

# Towards an Open Global Air Quality Monitoring Platform to Assess Children's Exposure to Air Pollutants in the Light of COVID-19 Lockdowns

Christina Last\*  
christina@skwire.co.uk  
Skwire  
UK

Prithviraj Pramanik\*<sup>†</sup>  
prithvirajpramanik@yahoo.co.in  
NIT Durgapur, India & Discovery  
Partners Institute  
USA

Nikita Saini\*  
nikitasain61@gmail.com  
CitiGroup  
India

Akash Smaran Majety\*  
akashsmaran@gmail.com  
Humaine  
USA

Do-Hyung Kim  
Manuel García-Herranz  
dokim@unicef.org  
mgarciaherranz@unicef.org  
Office of Innovation, UNICEF  
USA

Subhabrata Majumdar  
subho@att.com  
AT&T Inc.  
USA

## ABSTRACT

This ongoing work attempts to understand and address the requirements of UNICEF, a leading organization working in children's welfare, where they aim to tackle the problem of air quality for children at a global level. We are motivated by the lack of a proper model to account for heavily fluctuating air quality levels across the world in the wake of the COVID-19 pandemic, leading to uncertainty among public health professionals on the exact levels of children's exposure to air pollutants. We create an initial model as per the agency's requirement to generate insights through a combination of virtual meetups and online presentations. Our research team comprised of UNICEF's researchers and a group of volunteer data scientists. The presentations were delivered to a number of scientists and domain experts from UNICEF and community champions working with open data. We highlight their feedback and possible avenues to develop this research further.

## CCS CONCEPTS

• **Applied computing** → **Environmental sciences**; • **Human-centered computing** → **Ubiquitous and mobile computing**; **Systems and tools for interaction design**.

## KEYWORDS

PM2.5, Global Model, Air Quality Monitoring, Air Pollution, COVID-19

\*Equal Contribution

<sup>†</sup>Corresponding Author

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

*CHI '21 Extended Abstracts, May 8–13, 2021, Yokohama, Japan*

© 2021 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-1-4503-8095-9/21/05...\$15.00

<https://doi.org/10.1145/3411763.3451768>

## ACM Reference Format:

Christina Last, Prithviraj Pramanik, Nikita Saini, Akash Smaran Majety, Do-Hyung Kim, Manuel Garcia-Herranz, and Subhabrata Majumdar. 2021. Towards an Open Global Air Quality Monitoring Platform to Assess Children's Exposure to Air Pollutants in the Light of COVID-19 Lockdowns. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts (CHI '21 Extended Abstracts)*, May 8–13, 2021, Yokohama, Japan. ACM, New York, NY, USA, 7 pages. <https://doi.org/10.1145/3411763.3451768>

## 1 INTRODUCTION

There is a strong link between human health and exposure to high levels of air pollution [9, 11]. Deteriorating air quality has thus been a growing concern over the last decade [3, 24], in particular within urban agglomerations [4, 18]. In order to monitor the temporal and spatial variability in air pollution across regions where ground sensors are sparse, there is a need to develop data-driven approaches to predicting air pollution levels. Several such prediction methods use historic and current data from existing air quality monitoring stations [27]. Despite this progress, there is a need to advance current modelling approaches to account for the variation introduced from COVID-19 lockdown events at a global scale, with particular focus on modelling air pollution concentration in regions with sparse coverage of ground-based air quality monitoring sensors.

Long-term exposure to fine particulate matter (especially, PM2.5, i.e. particles with a diameter less than  $2.5\mu\text{m}$ ) are estimated to cause ~8 million excess deaths annually. The impact of air pollution is felt more acutely by the young, with 1 in every 4 deaths under 5 years is directly or indirectly related to environmental risks [25]. Globally, 93% of children live in places where air pollution levels exceed World Health Organization (WHO) guidelines [23], defined as PM2.5 values above  $10\mu\text{g}/\text{m}^3$  annual mean and  $25\mu\text{g}/\text{m}^3$  24-hour mean [21]. Therefore, there is an increasing need to develop a model that explains temporal and spatial variability in air pollution concentrations to accurately assess the global children populations' exposure to high levels of air pollution.

