Learning Global Variations in Outdoor $PM_{2.5}$ Concentrations with Satellite Images

Kris Y. Hong,[1] Pedro O. Pinheiro,[2] Scott Weichenthal[1]

[1]McGill University, Montreal, Canada; [2]Element AI, Montreal, Canada

Introduction

- Fine particulate matter ($PM_{2.5}$) kills millions annually with economic impacts measured in billions of dollars
- Cost-effective methods for estimating air pollution are needed to support pollution mitigation and health research
- Traditional geostatistical models for predicting exposures rely on detailed geographic information (e.g. traffic, land use) that are not always available
- Alternatively, this geographic information can be captured through satellite imagery

Methods

- 20,000 annual average measurements among 6000 global sites (spanning 2010-2016) were compiled from the WHO and grouped into ~156 x 156km geohash cells

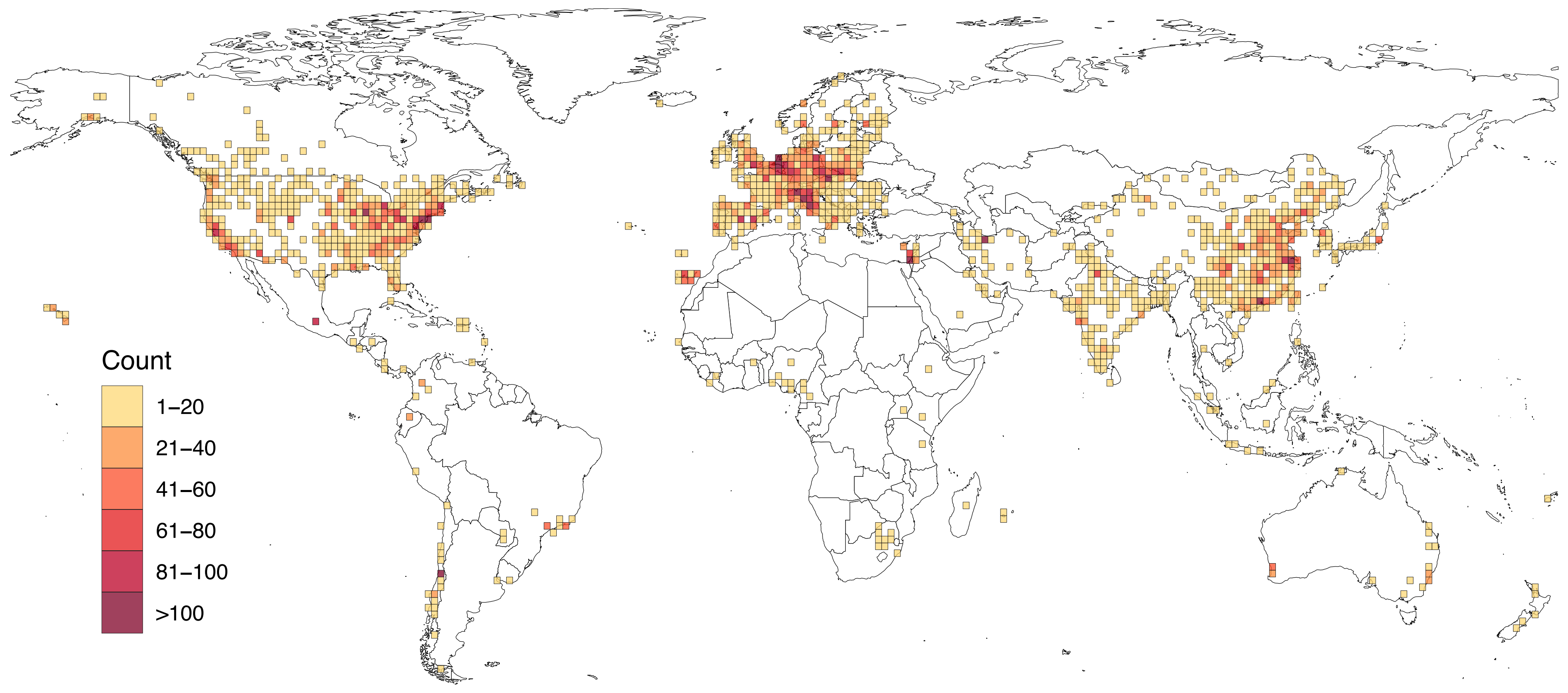


Figure 1: *Locations of global $PM_{2.5}$ monitoring sites grouped into 1200 geohash cells.*

- Zoom level 13 to 16 satellite images centred on measurement sites were downloaded from Google static maps

