Combining citizen science and deep learning for large-scale estimation of outdoor nitrogen dioxide concentrations

Peer-reviewed author version

Weichenthal, Scott; DONS, Evi; Hong, Kris Y; Pinheiro, Pedro O & Meysman, Filip J R (2021) Combining citizen science and deep learning for large-scale estimation of outdoor nitrogen dioxide concentrations. In: ENVIRONMENTAL RESEARCH, 196 (Art N° 110389).

DOI: 10.1016/j.envres.2020.110389
Handle: http://hdl.handle.net/1942/33188

Report from the Field

Combining Citizen Science and Deep Learning for Large-Scale Estimation of Outdoor Nitrogen Dioxide Concentrations

Scott Weichenthal[1]*, Evi Dons[2], Kris Y. Hong[1,3], Pedro O. Pinheiro[3], Filip J.R. Meysman[4]

[1] Department of Epidemiology, Biostatistics, and Occupational Health, McGill University, Montreal, Canada.
[2] Centre for Environmental Sciences, Hasselt University, Hasselt, Belgium.
[3] Element AI, Montreal, Canada.
[4] Department of Biology, University of Antwerp, Wilrijk, Belgium.

*Corresponding Author
Scott Weichenthal
Faculty of Medicine
Department of Epidemiology, Biostatistics, and Occupational Health, McGill University
1110 Pins Ave West, Montreal, QC, H3A 1A3
Email: scott.weichenthal@mcgill.ca
Tel: (514) 398-1584

Funding: This work was supported by an NSERC Discovery Grant and a CIHR Foundation Grant (Weichenthal PI) and a postdoctoral scholarship from FWO Research Foundation Flanders (Dons).

Conflict of Interest: None

Author Contributions: Scott Weichenthal was the lead author of the manuscript; all co-authors contributed to writing portions of the manuscript and approved the final version. Kris Hong and Pedro Pinheiro developed computer code, trained all deep learning models, and created the Figures included in the manuscript. Evi Dons and Filip Meysman were responsible for collecting the original NO_2 data used to train the models and contributed to writing portions of the manuscript.

Abstract

Reliable estimates of outdoor air pollution concentrations are needed to support global actions to improve public health. We developed a new approach to estimating annual average outdoor nitrogen dioxide (NO_2) concentrations using approximately 20,000 ground-level measurements in Flanders, Belgium combined with aerial images and deep neural networks. Our final model explained 79% of the spatial variability in NO_2 (root mean square error of 10-fold cross-validation=3.58 μg/m^3) using only images as model inputs. This novel approach offers an alternative means of estimating large-scale spatial variations in ambient air quality and may be particularly useful for regions of the world without detailed emissions data or land use information typically used to estimate outdoor air pollution concentrations.

Highlights:

- Citizen science can be used to collected air pollution data over broad spatial scales
- We combined NO2 data collected through citizen science with aerial image and deep learning models to predict annual average NO_2 concentrations in Flanders, Belgium
- The final models explained the majority of spatial variations in NO_2 using only image as input
- This method may be useful in regions lacking information on traditional model inputs including land use data or emissions information.

Keywords: Nitrogen dioxide; Citizen Science; Deep learning; Convolutional neural networks

1. Introduction

Outdoor air pollution causes millions of premature deaths around the world each year with estimated economic impacts measured in billions of dollars annually (1, 2). In order to address this global health concern, regulatory policies must be based on accurate and reliable population exposure data that suitably captures fine-scale spatial variations in pollutant concentrations over broad geographic areas. In particular, ambient nitrogen dioxide (NO_2) forms a key proxy for complex air pollution mixtures resulting from traffic emissions and other combustion sources, and elevated NO_2 levels have been associated with both acute and chronic illnesses causing pre-mature mortality (3). The current approach to modelling large-scale spatial variations in outdoor NO_2 relies on the application of statistical or chemical transport models that are fed with data streams from reference monitoring networks, complemented with satellite measurements, traffic and power plant emissions data, and land use information such as proximity to roadways, population density, green space, and active fire locations (4). However, considerable uncertainty remains with respect to ambient NO_2 concentrations in many portions of the world owing to sparse coverage of regulatory monitoring networks and/or missing data for parameters needed to estimate outdoor concentrations (4). Given the high cost of developing and maintaining conventional reference monitoring stations, many parts of the world will continue to have major gaps in air pollution data coverage until alternative approaches are developed that can provide accurate predictions in the absence of detailed emissions data and/or land use information (5). This is particularly true for developing countries where resources are limited. In this study, we developed a new approach to estimating ground-level NO_2 concentrations using a deep learning model trained

using aerial images and ground-level NO_2 data collected from a large-scale citizen science project in Flanders, Belgium (6). The resulting model provides reliable predictions of annual average outdoor NO_2 concentrations over a broad spatial scale using only digital images as input.