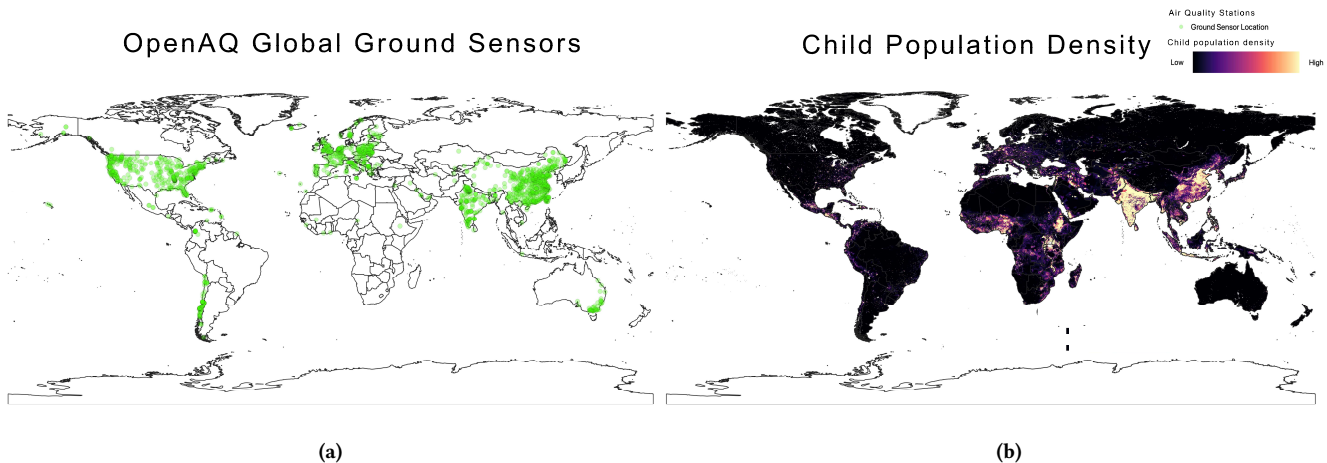


Figure 1: Global distribution of (a) Ground Level Air Quality Sensors, and (b) Child Population, Data source: Rees et al. [17].

## 1.1 Motivation

As seen in Figure 1a, ground-based air quality sensors are distributed worldwide heterogeneously, with particularly low concentration of sensors in West and East Africa. A comparison with the distribution of child population across the world (Figure 1b) indicates that it is difficult to assess children’s exposure to high levels of air pollution in areas with scarce ground-based observations, such as in highly populous locations like Western Africa and the Great Rift Valley. Responding to this challenge, UNICEF<sup>1</sup> and Solve for Good<sup>2</sup>—a non-profit foundation connecting philanthropic organizations with data science expertise—aim to build an open global model for air quality prediction.

Children are likely to have more acute respiratory illnesses than adults when exposed to high levels of air pollution, often with long-term consequences [2, 28]. Moreover, the detrimental effects of COVID-19 are compounded by poor air quality. COVID-19 is a lung disorder, to which air pollutants have been a major contributor [5, 7]. Motivated by such concerns—which are in tune with UNICEF’s focus on child health—the objectives of this collaborative research project are to:

- (1) Build a global model that augments accurate but limited air pollution data collected using ground sensors with machine learning (ML) predictions using global-level but coarse satellite data as input features.
- (2) Develop regional, “fine tuned” versions of the model to address inaccuracies within a global model to understand the variability of air pollution patterns across regions, specially with sparse ground sensors.
- (3) Measure children’s exposure to high levels of air pollution, in particular where children are exposed to pollution concentrations exceeding the WHO pollution standards during pre- and post-COVID-19 lockdown events.

- (4) Educate UNICEF’s regional offices, partner organisations and citizens on the benefits of data-driven air pollution modelling and use their feedback to improve our system.

## 1.2 Project Workflow

In the first stages of the work, UNICEF approached Solve for Good with challenge—that it is difficult for the UNICEF’s regional offices, and local partner organisations to monitor air pollution at a local level, resulting in insufficient evidence to justify investing in air quality improvement projects. (Figure 2: Stage 1 - Engagement).

After discussing different aspects of this challenge, the project team scoped a solution to fit the time and resources available (Figure 2: Stage 2 - Scoping). The project scope focused selecting a research question that responded to the challenge, and leveraged the data science expertise in the Solve for Good team of volunteers.

Following the scoping stage, UNICEF worked in collaboration with the Solve for Good researchers to develop a scientific methodology outlining the stages of data collection, pre-processing, analysis and visualisation (Figure 2: Stage 3 - Methodology). Key datasets were identified in this stage, such as satellite data and ground-based measurements of air quality (Section 2), and data such as the global child population and school locations.

During the development of the methodology, the project team integrated weekly feedback sessions to report on the project analysis (Figure 2 Stage 4: Feedback), giving UNICEF officials the opportunity to inform the research efforts with feedback from regional partners, as well as tailoring to the research needs of UNICEF right from the beginning.