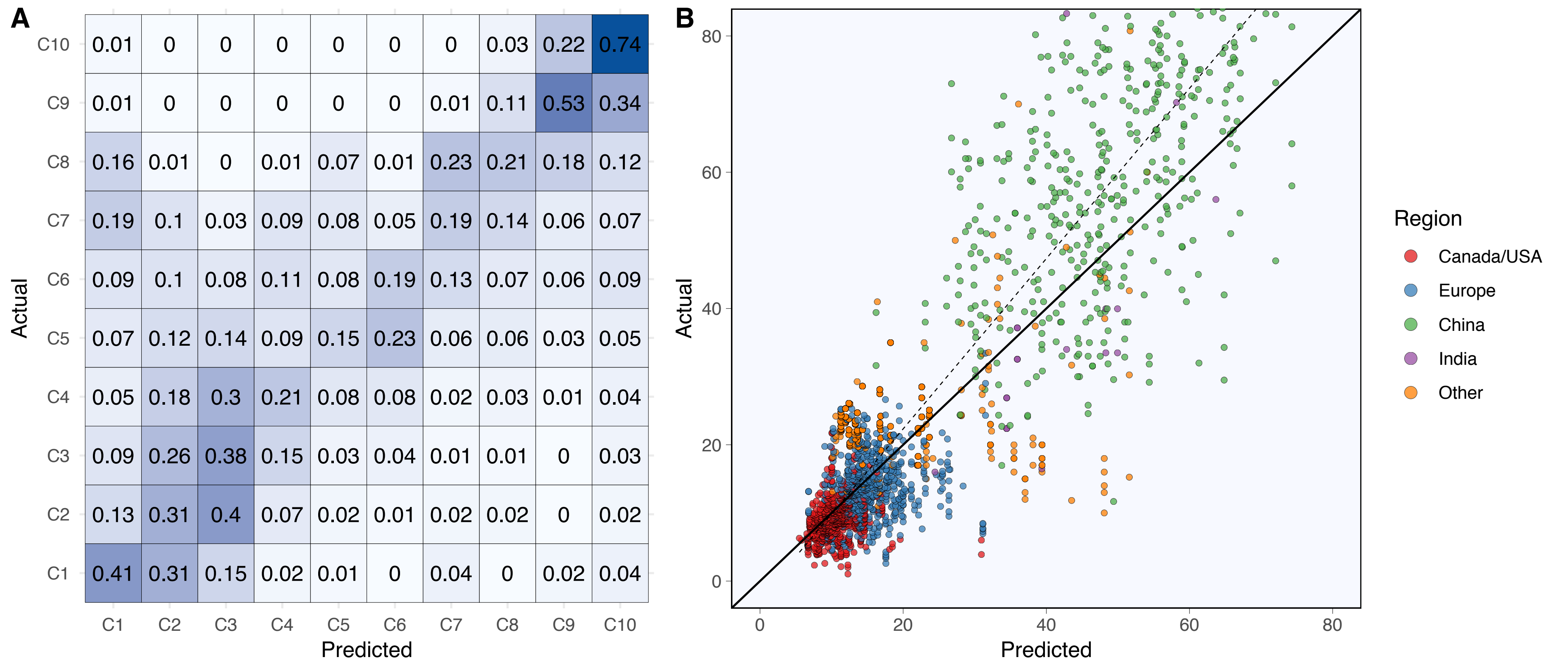
Figure 2: *From left to right, respectively: zoom level 13 (10 x 10km) through 16 (1.5 x 1.5km)*

- Data were randomly split into disjoint training (80%), validation (10%), and test sets (10%) by geohash cells
- Categorical (10 balanced classes split evenly by deciles of $PM_{2.5}$ distribution) and continuous models were developed
- Optimal configuration consisted of zoom level 13 images and an Xception base initialized with ImageNet weights
- Model performance was compared to "gold standard" DIMAQ geostatistical model from the Global Burden of Disease Study

Results

Classification accuracy: 33.7%
One-off classification accuracy: 65.7%

Test set SD: 23.82 $\mu g/m^3$
Regression RMSE: 13.01 $\mu g/m^3$



Actual \ Predicted	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
C10	0.01	0	0	0	0	0	0	0.03	0.22	0.74
C9	0.01	0	0	0	0	0	0.01	0.11	0.53	0.34
C8	0.16	0.01	0	0.01	0.07	0.01	0.23	0.21	0.18	0.12
C7	0.19	0.1	0.03	0.09	0.08	0.05	0.19	0.14	0.06	0.07
C6	0.09	0.1	0.08	0.11	0.08	0.19	0.13	0.07	0.06	0.09
C5	0.07	0.12	0.14	0.09	0.15	0.23	0.06	0.06	0.03	0.05
C4	0.05	0.18	0.3	0.21	0.08	0.08	0.02	0.03	0.01	0.04
C3	0.09	0.26	0.38	0.15	0.03	0.04	0.01	0.01	0	0.03
C2	0.13	0.31	0.4	0.07	0.02	0.01	0.02	0.02	0	0.02
C1	0.41	0.31	0.15	0.02	0.01	0	0.04	0	0.02	0.04