2. Methods

A database of approximately 20,000 annual average NO_2 measurements and corresponding latitude-longitude coordinates spread across Flanders, Belgium was compiled from the "CurieuzeNeuzen" ("Curious Noses") project. In brief, this citizen science project mass-mobilized families, schools, companies, and social organizations to measure NO_2 concentrations across Flanders, Belgium (Figure 1) (6). Data were collected over a 4-week period in May 2018 using low-cost passive samplers (in duplicate) and were subsequently scaled to annual average concentrations. Annual average outdoor NO2 concentrations for sites included in our analyses ranged from 10.9-75.3 μg/m³ (mean: 24.1 μg/m³; median: 21.9 μg/m³) across the study area.

Aerial images centered on each latitude-longitude pair for ground-level NO_2 data were downloaded from Google Static Maps using the *ggmap* package in the R statistical computing environment (7, 8). Three aerial images were downloaded for each coordinate differing by integer zoom levels of 16 (covering approximately 970 x 970m), 18 (240 x 240m) and 19 (120 x 120m). All images were saved at a resolution of 256 x 256 x 3 to maintain a reasonable training time. The subsequent database of image-NO_2 pairs was split into 11 equal sized folds for model training and evaluation.

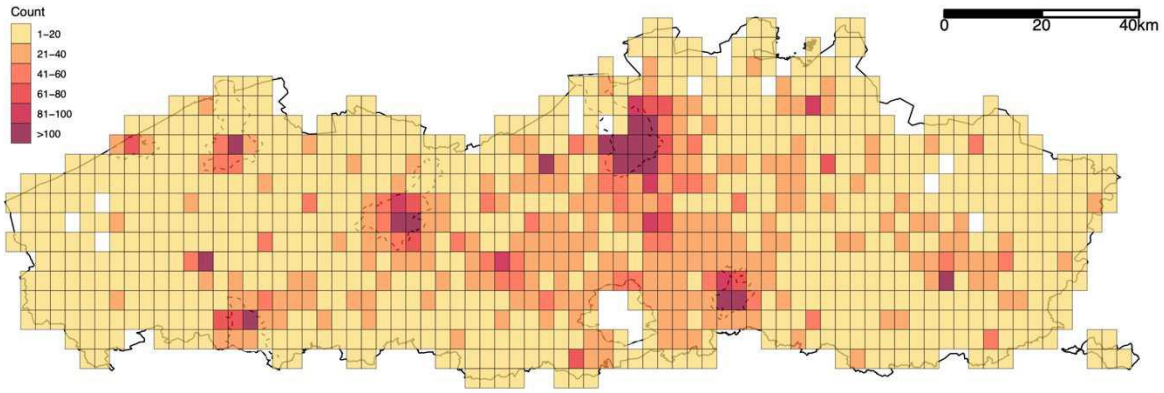

Fig 1. Monitoring locations for annual average NO_2 concentrations across Flanders, Belgium (2018). Major urban areas are indicated by dashed outlines.

Convolutional neural networks (CNNs) with two input images of different zoom levels (i.e. different spatial scales, one image to capture local information and one image to capture regional information) were trained to predict annual average NO_2 concentrations (Figure 2).