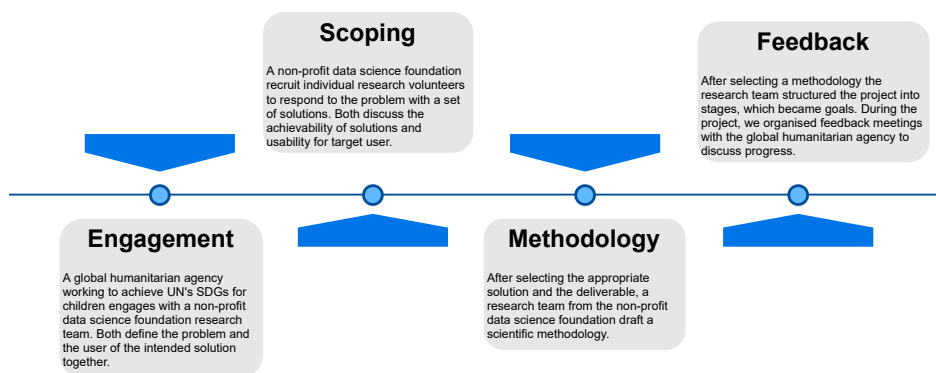
## 2 DATA

In the interest of using open data, we used the data provided by OpenAQ<sup>3</sup>, and collected global air quality data from air quality monitoring stations across the globe (sample size  $n=1601$ ), measuring ground-level PM<sub>2.5</sub> concentrations from January 2019 to September 2020. These ground-level values would act as our output feature

<sup>1</sup><https://www.unicef.org/>

<sup>2</sup><https://www.solveforgood.org/>

<sup>3</sup><https://openaq.org>



**Figure 2: The project process 1. Identifying the problem, 2. Scoping a potential solution, 3. Developing a scientific methodology, and 4. Receiving feedback from project partner (UNICEF) and update project process.**

(target variable), which we predicted using a machine learning (ML) model trained on weekly averaged value for AOD, NO<sub>2</sub>, Population Density and Precipitation (acquired from the Google Earth Engine<sup>4</sup>) as input features (predictor variables). We collected the input features for the time frame 1st January 2019 to 11th September 2020 from the satellites given in Table 1.

Since the raw satellite data were collected at different temporal and spatial granularities (Table 1), we first standardized the data temporally and spatially to obtain our input features. The PM<sub>2.5</sub> values from the OpenAQ data had a mean (M) of 22.21  $\mu\text{g}/\text{m}^3$  with a standard deviation (SD) of 29.85. The data distribution had half of the values below 12.66  $\mu\text{g}/\text{m}^3$  and three fourths of the values under 27.74  $\mu\text{g}/\text{m}^3$ . For output feature values, only PM<sub>2.5</sub> values between 0 and 3000 were considered, and resampled on a weekly basis<sup>5</sup> to calculate the weekly average PM<sub>2.5</sub> concentration.

Along with the satellite and PM<sub>2.5</sub> measurement we considered the city wide lock-down information (Footnote: 12) from Ox-CGRT/ACAPS Lockdown Data (Footnote: 11, 10) for 192 countries, but removed them later from the final feature set because of location heterogeneity issues (Further discussion in Section 4).

We generated a VRT driver<sup>6</sup> to enable faster processing of the vast amount of spatial data. We obtained weekly averages of the features at two different geographic levels—local and city-wide. At the local level, we derived the values using a local mask defined as a 75m buffer around each ground sensor location to capture intersecting pixel edges, and take their average. At the city level, we extracted the city-wide weekly averages using corresponding city masks. The final pre-processing step was to aggregate the sensor point data (along with latitude and longitude upto 4 decimal points) with the city and country identifiers. In Figure 4a, we summarize the full data preparation process.

<sup>4</sup><https://earthengine.google.com>

<sup>5</sup>Considering Monday as the starting day of the week

<sup>6</sup>A format driver for GDAL that allows a virtual GDAL dataset to be composed from the other GDAL datasets, <https://gdal.org/drivers/raster/vrt.html>

## 3 METHODS

### 3.1 Model

To achieve the first objective (Section 1.1), we developed a stacked ensemble Machine Learning model (Figure 4b). The model enabled us to leverage the capabilities of a range of well-performing models to improve the accuracy of PM<sub>2.5</sub> predictions over any individual model. We used LightGBM, XGBoost and Random Forest as the base models to generate predictions of weekly averaged PM<sub>2.5</sub> concentrations. Base models were trained on 80% of the OpenAQ PM<sub>2.5</sub> concentration data. Predictions from base models were aggregated to be used as the input features for the meta-model. We used linear regression with a 5-fold cross-validation to train the meta-model.

