Air Quality Forecasting for Africa

Principal Investigator

Name: R Subramanian
Agency/institution: Kigali Collaborative Research Center (KCRC); OSU-EFLUVE, CNRS, France;
Mailing address: Kigali Collaborative Research Center
4th Floor, Telecom House
Boulevard de l'Umuganda
Kacyiru-Kigali
Rwanda
Email: info@kcrc.rw
subu@cmu.edu

Collaborators

Name: Carl Malings (co-principal investigator)
Agency/institution: OSU-EFLUVE, CNRS, France
Mailing address: Université Paris-Est Créteil Val de Marne (UPEC)
Faculté de sciences et technologie
Bâtiment P, P2 Niveau 2
61, avenue du Général de Gaulle
94010 Créteil Cedex
France
Email: Carl.Malings@lisa.u-pec.fr
Phone: +1-516-864-9696

Name: Paulina Jaramillo
Agency/institution: Carnegie Mellon University; Kigali Collaborative Research Center (KCRC)
Email: pjaramil@andrew.cmu.edu

Name: Tim Brown
Agency/institution: Carnegie Mellon University; Kigali Collaborative Research Center (KCRC)
Email: timxb@andrew.cmu.edu
Phone: +2500784840

Name: James Chege
Agency/institution: Code for Africa
Mailing address: 84 Shortmarket St.
3rd Floor
Cape Town 8000
South Africa
Email: chege@codeforafrica.org
Phone: +254721203802

Executive Summary

Air quality has been identified as the single largest environmental risk factor to human health. Poor air quality is estimated to be responsible for at least seven million premature deaths annually. This issue is especially critical in developing nations in Africa, which have some of the highest estimated levels of ambient pollution, but the poorest infrastructure in place for monitoring and tracking air quality. The health and economic impacts of poor air quality in Africa, exacerbated by the lack of monitoring and mitigation infrastructure, are only projected to get worse in the future as further industrialization and climate change contribute to diminishing air quality globally, and for Sub-Saharan Africa in particular.

Traditional air quality measurements are made by highly accurate but expensive instruments measuring the various atmospheric gases and particles that contribute to poor air quality. It is these traditional data sources that are especially scarce for Africa. Several additional sources of air quality information exist to supplement these, however. Low-cost sensor systems, including electrochemical gas sensors and optical particulate matter detectors, represent an emerging technology with the potential to greatly expand access to in situ air quality measurements for national governments as well as citizens, scientists, and community groups with an interest in air quality monitoring. Satellite data provides the possibility of remotely measured air quality, covering a large fraction of the globe at relatively little marginal cost to end users once the systems have been deployed. Computational air quality models also seek to cover wide spatial domains, with more accurate ground level concentration estimates, but requiring a commensurate increase in input data and computational time. Each of these information sources have their own pros and cons, but there exists the opportunity to combine these multisource data and create a comprehensive picture of air quality, where the benefits of certain data sources can cover for the drawbacks of others.

This project proposes the development of a system for assimilating air quality data from disparate sources using Bayesian learning and inference methods. In doing so, it will create a multisource database for timely local information on ambient air quality, allowing scientists, policy makers, and the general public to have access to the information they need to better understand and respond to their air quality situation. This project is envisioned in three broad phases. The first phase, or the technical phase, involves the development of the several components necessary for the proposed system. Broadly, this includes capacities to interface with existing free online sources of air quality data (e.g. from traditional measurement stations, low-cost sensor, satellite data, and computational simulations), to integrate these data into a unified database, to perform Bayesian learning and appropriately quantify the uncertainties of different data sources, to perform Bayesian inference for spatial interpolation and temporal forecasting of air quality using these multisource data, and to present and visualize the resulting information. The second phase, or implementation phase, involves combining these components into a coherent system which allows users to access these data online, especially through applications optimized for smartphone use. The third phase, or usage phase, refers to the deployment of this system, including a pilot deployment which will focus on a relatively small geographical area, such as the city of Kigali, Rwanda, Nairobi, Kenya, and Lagos, Nigeria and seek feedback from stakeholders including local community groups, educators, scientists, and policy makers.

The primary end goal of this project is the creation of a system which, in a semiautonomous manner, will gather recent air quality information from a number of freely available sources, integrate these data in such a way that respects the relevant uncertainties associated with each source, and provide users with comprehensive air quality information covering historical trends as well as the current situation and short term (several hours or days ahead) forecasts based on a statistical interpretation of the available information. In all these cases, accurate quantification and presentation of the uncertainties in both measurements and estimates will be of foremost importance. Development of this project will involve the participation of computer scientists and programmers, air quality researchers and data technicians, social scientists, and local experts and community stakeholders. To achieve this, the project will draw extensively on local expertise, especially students, scientists, and developers from the Kigali Collaborative Research Center, the University of Rwanda College of Science and Technology, CMU-Africa, and the sensors.AFRICA project incubated by Code for Africa. In particular, the project will work with local students to develop the necessary software applications, thereby building capacity, providing opportunities for hands-on technical experience, and encouraging entrepreneurship.