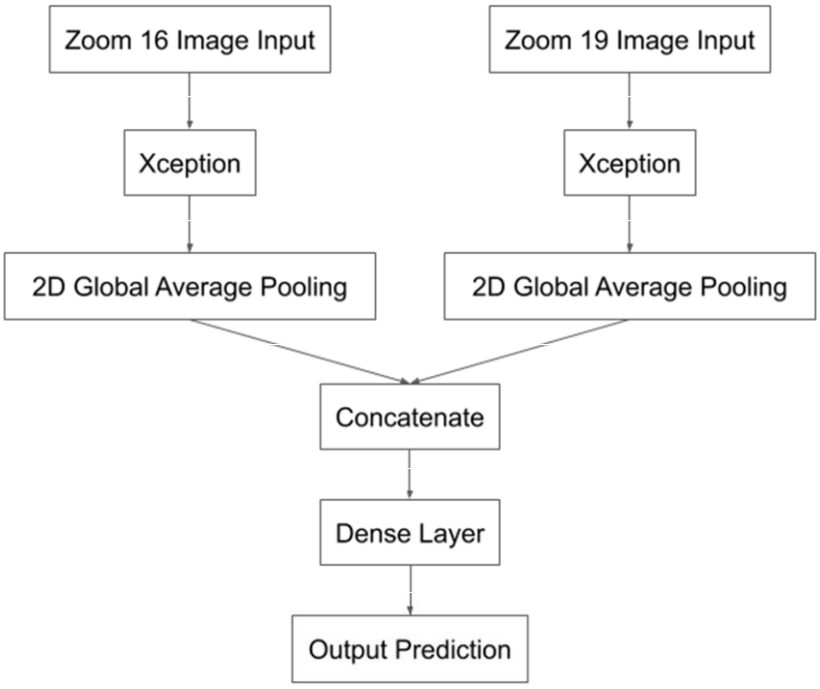

Fig 2. Summary of model architecture used to predict NO2 concentrations from images

Each input image was fed into an independent Xception convolutional base (9, 10) for feature extraction, followed by a 2D global average pooling operation. The outputs of the two CNNs were then concatenated and mapped to a linear activation layer to predict outdoor NO_2 concentrations. Data augmentation was used to improve the generalizability of the model by randomly flipping the aerial images during model training. Models were trained with the Nadam optimizer (using the default values for b1=0.9 and b2=0.999 as recommended by the original paper) (11) using root mean square error (RMSE) loss, and model performance was evaluated based on minimizing the RMSE of the predictions. These design choices were found to have the best performance in previous approaches to modelling outdoor air pollution using CNNs (12, 13).

Four key hyperparameters were identified that affected final model performance including initial model weights, image zoom levels (i.e. spatial scale), batch size, and learning rate. The initial model weight parameter describes the convolutional weights assigned at the start of model training. We considered two options for initial model weights: weights from a network pre-trained for image classification on ImageNet (14) and weights from the IMAGE-$PM_{2.5}$ model which was trained to estimate global variations in outdoor $PM_{2.5}$ concentrations using satellite images as input (12). The values of each hyperparameter are shown in Figure 3A. All hyperparameter combinations were used to train models for up to 500 epochs using a training dataset composed of the 1st through 10th folds, and a validation dataset composed of the 11th fold. These models were trained using an early-stopping callback to prematurely stop model training if the prediction accuracy did not improve for 50 epochs to prevent overfitting. The set of hyperparameters that achieved the lowest validation RMSE was retained and the number of epochs required to reach the minimum validation RMSE was noted. The lowest RMSE

of 3.30 μg/m^3 was achieved after training for 155 epochs using a combination of zoom level 16 and 19 images (spatial scales of approximately 1km^2 and 0.015km^2, respectively), with IMAGE-$PM_{2.5}$ weights along with a batch size of 8 and learning rate of 0.0001 (Figure 3A).

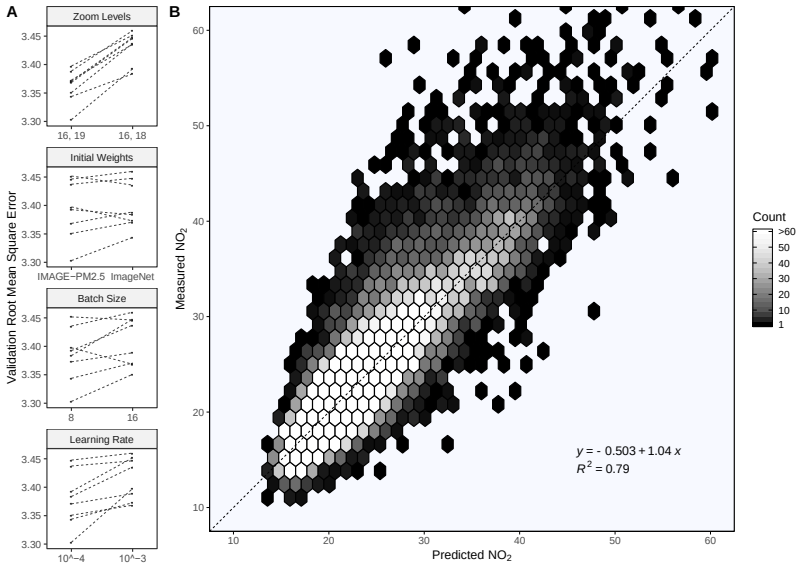

Figure 3. (A) Results of the hyperparameter optimization. Each set of connected dots represents the difference in performance on the validation set holding every other parameter constant. (B) Measured versus predicted annual average NO_2 concentrations (μg/m^3) using the best performing set of hyperparameters determined in A.