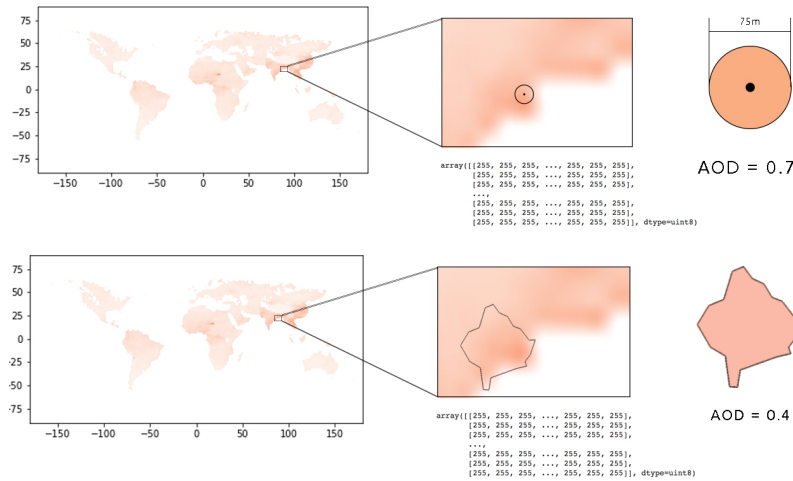
To reduce the generalization errors of the individual learners, we used a stacking regressor (stacked base and meta models). Feature importances of the base learners showed high variability, with AOD being the most important feature in XGBoost and Random Forest, LightGBM on the other hand had average population density as the most important feature. An ensemble approach reduced this variance making the model more robust and stable in performance.

To validate the model at global level, we generated 10<sup>6</sup> land-based locations [10], randomly distributed to ensure coverage of areas with little ground-based observation data of PM<sub>2.5</sub> concentrations. We extracted satellite data using a VRT driver to generate features for these locations to predict the weekly average PM<sub>2.5</sub> concentrations using our pre-trained ensemble model. Overall, our model achieved a median absolute error of 5.60  $\mu\text{g}/\text{m}^3$ , with Root Mean Squared Error (RMSE) of 22.8 and  $R^2 = 0.49$ . On a more granular level, the model performed best in Myanmar (RMSE = 9.9,  $R^2 = 0.8$ ) and worst in Ghana, attributed to the lack of training data. The model performed fairly well at locations in Guatemala, Iraq, Kyrgyzstan and Chile, with  $R^2 > 0.45$ , and RMSE values in the range of 6-14.

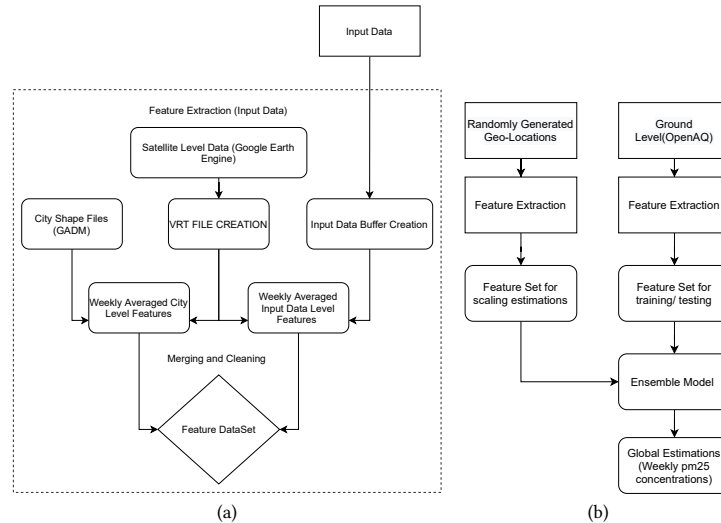
As an example, Figure 5 describes the spatial variability of PM<sub>2.5</sub> predictions the week beginning 20<sup>th</sup> March 2020 in two locations in Vietnam. Figure 5 (a) air pollution in Hà Nội, with weekly average PM<sub>2.5</sub> concentrations (M = 15.25  $\mu\text{g}/\text{m}^3$ , SD =  $\pm 10.52$ ) in urban setting generates relatively high air pollution levels in comparison

| Data Name          | Satellite Name  | Spatial Resolution | Temporal Resolution | Description  |
|--------------------|---|--------------------|---------------------|--|
| Precipitation      | GSMaP Operational: Global Satellite Mapping of Precipitation (Precipitation)          | 0.1 Arc Degrees    | Daily               | Snapshot of hourly precipitation rate, source: GSMaP Operational   |
| AOD                | MCD19A2.006: Terra & Aqua MAIAC Land Aerosol Optical Depth Daily 1 Km (AOD)           | 1000 Meters        | Daily               | Blue band (0.47 $\mu\text{m}$ ) aerosol optical depth over land, source: MCD19A2.006   |
| NO <sub>2</sub>    | Sentinel-5P NRTI NO <sub>2</sub> : Near Real-Time Nitrogen Dioxide (NO <sub>2</sub> ) | 0.01 Arc Degrees   | Daily               | Total vertical column of NO <sub>2</sub> = ratio of NO <sub>2</sub> slant column density and total air mass factor, source: Sentinel-5P NRTI NO <sub>2</sub> |
| Population Density | GPWv411: Population Density Gridded Population of the World (Population Density)      | 30 Arc Seconds     | Daily               | The estimated number of persons per square kilometer, source: GPWv411  |