Figure 3: *Measured versus predicted $PM_{2.5}$ concentrations (C1: low; C10: high) in the test set*

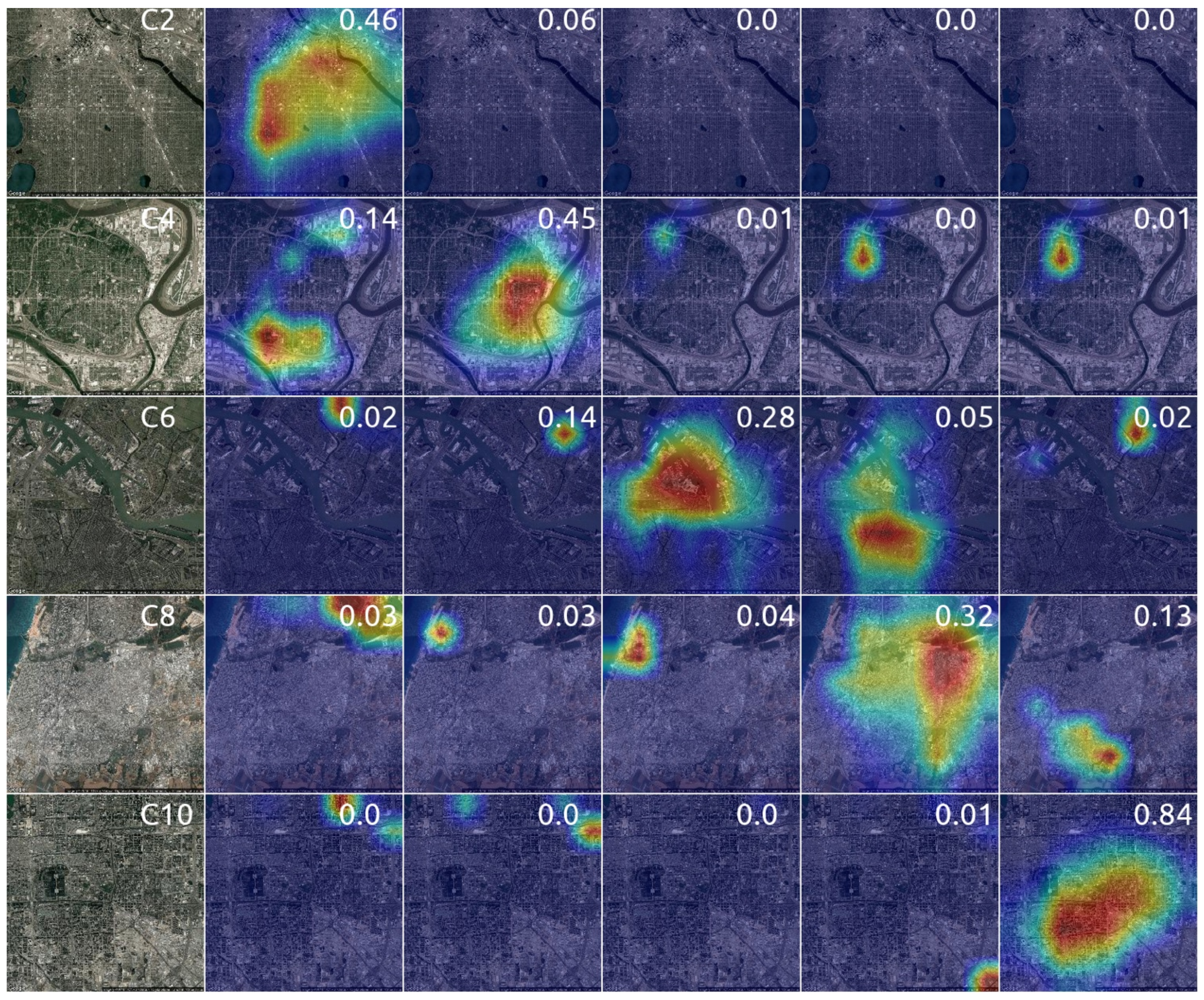


Figure 4: *Gradient-weighted class activation maps for images correctly classified by the final model. Column 1: Original Column 2-6: GRAD-CAM for classes 2, 4, 6, 8, 10, respectively. Numerical values indicate predicted probability of the image belonging to the respective class*