Although the development of this project will focus on the specific needs of African nations, the project itself will have global applicability. The techniques and tools for multisource data aggregation developed as a part of this project will be broadly applicable to other domains seeking to utilize multisource data. The insights generated from such a rich set of data will also be of interest to air quality scientists globally. Many project outcomes can be expressed in terms of various international development goals and policy mandates. With respect to the UN 2030 Agenda for Sustainable Development, the project will support the good health and wellbeing of people by providing timely, actionable information on local air quality which can be used to make decisions about where and when to limit exposures in the near term and where efforts at air quality improvement and pollution mitigation might be focused in the long-term. It also supports quality education by providing a broadly accessible resource of air quality information which educators can use to engage with students on issues of environmental quality and justice, as well as directly supporting the education of the computer science students who will be involved in the project. In terms of the Paris Climate Agreement, the strong link between climate change and air quality means that a system for gathering and interpreting air quality data will also be useful in helping nations track various greenhouse gas levels and levels of other airborne contributors to global warming such as black carbon. With respect to the Sendai Framework for Disaster Risk Reduction, this project provides a focus on monitoring and assessment of hazards posed by poor air quality. It also represents a platform for the free dissemination of such information, substantially increasing access to risk warning via a website application as well as a SMS-based air quality warning tool which will be implemented as part of the project. The different needs of different classes of users of the system, ranging from the general public, community groups, and air quality activists to educators, scientists, and policy makers, will also be considered and incorporated at all phases of the project. Finally, with its focus on the specific needs and concerns of African nations, the project will integrate into the goals of AfriGEOSS by increasing accessibility for earth observation data in Africa and by combining it with other data sources. This includes a special emphasis on the inclusion of low-cost sensor data with other data sources, since these low-cost sensors represent a highly cost-effective way to expand air quality data coverage in previously unmonitored or under-monitored regions such as those in Africa.

Project Outline and Goals

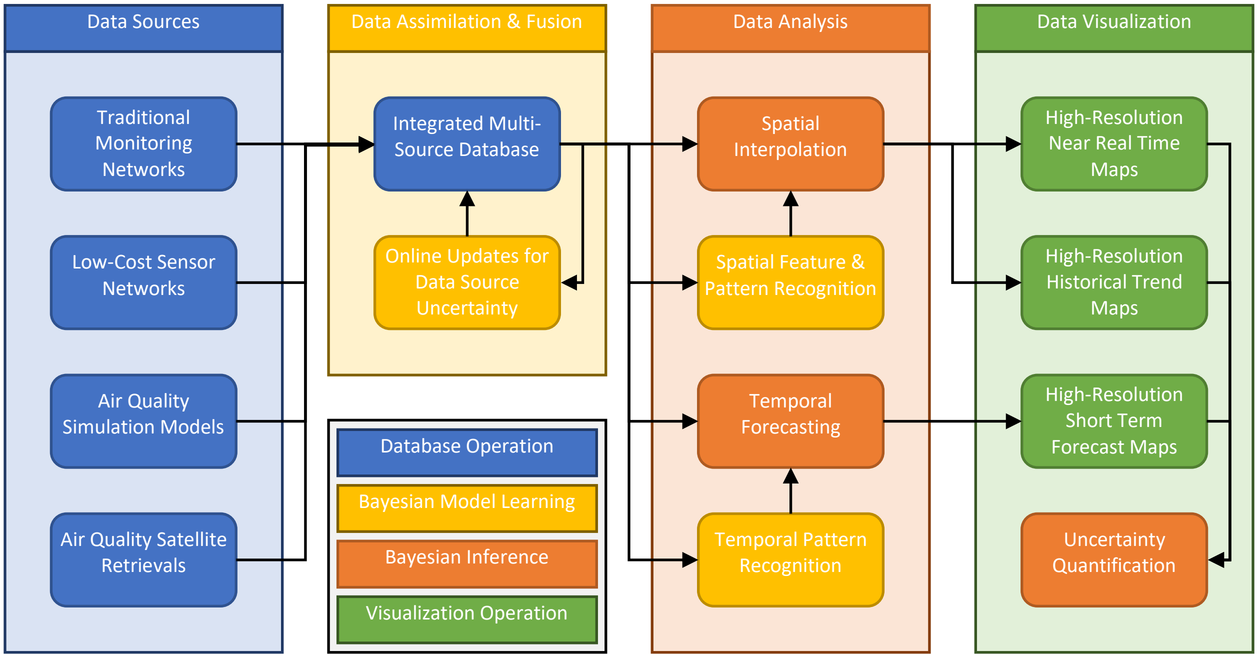


Figure 1: Schematic of the proposed project system.

The goal of this proposal is to use Amazon Web Services to develop a backend that assimilates data from various sources with associated uncertainties determined through expert input and machine learning, and produces real-time visualizations and short-term air quality forecasts to help residents in Sub-Saharan Africa (SSA) reduce their exposure to ambient air pollution. We envision the following system: the back end will have a database combining multiple data sources, including measurements of traditional monitoring stations, low-cost sensor systems, simulation results, and satellite observations. This will be coupled with a system for online learning to improve data assimilation. Starting from prior assumptions about data quality, the system will learn the relative uncertainties of different data sources to appropriately assign uncertainties to them as part of the combined database. From this database, the system front end will provide a web based interactive map of local air pollution in near real time, accessible by the general public using smartphones or other devices, and capable of providing various data visualization products, including:

- Easy to interpret qualitative near real time maps for the best estimates of current pollution (up to date to within the last hour), at a high spatial resolution (neighborhood-scale).
- Quantitative spatial interpolation maps and near-term forecasts (several hours to several days in advance) showing predicted pollution levels based on historical trends, simulations, and all available assimilated data.
- Long term averages and trends based on historical simulations and measured trends to allow for decision making about areas of chronic pollution and impacts of policy and technology changes over time.
- For all of the above, uncertainty estimates will be available, and options for average or different worst- or best-case scenarios depending on residual uncertainties will be possible.