**Table 1: Different satellite data sources, along with brief descriptions.**



**Figure 3: Top panel illustrates the "local" extraction of AOD from a 75m buffer around a point. Bottom panel illustrates the "city-level" extraction of AOD for each city globally. AOD = Aerosol Optical Depth is a well-known proxy for PM2.5 [12].**



**Figure 4: Flow diagrams for (a) Data Preparation, and (b) Modelling Process.**

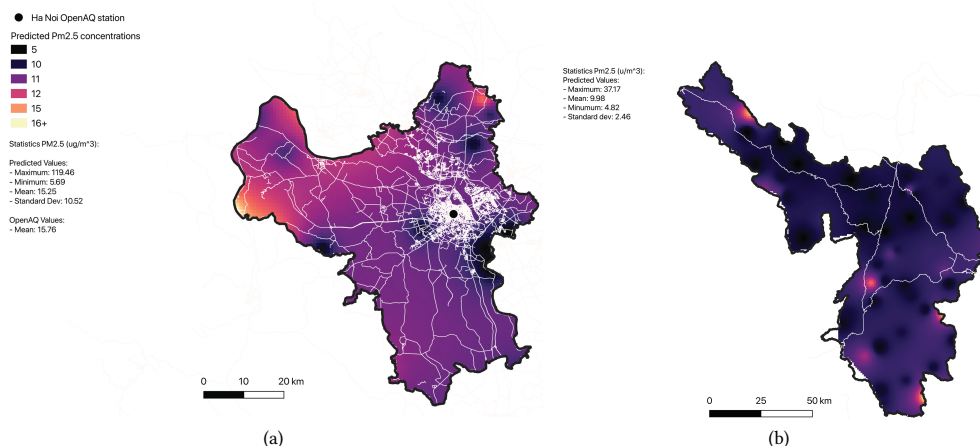


Figure 5: Comparative analysis of (a) Hà Nội (an urbanised location in northern Vietnam) and (b) Điện Biên (a rural region in northern Vietnam) and their predicted PM2.5 concentrations

to the region of Điện Biên 5 (b) ( $M = 9.95 \mu\text{g}/\text{m}^3$   $SD = \pm 2.49$ ) where the concentrations are much lower.

### 3.2 Stakeholder involvement

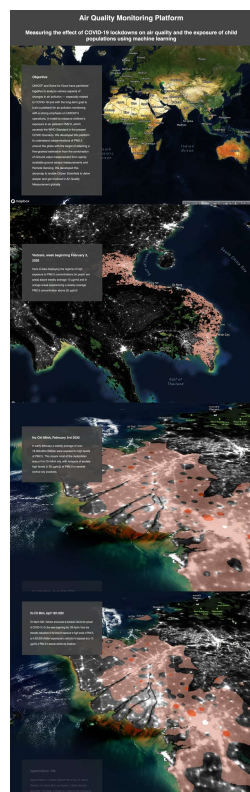


Figure 6: Mapbox story map showing the results of the analysis.

In the first phase of the project process (Engagement, Figure 2), our team consulted with 4 members from UNICEF—1 climate and energy specialist, and 3 data scientists—to sketch out UNICEF’s requirements and a timeline for the project. Two of the partner data scientists were involved in the second and third stages (Scoring and Methodology, respectively, in Figure 2). UNICEF’s ongoing involvement served two purposes: first, their domain knowledge helped data gathering, processing, and model building. Secondly, their continuous feedback during the weekly status meetings ensured that the final deliverable (global model and its outputs) was tailored to UNICEF’s needs.