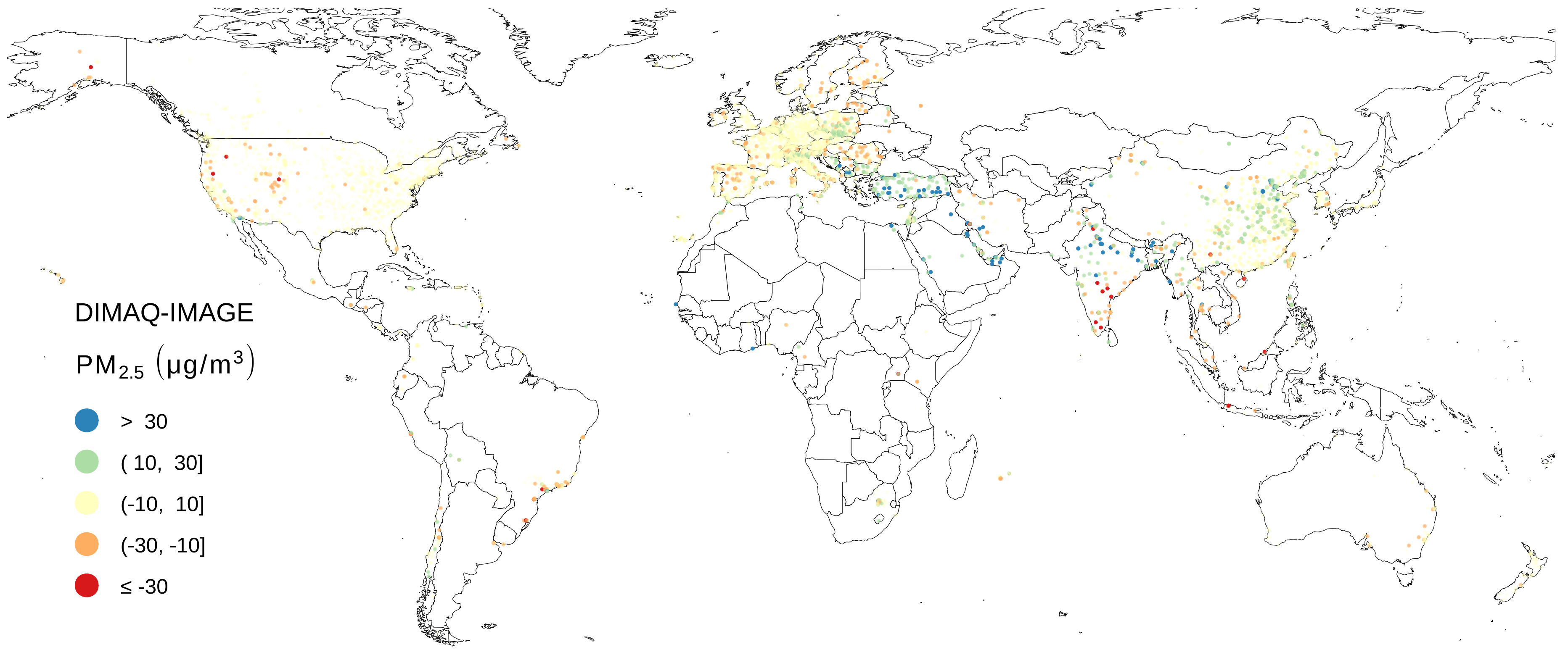


Figure 5: *Difference in predicted $PM_{2.5}$ between the IMAGE-$PM_{2.5}$ and DIMAQ models*

Discussion

- The IMAGE-$PM_{2.5}$ model offers a fast cost-effective method for estimating global variations in annual average $PM_{2.5}$
- Model could be improved with timestamped hi-res imagery
- Satellite images could serve as a predictor for other exposures

Weichenthal Research Group
scottweichenthal.weebly.com

References

Shaddick, G., Thomas, M. L., Green, A., Brauer, M., Donkelaar, A., Burnett, R., Chang, H. H., Cohen, A., Dingenen, R. V., Dora, C., et al. Data integration model for air quality: a hierarchical approach to the global estimation of exposures to ambient air pollution. Journal of the Royal Statistical Society: Series C (Applied Statistics), 2018

Download the paper