Trends in country-level exposures to particulate matter air pollution

Amelia Green [1] Gavin Shaddick [1] Matthew L. Thomas [1] Michael Brauer [2] Aaron van Donkelaar [3] Rick Burnett [4] Howard H. Chang [5] Aaron Cohen [6] Rita Van Dingenen [11] Carlos Dora [7] Sophie Gumy [7] Yang Liu [8] Randall Martin [3] Lance A. Waller [5] Jason West [9] James V. Zidek [10] Annette Prüss-Ustün [10]

[1]Department of Mathematical Sciences, University of Bath, Bath, UK [2]School of Population and Public Health, The University of British Columbia, Vancouver, British Colombia, Canada [3]Department of Physics and Atmospheric Science, Dalhousie University, Halifax, Nova Scotia, Canada [4]Health Canada, Ottawa, Ontario, Canada [5]Department of Biostatistics, Rollins School of Public Health, Emory University, Atlanta, Georgia, USA [6]Health Effects Institute, Boston, Massachusetts, USA [7]World Health Organisation, Geneva, Switzerland [8]Department of Environmental Health,Rollins School of Public Health, Emory University, Atlanta, Georgia, USA [9]Department of Environmental Sciences and Engineering, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA [10]Department of Statistics, University of British Columbia, Vancouver, British Columbia, Canada [11]Institute for Environment and Sustainability, Joint Research Centre, European Commission, Italy

Overview

Air pollution has become a growing concern in the past few years, with an increasing number of acute air pollution episodes in many cities worldwide. As a result, data on air quality is becoming increasingly available and the science underlying the related health impacts is also evolving rapidly. To date, air pollution, both ambient (outdoor) and household (indoor), is the biggest environmental risk to health, carrying responsibility for about one in every nine deaths annually. Ambient (outdoor) air pollution alone kills around 3 million people each year, mainly from noncommunicable diseases. Only one person in ten lives in a city that complies with the WHO Air quality guidelines. Air pollution continues to rise at an alarming rate, and affects economies and people's quality of life.

Introduction

The recently developed Data Integration Model for Air Quality (DIMAQ) integrates data from from multiple sources, including satellite observations of aerosols in the atmosphere, ground measurements, chemical transport model simulations, population estimates and land-use data, to provide information on population-weighted exposures to ambient air pollution, defined as the population-weighted annual average of $PM_{2.5}$; particles with an aerodynamic diameter less than 2.5 micrometers.

This results in a wealth of information on levels of air pollution around the world and highlights areas within countries that exceed WHO air quality limits. Such information is vital for health impact assessment, policy support and developing mitigation strategies.

Statistical Modelling

The DIMAQ model was used to calibrated ground measurements, where available, against data from the other sources. Coefficients in the calibration model were estimated for each country. Where data were insufficient within a country, information was 'borrowed' from a higher aggregation (region) or, if enough information was still not available, from an even higher level (super-region). This was implemented within a Bayesian Hierarchical modelling framework.

The results of the modelling comprise a posterior distribution for each grid cell, rather than just a single point estimate, allowing a variety of summaries to be calculated. The primary outputs here are the median and 95% credible intervals for each grid cell.

Due to both the complexity of the models and the size of the data, notably the number of spatial predictions that are required, recently developed techniques that perform 'approximate' Bayesian inference based on integrated nested Laplace approximations (INLA) were used. Computation was performed using the R interface to the INLA computational engine (R-INLA). Fitting the models and performing predictions for each of the ca. 1.4 million grid cells required the use of a high performance computing cluster (HPC) making use of high memory nodes.

Satellite estimates, populations and quantities estimated using the GEOS-Chem model were available for 1990, 1995, 2000, 2005, 2010, 2011, 2012, 2013 and 2014. Population estimates for 2000, 2005, 2010, 2015 and 2020 were available from GPW version 4. For 1990 and 1995 data were extracted from GPW version 3, as in GBD2013. As with populations for 2014, values for each cell for 2011, 2012, 2012 and 2013 were obtained by interpolation using natural splines with knots placed at 2000, 2005, 2010, 2015 and 2020.

These were used as inputs to the final model, enabling estimates of exposures to be obtained for each of these years respectively. For 2015, predictions of exposures (and associated measures of uncertainty) were obtained by fitting smoothing splines within each cell to the medians and limits of the 95% credible intervals for each cell from 2010 to 2014 and extrapolating to 2015. The smoothing splines were fit within a Generalised Additive Modelling framework with the degrees of freedom calculated for each cell by multivariate generalised cross validation.