In the expert feedback and Q&A session, we presented the methodology and results to 21 scientists and domain experts from UNICEF, and one open data and public policy expert from Solve for Good. To communicate the results of the research to UNICEF’s regional offices and citizens (satisfying objective 4), we developed a story map (Figure 6) using Mapbox, demonstrating the

changes in predicted PM2.5 concentrations before and after COVID-19 lockdown events globally. The story map contains layers showing the child population density, and overlays polygons indicating geographic regions with PM2.5 concentrations above the WHO recommended limits (PM2.5 exposure of  $10 \mu\text{g}/\text{m}^3$  annual mean or  $25 \mu\text{g}/\text{m}^3$  24-hour mean). Figure 6 indicates the reduction in exposure to high PM2.5 concentrations by comparing a pre-lockdown weekly average in early February 2020 to a week in mid-April 2020 for in Ho Chi Minh City, Vietnam. When presenting this story map in the Q&A session, we used a between-subject study design to collate themes that emerged from the feedback received on implementation of the results.

### 4 EXPERT FEEDBACK

A number of common themes emerged from the ensuing discussion and Q&A on how to help translate the insights from our model into practical local-level policies.

*Data scarcity:* The model results indicate that scarce air quality monitoring stations (Figure 1a) reduce model accuracy, as predictions are less reliable at locations with no proximate air quality monitoring stations, or areas with high spatial or temporal variations of pollutant levels—often due to seasonal local factors such as festivals [14] or agricultural practices [8]. In the absence of air quality monitoring stations in close proximity, ground or near ground-level air quality data can be obtained by collaborating with local corporations or nonprofits. For example, in line with objective 2 (to develop a regional, “fine tuned” versions of the model). In response to this challenge, a member of the expert feedback group suggested UNICEF’s global field presence can leveraged to obtain more granular non-regulatory grade sensor, and Plume Labs<sup>7</sup> agreed to share high-quality sensor data on PM2.5 concentrations across Lima, Peru.

*Links to COVID-19 infection rates:* Through a discussion with UNICEF’s health experts, one significant suggestion to account for

<sup>7</sup><https://plumelabs.com>



the association between COVID-19 incidences and poor air quality was to systematically correlate and compare administrative data from public health systems with levels of air quality in a geographic area pre- and post-COVID-19. This would help distinguish child cases of air pollution related respiratory conditions from COVID-19 cases. As we work towards a broader understanding of child respiratory health metrics and their connection to air quality in the wake of COVID-19, obtaining region-wise risk classifications based on observational public health data can help us translate local air quality predictions into concrete policy interventions.

*Local heterogeneity of data and transparent predictions:* Explainable methods should drive transparency in models that use features from local data and predictions from a global model such as ours (demonstration video<sup>8</sup>). Based on user needs, such explanations could use simple parametric models (such as linear regression), or techniques with post-hoc explanations [13].

*Technical feedback:* A prominent point brought forward by a open data and public policy expert from the non-profit foundation was the difficulty in validating PM2.5 measurement in regions which are completely missing from the training set. In response to this feedback, we aim to create training and validation datasets to mimic local air quality patterns in regions that have sparse observational data. To this end, encouraging the participation of more regional offices could provide the technical team with local data sources. Additionally, instead of splitting the data randomly by location to obtain train/validation sets (current approach), a stratified sampling of locations can be undertaken—with sampling weights for a location as a function of the number of air quality monitoring stations within a certain radius. This can potentially diversify the test regions under consideration and improve model prediction quality in data-poor regions. Mapping the global model predictions to standard air quality classifications can have more impact on downstream tasks like exposure measurement and public health analytics. For example, rather than considering raw PM2.5 predictions from our meta-regression model, one can classify them using the Air Quality Index look up table<sup>9</sup>, to indicate the severity of exposure over a fixed period of time. To evaluate this approach, the accuracy of this classification can be assessed by comparing with classifications of actual PM2.5 values using F1 Scores.

*Auxiliary data sources:* A major motivation of this project was to understand how lockdown policies in different countries have affected the exposure of children to air pollutants like PM2.5 and NO<sub>2</sub>. In response to this objective, future versions of our model could benefit from the incorporation of lockdown-related input features. We originally intended to use features from the ACAPS COVID-19 Government Measures Dataset<sup>10</sup> along with feature rich Oxford COVID-19 Government Response Tracker (OxCGRT)<sup>11</sup>, but due to data heterogeneity and the absence of location-specific measures across different locations, we removed them from the final model. Some examples of such features are related to lockdown

dates and severity, social distancing, movement restrictions, relevant social, economic or public health measures, and the COVID-19 Government Response Stringency Index<sup>12</sup>.

To account for local variability within the global model predictions, we intend to incorporate land cover [15, 20, 26], and meteorological variables such as temperature, wind speed, wind direction, and cloud cover [1, 22]. We can capture these information using publicly available data sources such as the Copernicus yearly land cover classification [19] and observational meteorological data collected by the United States Environmental Protection Agency (US EPA)<sup>13</sup>.