A 10-fold cross-validation procedure was used to evaluate model performance using the optimal set of hyperparameters listed above; the 11th fold was excluded from the cross-validation procedure as hyperparameters were optimized on this portion of the data. Each of the 10 models were trained for 155 epochs rather than using the early-stopping callback, as the model would be blind to the test set in practice and would not know when to optimally stop training to avoid overfitting. The RMSE of 10-fold cross-validation is reported to estimate the performance of the model. Gradient-weighted class activation maps (GradCAMs) were subsequently used to highlight portions of images that played an important role in making model predictions (15). All analyses were conducted using Keras (version 2.2.4) and TensorFlow

(1.13.1) in Python (16, 17). Using parallel training across four GPUs model training took on average 5-minutes per epoch.

3. Results and Discussion

Overall, the model provided reliable estimates of annual average outdoor NO_2 concentrations across the study area (slope=1.040, 95% CI: 1.033, 1.049) and explained the majority of spatial variations in annual average outdoor NO_2 (R^2=0.79, RMSE=3.58 μg/m^3) (Figure 3B). The spatial resolution of these predictions is approximately 0.015 km^2, as determined by the approximate area covered by the zoom level 19 images used to make predictions. For comparison, when applied using the same dataset, land use regression (R^2=0.77) and dispersion modelling (R^2=0.58) approaches explained a slightly lower proportion of spatial variations in annual average outdoor NO_2 across the study area (18). Similarly, a recent global land use regression model for ambient NO_2 reported an R^2 value of 0.57 (RMSE=4.5 μg/m^3) for Europe (4) which is substantially lower than the values reported for our model above.

Gradient-weighted class activation maps (GradCAMs) were used to highlight portions of images that played an important role in making model predictions (15). Figure 4 shows GradCAMs for images captured at monitoring locations in the lower (Figure 4A) and upper (Figure 4B) quintiles of annual average outdoor NO_2 concentrations in Flanders, Belgium. In these figures, the original image is shown in the column on the left and the GradCAM is displayed in the column on the right. Each row corresponds to a different quintile of annual average ambient NO_2 concentration. In Figure 4A, our model correctly classified the image as being in the lowest quintile of NO_2 concentration and the associated GradCAMs reveal a number

of interesting points with respect to what our model may be learning to make predictions. For example, in the top row of Figure 4A we see that the green areas around the roadway are highlighted but not the roadway itself. In addition, if we look at the bottom row of Figure 4A, we see that nothing is highlighted; this indicates that the model did not consider anything about this image to be particularly relevant to the highest quintile of exposure. Similarly, Figure 4B shows an image that was correctly classified as being in the highest category of outdoor NO_2 concentration. In this example, the streets and block configurations are highlighted in the bottom row suggesting that these features were important in classifying the image as being in the upper quintile of exposure. Alternatively, nothing is highlighted in the top row of Figure 3B suggesting that this image had few features relevant to classification in the lowest quintile. These visualizations are reassuring in that they are consistent with our understanding of sources of outdoor NO_2 as we would expect concentrations to be lower in green areas and higher in areas with more streets/traffic.