The system front end will be optimized for use on smartphones, due to the relative prevalence of such devices in Africa for internet access (versus desktop or laptop computers). Furthermore, a

SMS based air quality alert system will be developed to provide real time information to interested individual users about their local air quality using this system, to avoid the need to have access to the full visual interface via smartphone in order to receive this information.

The fundamental goal of this project is to develop, test, and deploy the system as outlined above and illustrated in Figure 1. This implies several subsidiary goals as well. One key scientific and technical goal is to develop and verify the methodologies for Bayesian data assimilation which will combine data from multiple sources into a single coherent database. Solutions to this open technical problem will have wider application and interest to the physical and computer science communities. Additionally, tools for spatial interpolation and temporal forecasting of air quality will be of interest in the broader global air quality community, as the specific forms of these models will be of interest and utility for other air quality modeling and forecasting projects worldwide. Another major goal associated with the deployment of the system is to gain a better understanding of how users will interact with and make use of the system. Data will be gathered on what information different types of users (ordinary citizens, policy makers, educators, scientists, etc.) find most useful or access most frequently. Discussions with users will identify how they might like to see the information better presented to improve its clarity and utility for their purposes.

This project will necessitate the collaboration of many individuals with diverse backgrounds, including computer scientists and programmers, air quality researchers and data technicians, social scientists, and local experts and community stakeholders. The CMU Africa campus in Kigali, Rwanda (www.africa.engineering.cmu.edu) has Masters programs in electrical and computer engineering (MSECE) and information technology (MSIT), in which students learn data science, machine learning, and software development. As part of these degree programs, students take practicum courses and collaborate on research projects. This provides opportunities for the students to collaborate on the development of the systems envisioned for this project as part of their coursework. These students will develop the necessary software applications, thereby building local capacity, providing opportunities for hands-on experience, and encouraging entrepreneurship. Several members of the project team (Subramanian, Brown, Jaramillo) are also fellows at the Kigali Collaborative Research Center (KCRC; www.kcrc.rw). They work with and mentor students and researchers at CMU Africa and at the University of Rwanda College of Science and Technology (cst.ur.ac.rw), and we will use their connections to support interactions with local stakeholders. Project members Subramanian and Malings are supported by the French National Research Agency (ANR) through the "Make Air Quality Great Again" project, which will support other aspects of project development not covered by the Amazon Web Services grant. Code for Africa (CfA) is the continent's largest federation of independent civic technology laboratories, which use open source software and open data to build digital democracy services that give citizens timely and unfettered access to actionable information that empowers them to make informed decisions and that strengthens civic engagement for improved public governance and accountability. Backed by CfA, sensors.AFRICA is a network of citizen sensor projects with air quality sensors deployed in Kenya, Nigeria, South Africa, and Tanzania.

Background

Key atmospheric pollutants

Air quality is the largest environmental risk factor for human health; exposure to outdoor air pollution is estimated to cause about 4.2 million premature deaths annually (WHO, 2016, 2018a), Air quality guidelines by the WHO describe particulate matter, ozone, sulfur dioxide, and nitrogen dioxide as the main pollutants of concern (WHO, 2006).

Particulate matter (PM) mass is one of the most commonly tracked pollutants, as it has the largest health impact among the pollutants listed above. PM represents a mixture of solid and liquid substances, including organic aerosol, sulfates, nitrates, ammonia, mineral dust, and black carbon (e.g. from fuel combustion). PM can be naturally occurring or anthropogenic, including emissions from the burning of fuels, and from cooking, which is a major source in the developing world. PM mass concentration is typically tracked as both PM_{10}, the total mass of PM with a diameter of 10 micrometers or less, and $PM_{2.5}$, the total mass of PM with diameter below 2.5 micrometers (and a subset of PM_{10}). PM_{10} is the inhalable fraction of ambient PM, and particles smaller than 2.5 µm ($PM_{2.5}$) can reach deep into the lungs and eventually enter the bloodstream. These particles are closely linked to lung cancer as well as cardiovascular and respiratory diseases (e.g. Schwartz et al., 1996; Pope et al., 2002; Brook et al., 2010). Even low PM concentrations can have noticeable health impacts (Bell et al., 2007; Apte et al., 2015), especially among low-income communities (Di et al., 2017; Ren et al., 2018).

Ozone (O_3) forms as a result of photochemical reactions of nitrogen oxides and volatile organic compounds emitted by vehicles and industry. Thus, ozone can be used as a proxy for some of these other pollutants. However, it is also more directly harmful, causing lung inflammation and triggering asthma in susceptible individuals, as well as contributing to photochemical smog.