Results

Predictions from the DIMAQ model for 2014 can be seen in Figure 1. The point estimates shown here give a summary of air quality for each grid cell and clearly show the spatial variation in global $PM_{2.5}$. For each grid cell, there is an underlying (posterior) probability distribution which incorporates information about the uncertainty of these estimates.

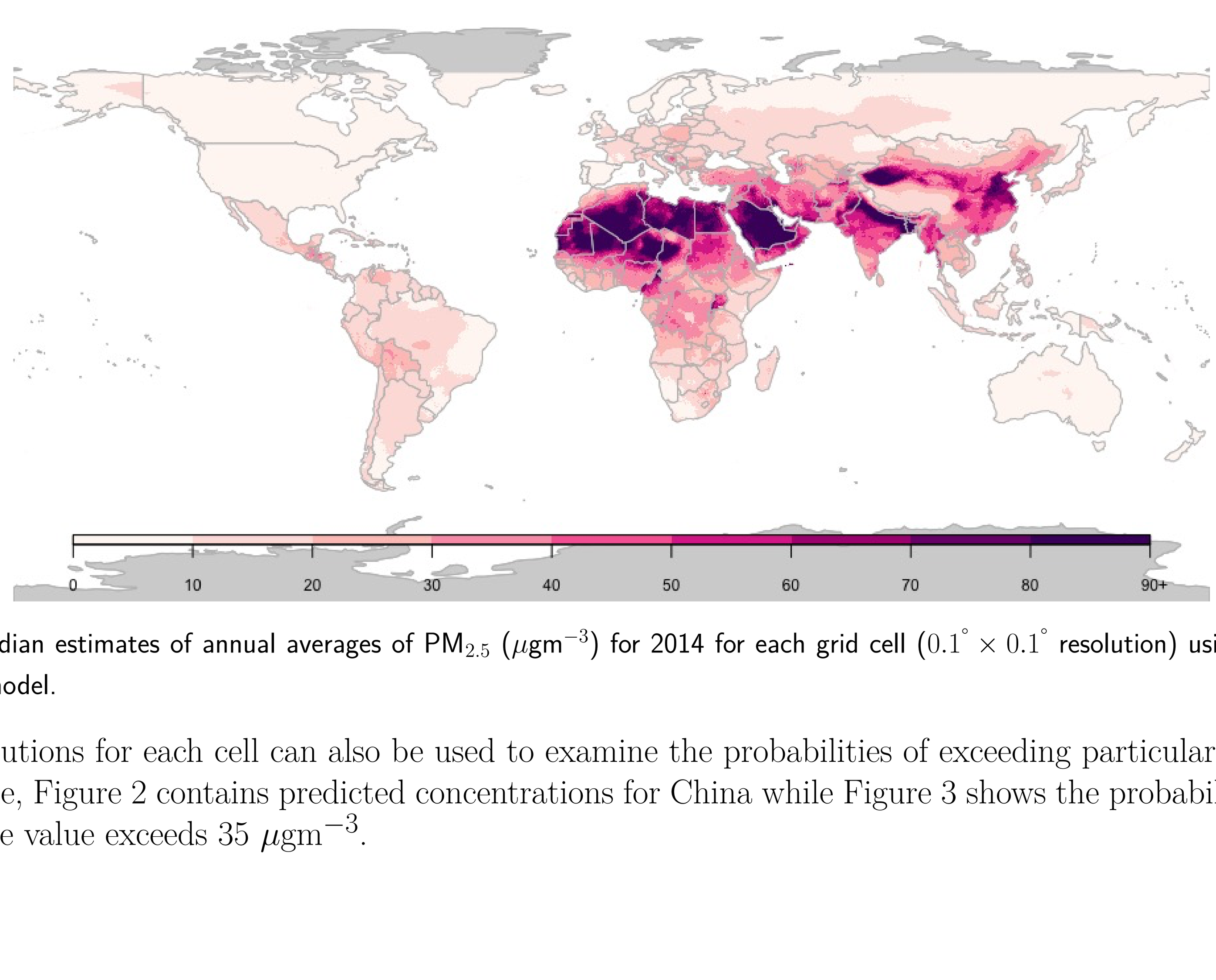

Figure 1: Median estimates of annual averages of $PM_{2.5}$ (μgm^{-3}) for 2014 for each grid cell ($0.1^\circ \times 0.1^\circ$ resolution) using a Bayesian hierarchical model.