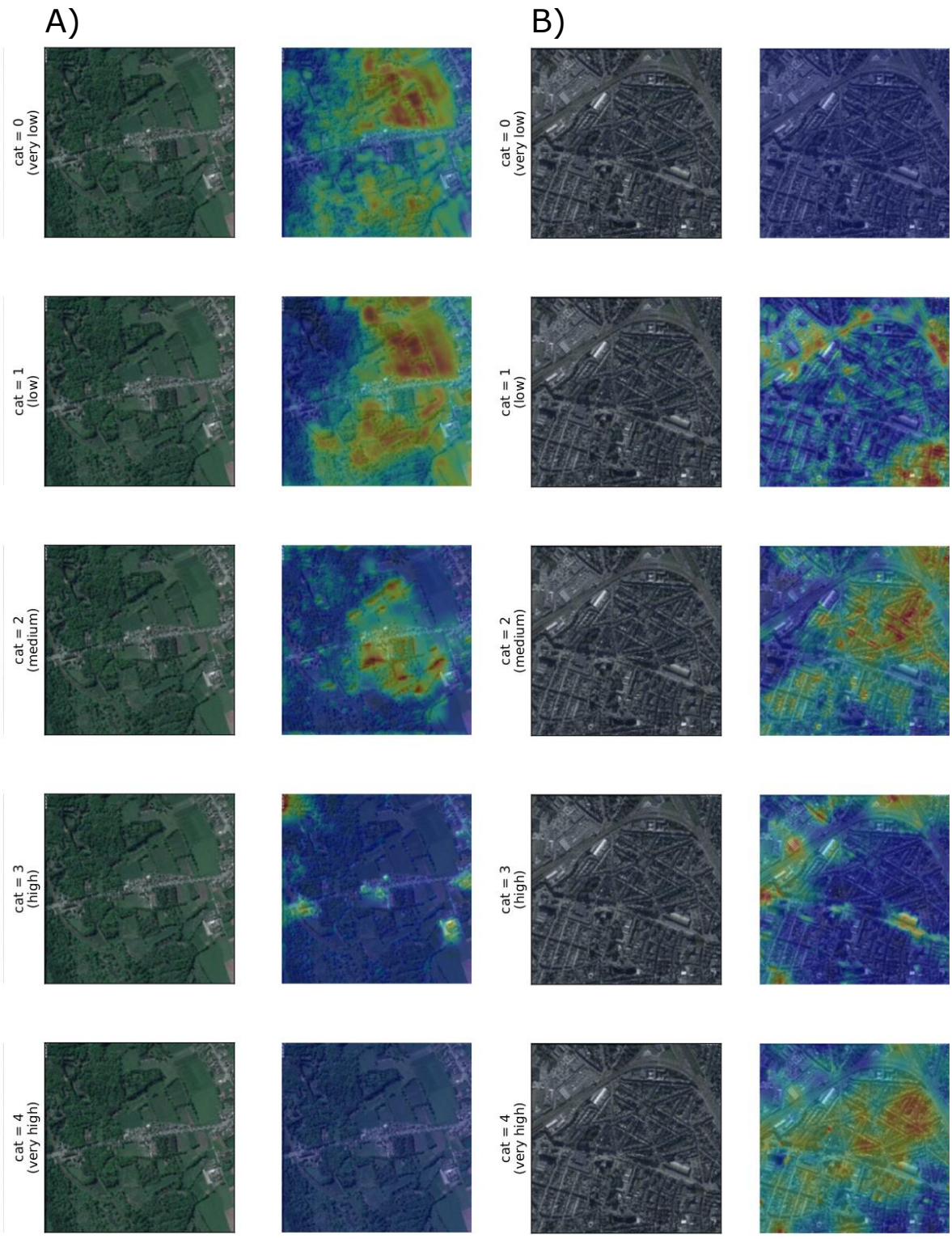

Figure 4. Aerial images and gradient-weighted class activation maps (GradCAMs) for sites in the lower (A) and upper (B) quintiles of annual average outdoor NO_2 concentrations in Flanders, Belgium.

Overall, our findings suggest that convolutional neural networks trained using ground level measurements and digital imagery can provide reliable estimates of annual average outdoor NO_2 concentrations without relying on typical model inputs including emissions data, chemical transport models, and/or geographic information systems (4, 19, 20). In general, the deep learning approach can be conceptualized as a modified version of traditional land use regression models whereby important combinations of land use features (captured in images) are learned automatically by the model to make predictions without having to rely on pre-specified geographic information system data or other data sources such as emissions data which are often missing in many locations around the world. Indeed, the combined use of aerial images and deep learning models has several advantages as aerial images are increasingly available on a global scale and thus model predictions are not limited to locations with access to suitable model inputs. While additional geolocated NO_2 data are needed to expand model coverage, particularly for locations with different physical appearances and distributions of outdoor NO_2, a transfer learning approach (i.e. starting model training using weights from our existing model) can be used to greatly reduce the number of samples required to facilitate this process (21). Moreover, as shown here, the process of ground-level data collection can be scaled efficiently in the context of citizen science projects. Importantly, the deep learning approach to model development is amenable to continuous improvement as new information can be added to update the model and expand generalizability across space and time as more data become available. Moreover, this approach could also facilitate the development of hybrid models whereby deep learning model predictions based on images are used as additional parameters in traditional land use regression modelling framework.