Nitrogen dioxide (NO_2) is related to the production of $PM_{2.5}$ and ozone, and is emitted mainly from combustion, including power generation and use of vehicles. Thus, it can also be used as a proxy for combustion emissions generally. It is toxic in high concentrations, and at lower concentrations it contributes to reduced lung function and bronchitis symptoms in asthmatic children.

Sulfur dioxide (SO_2) is emitted through the burning of fossil fuels (for vehicles, heating, and power generation) and processing of ores containing sulfur. Exposure to SO_2 causes irritation of the eyes and lungs, causing coughing and aggravation of chronic bronchitis and asthma. Higher SO_2 levels are correlated with increased hospital emissions and mortality from cardiac diseases. Furthermore, it combines with water to produce acid rain, causing damage to infrastructure and deforestation.

Besides the pollutants mentioned by the WHO, several other pollutants are of interest for air quality and environmental monitoring. Carbon Monoxide, emitted through fuel combustion, reduces oxygen supply to the body, causing loss of consciousness or death at high concentrations. It is especially dangerous when emitted by cookstoves or heaters in poorly ventilated buildings. Carbon dioxide, though typically not directly dangerous, is the most prevalent greenhouse gas. Black Carbon, emitted from vehicles, power plants, and cookstoves, is associated with respiratory and cardiovascular diseases, and is the second largest climate forcing agent after carbon dioxide (Bond et al., 2013). Volatile organic compounds (VOCs) represent a variety of compounds which contribute to poor health (in indoor environments) and photochemical smog (outdoors). This category also includes methane, a major contributor to global warming (although methane is sometimes also considered separately because of this).

Air quality impacts and measurements in Africa

Open burning and residential use of biofuels for heating and cooking are the main anthropogenic sources of air pollution in Africa (Liousse et al., 2010). Additional natural sources include dust from the Sahara desert and smoke from grassland fires (Naidja et al., 2018). In Africa, about 10%

of communities assessed by the WHO met recommended air quality guidelines, which is below the average of 18% globally, and well below the 40-80% rates found in high income countries in Europe and North America (WHO, 2018b). A study of infant mortality in SSA found that a 10 $\mu g/m^3$ increase in $PM_{2.5}$ concentrations was associated with a 10% increase in infant mortality, and that 22% of infant deaths could be attributed to high $PM_{2.5}$ concentrations (Heft-Neal et al., 2018). It is further estimated that every child under the age of five in SSA has been exposed to unsafe levels of air pollution, which is expected to have lifelong consequences in terms of a greater risk of chronic respiratory and cardiovascular diseases (Matshidiso Moeti, 2018). In 2013, gross domestic product growth in SSA was reduced by an estimated 3.8 percent, or $114 billion, due to the impacts of poor air quality on the economy (World Bank, 2016).

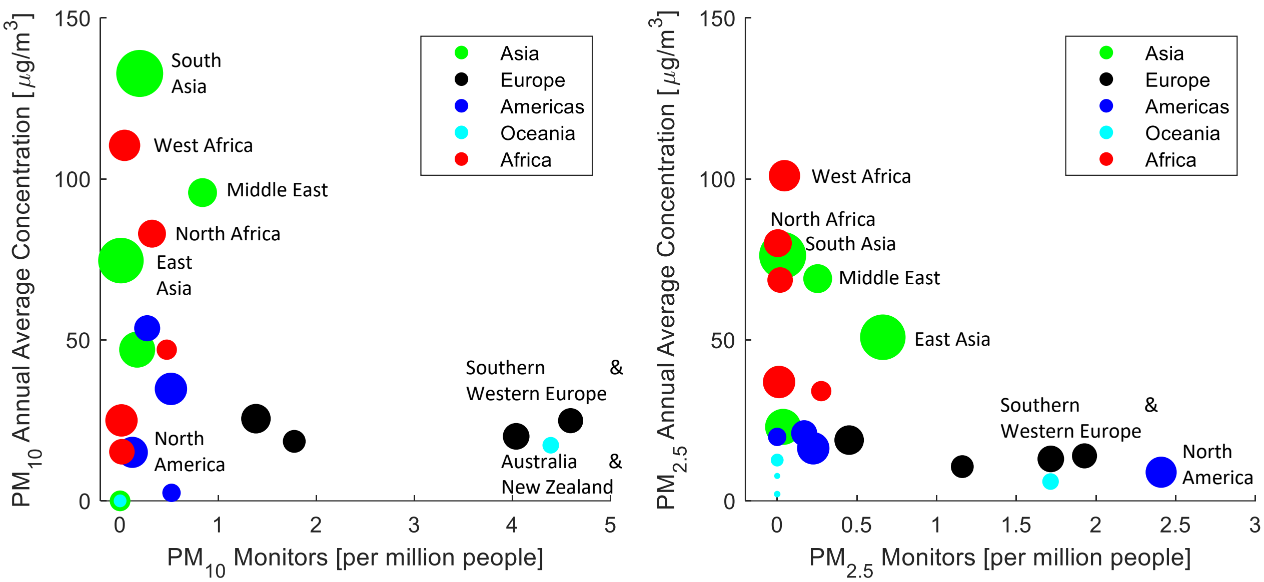


Figure 2: Scatterplot comparing estimated annual average pollutant levels (for PM_{10} and $PM_{2.5}$) versus number of monitoring stations of these pollutants per million people, broken down into several global region. Sizes of the points roughly correspond to the population of each region.