The effect of population on children’s respiratory health is well-documented. For example, in cities with populations more than 10 million, children are exposed to air pollution levels that are 2 – 8 times higher than WHO-acceptable limits [6]. Through UNICEF’s ties with local governments, we have obtained school locations in Niger, Sierra Leone, and Colombia and global child population data. We intend to augment our model with input features from these data [16] to prioritize locations with higher child population and high predicted PM2.5 levels for post-COVID-19 monitoring. Another potential contributor to children’s exposure to high levels of air pollution is *intra-day* variability. Alternative methods of feature aggregation may be considered such as aggregating PM2.5 values using regional travel behaviour patterns to account for daytime (such as rush-hour periods) and nighttime discrepancy in pollutant levels, as well as the variation at school locations and residential areas.

## 5 CONCLUSION

Due to the lack of ground-based sensors and the heterogeneity of sensor distribution around the world, there is a need to develop a global model of air pollution to understand the effects of the COVID-19 lockdowns. Through this work, we developed an initial model along with the expert comments from a multi-disciplinary panel, and charted a roadmap for progress towards the broader goal of developing locally-tuned models through iterative feedback from relevant stakeholders. By presenting our motivation and findings to the CHI community, we wish to draw their attention to the issue of equitable global air quality monitoring in the post-COVID-19 world, and highlight the role of collaborative research teams using multi-disciplinary methods to develop this research further.

## ACKNOWLEDGMENTS

This work is an outcome of the research and development undertaken through the Solve For Good platform with UNICEF as the primary stakeholder. The authors wish to thank UNICEF Office of Innovation for the computational resources and funding support to present the work. We thank qAira & PlumeLabs for the regional air quality data in Peru. Additionally, one of the authors (PP) is supported by Visvesvaraya PhD Scheme of Ministry of Electronics & Information Technology, Government of India, being implemented by Digital India Corporation and Fulbright Fellowship.

<sup>8</sup><https://youtu.be/2JVDRpLRX2k>

<sup>9</sup><https://www.epa.gov/wildfire-smoke-course/wildfire-smoke-and-your-patients-health-air-quality-index>