The distributions for each cell can also be used to examine the probabilities of exceeding particular thresholds. For example, Figure 2 contains predicted concentrations for China while Figure 3 shows the probability for each cell that the value exceeds 35 μgm^{-3}.

Results (ctd.)

The profile of air pollution ($PM_{2.5}$) in China contains three distinct components:

- a land mass with low levels of air pollution;
- a much larger proportion of the total land mass with (comparatively) high levels; and
- a substantial area with very high levels.

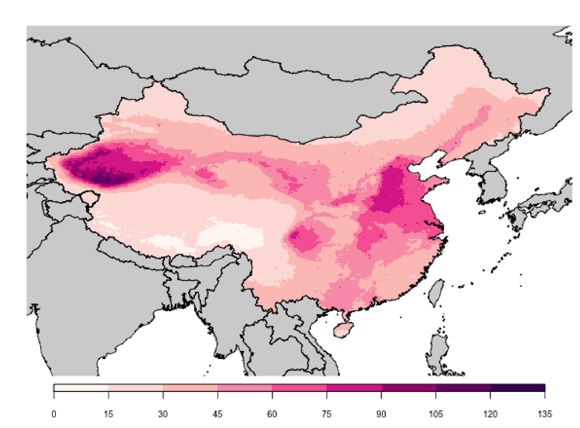

Figure 2: Medians of posterior distributions for estimates of annual mean $PM_{2.5}$ concentrations (μgm^{-3}) for 2014, in China.

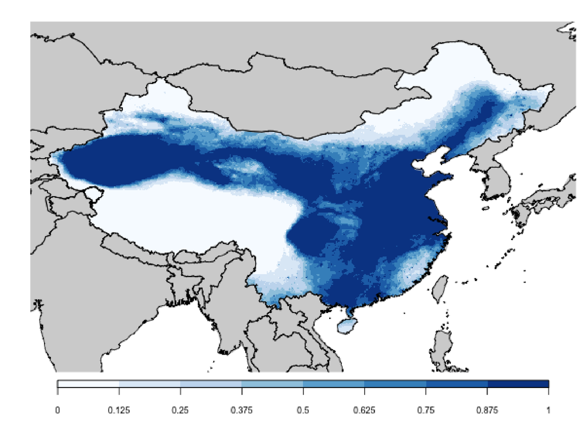

Figure 3: Probability of exceeding 35 μgm^{-3} using a Bayesian hierarchical model for each grid cell ($0.1^\circ \times 0.1^\circ$ resolution) for 2014, in China.

The distribution of estimated exposures shown in the map of median values of the posterior distributions, shown in Figure 2, can also be seen in Figure 4. These can be translated into distributions of population exposures by matching estimated concentrations with population estimates. Figure 5 shows the distribution of estimated population level exposures, calculated by multiplying the estimate in each grid cell by the estimated population.

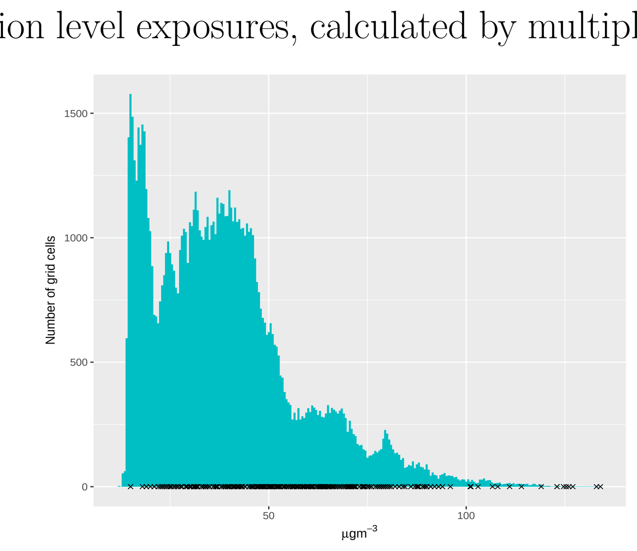

Figure 4: Estimated annual average concentrations of $PM_{2.5}$ by grid cell ($0.1^o \times 0.1^o$ resolution). Black crosses denote the annual averages recorded at ground monitors.