African countries have some of the highest modeled annual average PM_{10} and $PM_{2.5}$ levels, yet are simultaneously among those with the lowest number of in situ measurements per capita. In Figure 2, we present information about the density of monitoring in different regions versus the estimated average annual PM_{10} and $PM_{2.5}$ concentrations for these regions, based on information from the Global Health Observatory (WHO, 2017). Countries in Africa, especially in West and North Africa, have some of the highest average PM pollution levels, second only to regions like South and East Asia and the Middle East. However, these regions also have the fewest per capita air pollution monitoring stations for either PM_{10} or $PM_{2.5}$. While countries like China and India have significant air pollution problems, they also have at least modest monitoring networks, e.g. about 350 PM_{10} monitoring sites in India, and more than 1000 $PM_{2.5}$ monitoring sites in China (although this is still well below per capita monitoring levels in developed countries in Europe, for example). For all of Africa, on the other hand, there are less than 150 PM_{10} and 50 $PM_{2.5}$ monitoring stations, and just Egypt and South Africa together account for more than half of these stations, leaving most of the continent quite sparsely monitored.

Quantifying air pollution levels and identifying their sources and dispersion in the atmosphere are the first steps toward pollution mitigation activities. However, many countries in Africa lack even a basic air quality monitoring network for this purpose (Petkova et al., 2013). In 2018, only 41 cities across 10 countries in SSA tracked their air quality (Matshidiso Moeti, 2018). Furthermore, global efforts like the Global Burden of Disease Project (Brauer et al., 2012), which aims to identify the contribution of air pollution and other factors to human mortality and disease, are hindered by the lack of ground based air quality monitoring for Africa, and therefore may misrepresent the burden of disease for the continent.

Recent work (REMA, 2018) in Rwanda towards the development of a national air quality control strategy has identified vehicle emissions as the largest contributor to poor air quality in urban areas, while rural areas are most impacted by domestic stove emissions. However, the lack of a national air quality monitoring network was identified as a major impediment to future work, and it was recommended that such a network be deployed to better understand the spatial distribution of air pollution and to track air quality changes to assess the impacts of any future air quality control strategies (REMA, 2018). The government of Rwanda has purchased eight low-cost RAMP sensors systems to meet this need following the success of a pilot deployment project (see "low-cost sensors" below).

Several research groups have established long term air quality monitoring networks for Africa. These include the International Network of Deposition and Atmospheric chemistry in Africa (INDAAF) network (https://indaaf.obs-mip.fr/), Figure 3, which maintains a total of 13 stations across the continent measuring the gases NO_2 and ozone, as well as PM_{10}. The Rwanda Climate Change Observatory at Mount Mugogo has been recently established with support by MIT to monitor greenhouse gases and air pollution, including black carbon particles, carbon dioxide, carbon monoxide, nitrous oxide, and ozone (DeWitt and Gasore, 2018). The Henties Bay Aerosol Observatory (HBAO) monitors total suspended particulate matter (TSP) and black carbon mass (http://www.hbao.cnrs.fr/).

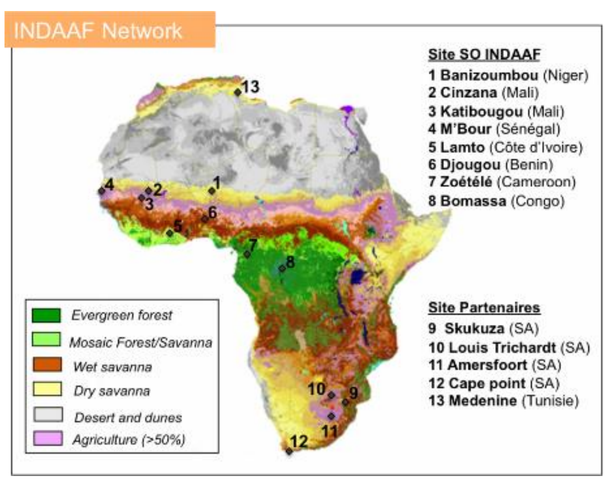


Figure 3: Map of the INDAAF Network in Africa. Each site includes passive gas and PM samplers which must be periodically sent elsewhere in the world (typically the Laboratoire d'Aerologie in Toulouse, France) for analysis (https://indaaf.obs-mip.fr/).

At current rates of development in Africa, anthropogenic emissions from industrial activity and transportation will overtake residential pollution sources as the main contributors to poor air quality in African nations by 2030 (Liousse et al., 2014). Climate change will also likely contribute to poorer air quality in the future. Increased use of energy for cooling will lead to increased emissions from power generation worldwide, including in Africa (Abel et al., 2018). Increasing frequency and severity of dust storms will loft more particulate matter into the atmosphere, independent of the pollution impacts of human activity (UNEP, 2016). Different rainfall patterns could lead to less effective and/or consistent removal of this PM from the atmosphere through the "washing out" effect of rainfall (Silva et al., 2017). These climate related trends will only exacerbate the impacts of increasing urbanization and industrialization on air quality in Africa.