It is difficult to overstate the potential impact that this could have on the future of air pollution exposure science: given the right infrastructure and data-sharing mechanisms, every new study containing geolocated air pollution data could contribute to improving a single global model. The efficiencies gained through this approach could be enormous and the resulting models could have a transformative impact on our current understanding of spatial/temporal variations in outdoor air pollution concentrations in regions currently lacking basic data on environmental pollutants. While we cannot speculate with respect to the effectiveness of this approach for all air pollutants, we previously demonstrated that images can be used to predict spatial variations in long-term average outdoor PM2.5 (22) and ultrafine particle concentrations (23) and more recently short-term variations in UFPs number concentrations, UFP size, and noise (24). Therefore, it seems likely that images may provide useful information in predicting exposures for environmental pollutants with source information (or surrogate measures of sources like land use characteristics) that can be captured in images. Of course, deep learning models also have limitations, most notably related to interpretation as it is not always clear how predictions are made given the large number of parameters in these models (25). Convolutional neural networks are somewhat advantageous in this regard, and the GradCAMs employed in this study suggest that our NO_2 model is learning from features that are consistent with our understanding of spatial variations in sources of NO_2 including green space and traffic/street characteristics. This provides reassurance that the model is capturing meaningful information related to the presence/absences of known sources of NO_2.

4. Conclusions

In summary, we developed a new deep learning model to predict spatial variations in annual average outdoor NO_2 concentrations across Flanders, Belgium. This model relies only on digital images as input and predictive performance meets or exceeds current state-of-the-art approaches to modelling spatial variations in outdoor NO_2. A transfer learning approach can be used to efficiently expand model coverage in the future, continuously updating model weights as new data are collected. This procedure offers a novel means of expanding coverage to regions of the world missing detailed emissions data and/or land use information typically used to estimate outdoor air pollution concentrations. Predictions from these models may be used on their own or integrated into more complex ensemble models to leverage the overall strengths of different modelling approaches.

References:

1. GBD 2017 Risk Factors Collaborators. Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet* 2018; 392: 1923–1994.
2. Brauer M, Freedman G, Frostad J, van Donkelaar A, Martin RV, Dentener F, et al. Ambient air pollution exposure estimation for the global burden of disease 2013. *Environ Sci Technol* 2016; 50: 79-88.
3. Faustini A, Rapp R, Forastiere F. Nitrogen dioxide and mortality: review and meta-analysis of long-term studies. *Eur Respir J* 2014; 44: 744-53.
4. Larkin A, Geddes JA, Martin RV, Xiao Q, Liu Y, Marshall JD, Brauer M, Hystad P. Global land use regression model for nitrogen dioxide air pollution. *Environ Sci Technol* 2017; 51: 6957-6964.
5. Martin RV, Brauer M, van Donkelaar A, Shaddick G, Narain U, Dey S. No one knows which city has the highest concentration of fine particulate matter. *Atmos Environ* 2019; 3: 100040.
6. Irwin A. News Feature: Citizen science comes of age. *Nature* 2018; 562: 480-482.
7. Kahle D, Wickham, H. ggmap: Spatial visualization with ggplot2. The R Journal, 2013. URL https://journal.r-project.org/ archive/2013-1/kahle-wickham.pdf
8. R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.
9. Chollet F. Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition, 2017.
10. Kaiming H, Zhang X, Ren S, Sun J. "Deep residual learning for image recognition." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770-778. 2016.
11. Dozat T. Incorporating Nesterov Momentum into Adam. ICLR Workshop, 2016. URL https://openreview.net/pdf?id=OM0jvwB8jIp57ZJjtNEZ
12. Hong KY, Pinheiro PO, Weichenthal S. Learning global variations in outdoor $PM_{2.5}$ concentrations with satellite images. International Conference on Machine Learning AI for Social Good Workshop, Long Beach, United States, 2019. URL https://aiforsocialgood.github.io/icml2019/accepted/track1/pdfs/12_aisg_icml2019.pdf
13. Hong KY, Pinheiro PO, Minet L, Hatzopoulou M, Weichenthal S. Extending the spatial scale of land use regression models for ambient ultrafine particles using satellite images and deep convolutional neural networks. *Environmental Research* 2019; 176: 108513.

