 2016) provides a comprehensive assessment of risk factor exposure and attributable burden of disease. By providing estimates over a long time series, this study can monitor risk exposure trends critical to health surveillance and inform policy debates on the importance of addressing risks in context.

Methods We used the comparative risk assessment framework developed for previous iterations of GBD to estimate levels and trends in exposure, attributable deaths, and attributable disability-adjusted life-years (DALYs), by age group, sex, year, and location for 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks from 1990 to 2016. This study included 481 risk-outcome pairs that met the GBD study criteria for convincing or probable evidence of causation. We extracted relative risk (RR) and exposure estimates from 22 717 randomised controlled trials, cohorts, pooled cohorts, household surveys, census data, satellite data, and other sources, according to the GBD 2016 source counting methods. Using the counterfactual scenario of theoretical minimum risk exposure level (TMREL), we estimated the portion of deaths and DALYs that could be attributed to a given risk. Finally, we explored four drivers of trends in attributable burden: population growth, population ageing, trends in risk exposure, and all other factors combined.

Findings Since 1990, exposure increased significantly for 30 risks, did not change significantly for four risks, and decreased significantly for 31 risks. Among risks that are leading causes of burden of disease, child growth failure and household air pollution showed the most significant declines, while metabolic risks, such as body-mass index and high fasting plasma glucose, showed significant increases. In 2016, at Level 3 of the hierarchy, the three leading risk factors in terms of attributable DALYs at the global level for men were smoking (124·1 million DALYs [95% UI 111·2 million to 137·0 million]), high systolic blood pressure (122·2 million DALYs [110·3 million to 133·3 million], and low birthweight and short gestation (83·0 million DALYs [78·3 million to 87·7 million]), and for women, were high systolic blood pressure (89·9 million DALYs [80·9 million to 98·2 million]), high body-mass index (64·8 million DALYs [44·4 million to 87·6 million]), and high fasting plasma glucose (63·8 million DALYs [53·2 million to 76·3 million]). In 2016 in 113 countries, the leading risk factor in terms of attributable DALYs was a metabolic risk factor. Smoking remained among the leading five risk factors for DALYs for 109 countries, while low birthweight and short gestation was the leading risk factor for DALYs in 38 countries, particularly in sub-Saharan Africa and South Asia. In terms of important

Lancet 2017; 390: 1345–422

*Collaborators listed at the end of the Article

This online publication has been corrected. The corrected version first appeared at thelancet.com on September 18, 2017

Correspondence to: Prof Emmanuela Gakidou, Institute for Health Metrics and Evaluation, Seattle, WA 98121, USA

gakidou@uw.edu

ROYAL STATISTICAL SOCIETY

Journal of the Royal Statistical Society
Applied Statistics
Series C

Appl. Statist. (2017)

Data integration model for air quality: a hierarchical approach to the global estimation of exposures to ambient air pollution

Gavin Shaddick,
University of Exeter and University of Bath, UK

Matthew L. Thomas and Amelia Green,
University of Bath, UK

Michael Brauer,
University of British Columbia, Vancouver, Canada

Aaron van Donkelaar,
Dalhousie University, Halifax, Canada

Rick Burnett,
Health Canada, Ottawa, Canada

Howard H. Chang,
Emory University, Atlanta, Canada

Aaron Cohen,
Health Effects Institute, Boston, USA

Rita Van Dingenen,
European Commission, Ispra, Italy

Carlos Dora and Sophie Gumy,
World Health Organization, Geneva, Switzerland

Yang Liu,
Emory University, Atlanta, USA

Randall Martin,
Dalhousie University, Halifax, Canada

Address for correspondence: Gavin Shaddick, Department of Mathematics, University of Exeter, Streatham Campus, Exeter, EX4 4QF, UK.