Traditional air quality monitoring, especially in Africa, is too sparse to meet the growing challenges of development and climate change. More information is needed in growing urban areas to determine specific contributors to local air pollution, which can support the development city-specific air quality management plans. This project seeks to develop and deploy a system for integrating air quality data from disparate sources, including traditional ground based monitoring, satellite retrievals, computational and statistical air quality models, and low-cost sensor networks.

Low-cost sensors

Currently, in countries with air quality standards, regulatory monitoring is carried out by a network of fixed monitoring stations (e.g. Snyder et al., 2013). Due to their relatively high setup and operating costs, only a few are established, typically in major urban areas. Even in developed nations such as the United States, the Environmental Protection Agency (EPA) determines compliance at the county level, with many rural counties having at most a single monitoring site. However, air pollutant concentrations can vary greatly at smaller spatial scales, especially within urban areas with their large numbers and varieties of pollutant sources (Marshall et al., 2008; Karner et al., 2010; Tan et al., 2014). This variability means that even air quality estimates based on high accuracy, high cost monitoring stations may underrepresent the variability or extremes in local air pollution (Jerrett et al., 2005b). Of course, in regions where such a monitoring network is not present in the first place, this problem is only exacerbated.

Low-cost sensor packages represent one alternative for increasing the spatial density of air quality monitoring. Low-cost monitors combine several comparatively inexpensive air pollutant sensors (typically electrochemical or metal oxide sensors for gases and laser optical sensors for particulate matter) with an independent power source and wireless communication capability. Since each monitor has a relatively small purchase cost and is designed to operate autonomously for extended periods of time, low-cost monitors can allow a relatively larger number of measurement stations to be deployed for a similar cost to a more traditional monitoring network. This can be done for the purpose of supplementing traditional monitoring networks where they are available, or for establishing new networks in previously unmonitored areas; the latter is typically the use case for developing countries. These monitors also allow for increased community involvement and education in air quality monitoring, as it is often convenient to distribute these simple low-cost instruments to concerned citizens and community groups, allowing them to participate in, identify with, and directly benefit from the information that the instruments collect (Snyder et al., 2013; Loh et al., 2017; Turner et al., 2017). Several pilot deployments of low-cost sensor networks have occurred in Cambridge, UK (Mead et al., 2013), Imperial Valley, California (Sadighi et al., 2018; English et al., 2017), and Pittsburgh, Pennsylvania (Zimmerman et al., 2018).

Tradeoffs must of course be made in the use of low-cost sensors, which tend to have lower signal to noise ratios and higher cross sensitivities to other atmospheric pollutant and ambient environmental factors than traditional monitors (Popoola et al., 2016). Methods for reducing the impacts of these factors through improved calibration have been attempted (e.g. Masson et al., 2015; Spinelle et al., 2015; Cross et al., 2017; Hagan et al., 2018). In particular, the research team has recently focused on the development of the Realtime Affordable Multi Pollutant (RAMP) low-cost monitor. This monitor, jointly developed with SENSIT Technologies (USA), combines low-cost sensors measuring carbon monoxide, sulfur dioxide, nitrogen dioxide, and ozone with data storage and cellular communications capabilities. Calibration using random forest machine learning algorithms was found to be effective for most of the pollutant gases (Zimmerman et al. 2018), and it was found that the same generalized calibration models could be used across multiple monitors without sacrificing measurement accuracy (Malings et al., 2019a). Furthermore, RAMPs may be paired with low-cost PM sensors, such as PurpleAir's PA-II, which make use of Plantower PMS5003 optical sensors. Such PMS5003 sensors, alogn with Shinyei PPD4NS and Nova SDS011 sensors, are also used by sensors.AFRICA, which, with the support of CfA, locally assembles low-cost air quality sensors using off-the-shelf components. The sensors.AFRICA kits measure PM_1, $PM_{2.5}$, PM_{10}, temperature, and humidity. Academic research and empirical data shows the shortcomings of some optical sensors, where high relative humidity (above 50%) negatively impacts the sensor's performance. The PMS5003 was found to be better characterized and has been shown to perform well in field evaluation studies conducted by authorities such as the US EPA (AQ-SPEC, 2017) and academic groups (Malings et al., 2019b). In particular, using appropriate calibrations incorporating basic environmental factors, the accuracy of the lost-cost $PM_{2.5}$ sensors associated with the RAMPs could be improved to the point where measurement uncertainties were reduced to about ±4 μg/m^3 for hourly measurements and ±1 μg/m^3 for annual averages, a significant improvement over the un-corrected measurements (Malings et al., 2019b).