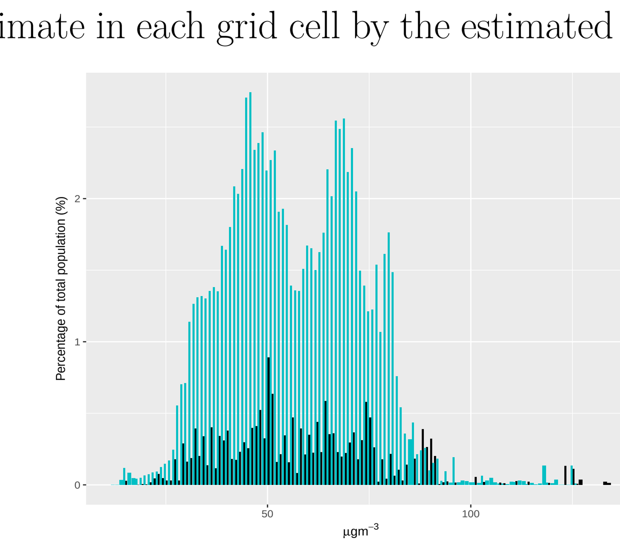

Figure 5: Estimated population level exposures (blue bars) and population weighted measurements from ground monitors (black bars).

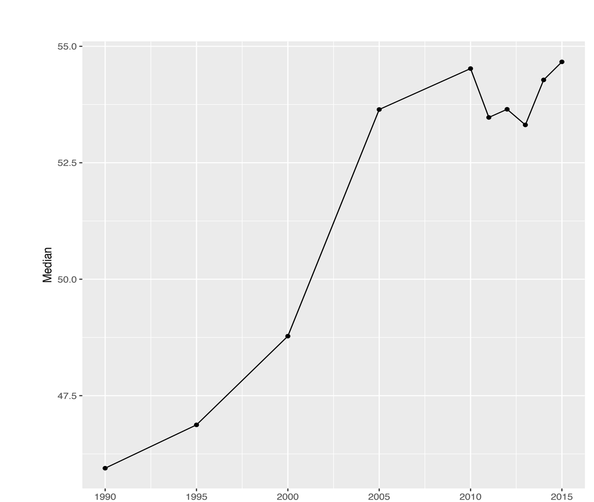

Figure 6: Average $PM_{2.5}$ concentrations (μgm^{-3}) for China, 1990-2015.

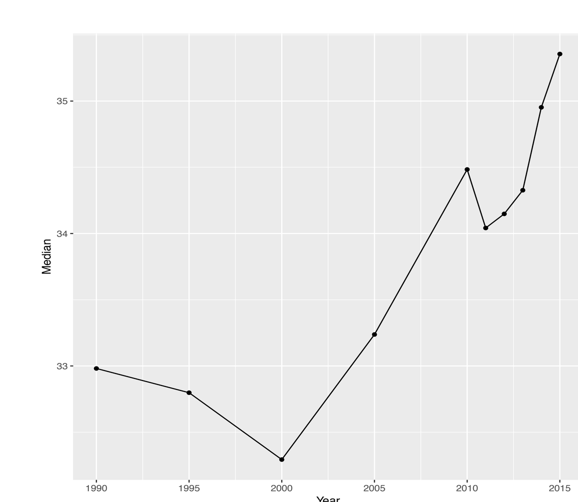

Figure 7: Population weighted average $PM_{2.5}$ concentrations (μgm^{-3}) for China , 1990-2015.

Figures 6 and 7 show trends in exposures for China with the population weighted exposures increasing at a faster rate (10μgm^{-3} over 30 years) than the unweighted exposures (2μgm^{-3} over 30 years). As in the distributions presented in Figures 4 and 5, there is a clear difference in levels of air pollution in urban and rural areas, with levels being higher in locations where large populations reside. This effect is likely to increase as populations migrate from rural to urban areas in many parts of the world.

Discussion

The DIMAQ model is fit within a Bayesian hierarchical framework which produces full posterior distributions for estimated levels of $PM_{2.5}$ for each grid-cell rather than just point estimates. Summaries of these posterior distributions can be used to give point estimates, e.g. mean, median, measures of uncertainty, e.g. 95% credible intervals, and exceedance probabilities, e.g. the probability of exceeding air quality guidelines. Such information is vital in creating accurate estimates of population exposures for use in health analysis, policy support and as a basis of mitigation strategies. Based on the posterior estimates for 2014, 92% of the world's population reside in areas where the WHO guideline of 10 μgm^{-3} is exceeded.