<sup>10</sup><https://www.acaps.org/covid-19-government-measures-dataset>

<sup>11</sup><https://github.com/OxCGRT/covid-policy-tracker>

<sup>12</sup><https://ourworldindata.org/grapher/covid-stringency-index>

<sup>13</sup><https://www.epa.gov/scram/meteorological-data-and-processors>

## REFERENCES

- [1] Christine Arciszewska and John McClatchey. 2001. The importance of meteorological data for modelling air pollution using ADMS-Urban. *Meteorol. Appl.* 8 (2001), 345–350.
- [2] American Lung Association. 2020. *Children and Air Pollution*. <https://www.lung.org/clean-air/outdoors/who-is-at-risk/children-and-air-pollution>
- [3] Bert Brunekreef and Stephen T Holgate. 2002. Air pollution and health. *The lancet* 360, 9341 (2002), 1233–1242.
- [4] Aaron J Cohen, H Ross Anderson, Bart Ostro, K Dev Pandey, Michal Krzyzanowski, Nino Künzli, Kersten Gutschmidt, C Arden Pope III, Isabelle Romieu, Jonathan M Samet, et al. 2004. Urban air pollution. *Comparative quantification of health risks: global and regional burden of disease attributable to selected major risk factors* 2 (2004), 1353–1433.
- [5] Eric S Coker, Laura Cavalli, Enrico Fabrizi, Gianni Guastella, Enrico Lippo, Maria Laura Parisi, Nicola Pontarollo, Massimiliano Rizzati, Alessandro Varacca, and Sergio Vergalli. 2020. The effects of air pollution on COVID-19 related mortality in northern Italy. *Environmental and Resource Economics* 76, 4 (2020), 611–634.
- [6] Devra L. Davis and Paulo H.N. Saldiva. 1999. *Urban Air Pollution Risks to Children: A Global Environmental Health Indicator*. Technical Report. World Resources Institute. Available at <https://www.wri.org/publication/urban-air-pollution-risks-children>.
- [7] Ingo E Ispording and Nico Pestel. 2020. Pandemic meets pollution: poor air quality increases deaths by COVID-19. (2020).
- [8] Hiren Jethva, Omar Torres, Robert D. Field, et al. 2019. Connecting Crop Productivity, Residue Fires, and Air Quality over Northern India. *Scientific Reports* 9, 16594 (2019).
- [9] Marilena Kampa and Elias Castanas. 2008. Human health effects of air pollution. *Environmental Pollution* 151, 2 (2008), 362 – 367. <https://doi.org/10.1016/j.envpol.2007.06.012> Proceedings of the 4th International Workshop on Biomonitoring of Atmospheric Pollution (With Emphasis on Trace Elements).
- [10] T. Karin. 2020. *toddkarin/global-land-mask: Release of version 1.0.0*. <https://doi.org/10.5281/zenodo.4066722> Available at <https://doi.org/10.5281/zenodo.4066722>.
- [11] Janice J Kim et al. 2004. Ambient air pollution: health hazards to children. *Pediatrics* 114, 6 (2004), 1699–1707.
- [12] N. Kumar, A. Chu, and F. Foster. 2011. An empirical relationship between PM2.5 and aerosol optical depth in Delhi Metropolitan. *Atmos Environ* 41, 21 (2011), 4492–4503.
- [13] W. James Murdoch, Chandan Singh, Karl Kumbier, et al. 2019. Definitions, methods, and applications in interpretable machine learning. *Proceedings of the National Academy of Sciences* 116, 44 (2019), 22071–22080.
- [14] Neha Parkhi, Dilip Chate, Sachin D. Ghude, et al. 2016. Large inter annual variation in air quality during the annual festival 'Diwali' in an Indian megacity. *Journal of Environmental Sciences* 43 (2016), 265 – 272.
- [15] Wallace E. Reed and John E. Lewis. 1978. *Land Use and Land Cover Information and Air-Quality Planning*. Technical Report 1099-B. Geological Survey. Available at <https://pubs.usgs.gov/pp/1099b/report.pdf>.
- [16] N. Rees. 2017. *Danger in the air: How air pollution may be affecting the brain development of young children around the world*. Technical Report. Data, Research and Policy Working Paper. Available at [https://www.unicef.org/environment/files/Danger\\_in\\_the\\_Air.pdf](https://www.unicef.org/environment/files/Danger_in_the_Air.pdf).
- [17] Nicholas Rees, Amy Wickham, Yoonie Choi, et al. 2019. *Silent suffocation in Africa: Air pollution is a growing menace*. Technical Report. UNICEF. Available at <https://www.unicef.org/reports/silent-suffocation-in-africa-air-pollution-2019>.
- [18] Dietrich Schwela. 2000. Air pollution and health in urban areas. *Reviews on environmental health* 15, 1-2 (2000), 13–42.
- [19] Copernicus Climate Data Store. 2019. *Land cover classification gridded maps from 1992 to present derived from satellite observations*. <https://cds.climate.copernicus.eu/cdsapp#!/dataset/satellite-land-cover?tab=overview>
- [20] L. Sun, J. Wei, D.H. Duan, et al. 2016. Impact of Land-Use and Land-Cover Change on urban air quality in representative cities of China. *Journal of Atmospheric and Solar-Terrestrial Physics* 142 (2016), 43–54.
- [21] Zander S. Venter, Kristin Aunan, Sourangsu Chowdhury, and Jos Lelieveld. 2020. COVID-19 lockdowns cause global air pollution declines. *Proceedings of the National Academy of Sciences* 117, 32 (2020), 18984–18990.
- [22] Claudia Vitolo, Marco Scutari, Mohamed Ghalaieny, et al. 2018. Modeling Air Pollution, Climate, and Health Data Using Bayesian Networks: A Case Study of the English Regions. *Earth and Space Science* 5, 4 (2018), 76–88.
- [23] WHO. 2018. *Ambient (outdoor) air pollution*. [https://www.who.int/news-room/fact-sheets/detail/ambient-\(outdoor\)-air-quality-and-health](https://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health)
- [24] World Health Organization. 2016. *Ambient air pollution: A global assessment of exposure and burden of disease*. Technical Report ISBN: 9789241511353. WHO. Available at <https://www.who.int/publications/m/item/ambient-air-pollution-a-global-assessment-of-exposure-and-burden-of-disease>
- [25] World Health Organization. 2018. *Air pollution and child health: prescribing clean air*. Technical Report WHO/CED/PHE/18.01. WHO. Available at <https://www.who.int/ceh/publications/air-pollution-child-health/en>.
- [26] G. Xian. 2007. Analysis of impacts of urban land use and land cover on air quality in the Las Vegas region using remote sensing information and ground observations. *International Journal of Remote Sensing* 28, 24 (2007), 5427–5445.
- [27] Yu Zheng, Furui Liu, and Hsun-Ping Hsieh. 2013. U-Air: When Urban Air Quality Inference Meets Big Data. In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (Chicago, Illinois, USA) (KDD '13). Association for Computing Machinery, New York, NY, USA, 1436–1444. <https://doi.org/10.1145/2487575.2488188>
- [28] J. Zvobgo. 2018. *Children are 37% more exposed to air pollution than adults*. <https://www.childinthecity.org/2018/10/19/children-are-37-more-exposed-to-air-pollution-than-adults>