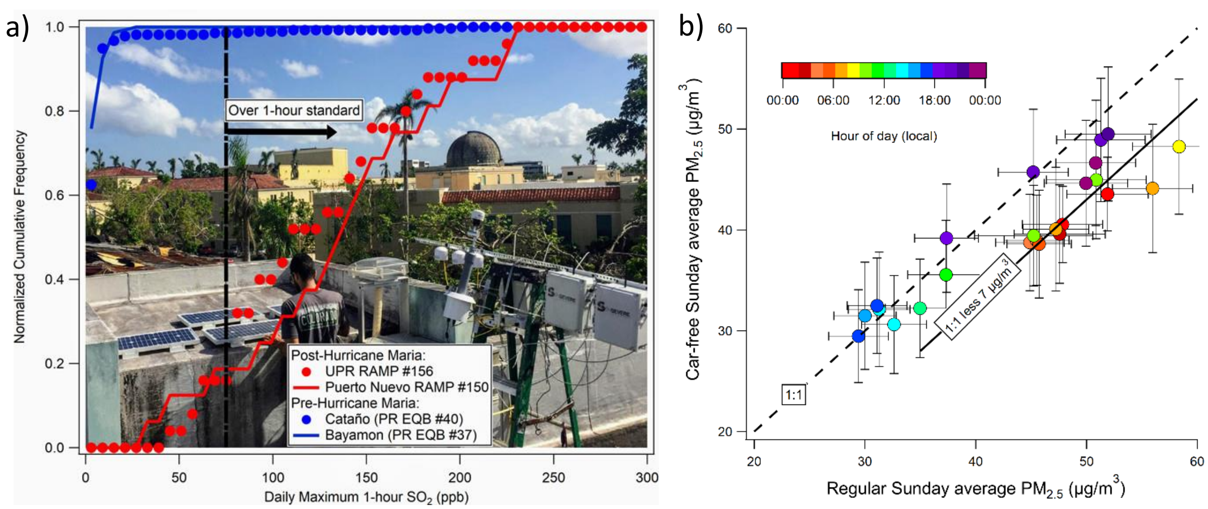


Figure 4: Results from low-cost sensor deployments. a) deployments in Puerto Rico following Hurricane Maria identified elevated levels of SO_2 due to the use of high-sulfur-content fuels for emergency generators. b) in Kigali, Rwanda, low-cost sensors identified the impact of Car-free Sunday policies as reducing ambient $PM_{2.5}$ levels by up to 7 μg/m^3 in the morning.

Data from low-cost sensor have yielded several new scientific insights (see also Figure 4). From 2016 to 2019, a network of about 50 RAMPs was deployed around Pittsburgh, Pennsylvania as a proof of concept for a dense low-cost monitoring network for an urban area. In the aftermath of Hurricane Maria, four RAMPs were deployed in Puerto Rico, where they assessed the impact of the hurricane and subsequent recovery efforts (during which diesel with higher than normal sulfur content was used to power emergency generators) on local air quality (Subramanian et al., 2018). RAMP deployments in Kigali, Rwanda showed high nighttime $PM_{2.5}$ levels resulting from cooking emissions, as well as significant diurnal variations in both CO and $PM_{2.5}$.

Satellite data

Satellites remotely monitor atmospheric pollution over wide areas of the earth's surface by measuring Aerosol Optical Depth (AOD) or other atmospheric properties. Multispectral analysis of IASI and GOME-2 measurements has provided a new capacity to estimate tropospheric ozone pollution levels down to the atmospheric boundary layer using satellite data (Cuesta et al., 2013). Some satellites, like the Sentinel-5P of the Copernicus Program with its TROPOMI instruments, can remotely monitor multiple air pollutants, including NO_2, ozone, sulfur dioxide, carbon monoxide, and airborne particulate matter ($PM_{2.5}$ and PM_{10}), as well as several greenhouse gases such as carbon dioxide and methane (http://www.tropomi.eu/). The US MODIS system provides AOD measurements across the entire Earth every two days (Engel-Cox et al., 2004). Satellite air quality data products from many sources are available through the GEOSS portal. Satellites thus have the potential to provide data on a number of pollutants of interest at a worldwide scale.

However, there exist several limitations to satellite data. Satellite coverage is not continuous, and for a given point there may be hours or days between satellite passes. Even during satellite passes, cloud cover or atmospheric irregularities can disrupt data acquisition, and some instruments only acquire daytime data. Satellites do not necessarily measure pollutant concentrations at or near ground level, where pollutants impact human health directly (Cuesta et al., 2013). Furthermore, the algorithms used to extract air quality information from satellite data are not necessarily applicable worldwide or under all weather and atmospheric conditions, and may introduce significant bias between the remotely measured and true concentration; this is especially the case for Africa, where dust and fine aerosols are typically found much higher into the atmosphere than in other parts of the world, making assumptions about the relationship between PM and AOD less applicable here (van Donkelaar et al., 2015). To compensate, in situ measurement data may be used to recalibrate the concentration estimates from satellites (e.g. Han et al., 2018); unfortunately, Africa's sparse network of in situ monitors makes this especially difficult. Satellite downtime and delays in processing satellite data also mean that these data are not typically available in real time. Finally, the spatial resolution of satellite retrievals is limited to lower resolutions, from about 1 km for MODIS AOD data to about 7km or more for other pollutants from TROPOMI (http://www.tropomi.eu/), meaning local scale variability cannot be captured. These shortcomings can be mitigated by coupling satellite data to ground based monitoring. This allows for "filling the gap" between satellite passes, as well as providing a point of reference to recalibrate the satellite data so they better agree with the ground based measurements at locations where these overlap (i.e. so that any biases introduced by the calibration methods can be identified and removed).