14. Krizhevsky A, Sutskever I, Hinton GE. "Imagenet classification with deep convolutional neural networks." In *Advances in neural information processing systems*, pp. 1097-1105. 2012.

15. Selvaraju R, Cogswell M, Das A, et al. Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. ArXiv, 1610.02391, 2017. Arxiv.org/abs/1610.02391v3

16. Chollet F. Keras: Deep learning library for theano and tensorflow. https://keras.io

17. Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, Corrado GS, Davis A, Dean J, Devin M, Ghemawat S, Goodfellow I, Harp A, Irving G, Isard M, Jozefowicz R, Jia Y, Kaiser L, Kudlur M, Levenberg J, Mané D, Schuster M, Monga R, Moore S, Murray D, Olah C, Shlens J, Steiner B, Sutskever I, Talwar K, Tucker P, Vanhoucke V, Vasudevan V, Viégas F, Vinyals O, Warden P, Wattenberg M, Wicke M, Yu Y, Zheng X. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.

18. Meysman FJR, De Craemer S, Lefebvre W, Vercauteren J, Sluydts V, Dons E, Hooyberghs H, Van den Dossche J, Fierens F, Huyse H. Large scale citizen science reveals the spatial structure of air pollution. EarthArXiv 2020. Doi:10.31223/osf.io/bryje

19. de Hoogh K, Gulliver J, van Donkelaar A, Martin RV, Marshall JD, Bechle MJ, Cesaroni G, Pradas MC, Dedele A, Eeftens M, Forsberg B, Galassi C, Heinrich J, Hoffmann B, Jacquemin B, Katsouyanni K, Korek M, Künzli N, Lindley SJ, Lepeule J, Meleux F, de Nazelle A, Nieuwenhuijsen M, Nystad W, Raaschou-Nielsen O, Peters A, Peuch V, Rouil L, Udvardy O, Slama R, Stempfelet M, Stephanou EG, Tsai MY, Yli-Tuomi T, Weinmayr G, Brunekreef B, Vienneau D, Hoek G. Development of West-European $PM_{2.5}$ and NO_2 land use regression models incorporating satellite-derived and chemical transport modelling data. *Environmental Research* 2016; 151: 1-10

20. Beelen R, Hoek G, Vienneau D, Eeftens M, Dimakopoulou K, Pedeli X, Tsai M, Künzli N, Schikowski T, Marcon A, Eriksen KT, Raaschou-Nielsen O, Stephanou E, Patelarou E, Lanki T, Yli-Tuomi T, Declercq C, Falq G, Stempfelet M, Birk M, Cyrys J, von Klot S, Nádor G, Varró MJ, Dedele A, Gražulevičiene R, Mölter A, Lindley S, Madsen C, Cesaroni G, Ranzi A, Badaloni C, Hoffmann B, Nonnemacher M, Krämer U, Kuhlbusch T, Cirach M, de Nazelle A, Nieuwenhuijsen M, Bellander T, Korek M, Olsson D, Strömgren M, Dons E, Jerrett M, Fischer P, Wang M, Brunekreef B, de Hoogh K. Development of NO_2 and NO_x land use regression models for estimating air pollution exposure in 36 study areas in Europe – the ESCAPE project. *Atmos Environ* 2013; 72: 10-23

21. Pan SJ, Yang Q. A Survey on Transfer Learning. IEEE Transactions on Knowledge and Data Engineering 2010; 22: 1345-1359

22. Hong KY, Pinheiro PO, Weichenthal S. Learning global variations in outdoor PM2.5 concentrations with satellite images. International Conference on Machine Learning AI for Social Good Workshop, Long Beach, United States, 2019a. URL https://aiforsocialgood.github.io/icml2019/accepted/track1/pdfs/12_aisg_icml2019.pdf

23. Hong KY, Pinheiro PO, Minet L, Hatzopoulou M, Weichenthal S. Extending the spatial scale of land use regression models for ambient ultrafine particles using satellite images and deep convolutional neural networks. Environmental Research 2019b; 176: 108513

24. Hong KY, Pinheiro PO, Weichenthal S. Predicting outdoor ultrafine particle number concentrations, particle size, and noise using street-level images and audio data. Environ Int 2020; 144: 106044.

25. Weichenthal S, Hatzopoulou M, Brauer M. A picture tells a thousand…exposures: opportunities and challenges of deep learning image analysis in exposure science and environmental epidemiology. *Environ Int* 2019; 122: 3-10