Modeling approaches

Air quality models, such as CHIMERE (Menut et al., 2013) or PMCAMx (Fountoukis et al., 2016) allow the simulation of past, present, and future air pollution. These models range in scale from global simulations to highly specialized local models, taking into account factors such as street canyons in urban areas. A big advantage of such models is the ability to predict temporal variations in air quality based on available and predicted meteorology. However, these models are computationally expensive and limits in the available input data will lead to unrealistic model outcomes. To compensate, model predictions are typically calibrated against in situ measurements such as those provided by traditional air quality monitoring networks (Jerrett et al., 2005a), much like what is done with satellite measurements. Even so, the models may not adequately capture local scale variability; analysis for RAMP data collected in Kigali, Rwanda has shown that, while global pollution models can capture background concentrations, they can significantly underestimate pollution in cities where population, emissions, and exposure are concentrated.

Simpler statistical models of air pollution, such as land use regression, attempt to explain air pollution as a function of the land usage, geographic, and demographic conditions of the surrounding areas and (in some cases) the time of day. However, these models tend to generalize poorly to regions different from those where they were developed (Hoek et al., 2008; Sally Liu et al., 2012) and cannot account for transient pollution events (e.g. from fires or dust storms). Models can fill in gaps in other data sources, including satellite data and locations unmonitored by traditional networks. However, all models require calibration against in situ measurements, meaning their applicability will continue to be constrained by the availability of data sources.

Bayesian probabilistic inference and learning

Air quality data from traditional in situ monitoring networks, low-cost sensor systems, satellite retrievals, and computational models all have certain benefits and drawbacks. In many cases, the shortcomings of one system can be mitigated, at least in part, if data from other sources is available for calibration and to “fill in” gaps in the data. Bayesian probabilistic modeling provides a consistent theoretical framework under which such a multisource data fusion can be accomplished (Waltz and Llinas, 1990; Castanedo, 2013). In general, Bayesian probabilistic modeling implies the development of a statistical model of the quantity or quantities of interest (in this case, the concentrations of various air pollutants), serving as a “prior” model. Various sources of evidence or observations of the quantities of interest can be used to update this model, with the uncertainties in this evidence treated appropriately, resulting in a “posterior” model.

The use of probabilistic models to describe phenomena which vary in space and time (so called spatiotemporal random field modeling), together with the use of Bayesian techniques to update these probabilistic models using uncertain observations of the random fields, has an extensive history (see e.g. Cressie and Wikle, 2011). Malings et al. (2017) used a form of Bayesian probabilistic modeling know as a Gaussian Process modeling framework for the modeling of urban temperatures in heat wave scenarios via a probabilistic prior model. This model was then combined with temperature data from a variety of sources (such as long term regional temperature forecasts from computational weather models, high resolution urban microclimate simulations, and in situ weather station temperature measurements). The resulting posterior urban temperature models could be used to make improved predictions about the severity and duration of heat wave events, taking into account information from these various sources to better represent the intensity and duration of the heat wave overall as well as the specific effects in different areas of a city due to the urban heat island effect (Malings et al., 2017, 2018). This work provides a complete case study for the use of Bayesian spatiotemporal probabilistic modeling for phenomena such as air temperature (or even air quality) which vary over space and time and which can be measured by various instruments at different spatial and temporal resolutions.

Finally, Bayesian probabilistic modeling allows for preposterior analysis, or an analysis of the impacts of various information sources on the reduction of uncertainties between the prior and posterior models (Berger, 1993). Such an analysis can evaluate the relative “value” of different sources of information, in terms of how their use leads to an improvement and reduction in uncertainty with respect to the prior model. Utilizing such an approach, not only can a Bayesian model incorporate information from different sources with different levels of uncertainty, but over time the system can "learn" the relative uncertainties, and thus the relative benefits, of different sources of information (Bergerson and Muehleisen, 2015; Maddox et al., 2019). Initially, such a system would have to be provided with an assumed "best guess" as to the uncertainties of different information sources. Over time, the system will be presented with more data from each source, and will be able to compare the evidence provided by each source to the "aggregate truth" derived from all data sources. In doing so, it would recalibrate its estimates of the uncertainties and/or biases inherent in each source. Thus, when new data were obtained from this source, the system would be able to more appropriately handle them in terms of their uncertainties compared with other sources and therefore in terms of their potential to improve the final aggregate truth model.

Identification of needs and project contributions

No one source of air pollution information is sufficient to provide citizens, scientists, and policy makers with the measurement accuracy, spatial coverage and resolution, and timeliness to meet all of their data needs. Satellites may only provide coverage once or twice per day; computational models can provide higher temporal resolution, but both these and satellite data require calibration against ground-based monitors. This situation is exacerbated for Africa, where high accuracy in situ monitors are rare, global satellite information coverage may not accurately reflect local conditions, real time simulation models are too computationally costly to run, and low-cost sensors to fill gaps in the above systems are still an emerging and relatively inaccurate technology. However, together, these sources can fill in the gaps and cover the weaknesses of each other, with Bayesian probabilistic modeling providing a theoretical framework for integrating these data.

What is needed, and what we will develop in this project, is a practical implementation of a system as outlined above. It will interface with multiple open source online databases and correctly interpret and integrate their data, creating a single unified estimation of air quality across space and time and for different pollutants. This system will also be able to "learn" and adapt, for example by comparing data from different sources to appropriately assign uncertainties to these data based on how often and to what extent the different sources agree or disagree with one another.

Project Implementation Plan

Phase 1: Technical

Task 1: Gathering and interfacing data from multiple sources

Multiple sources of freely available air quality data for Africa will be considered for this project. This includes, but is not limited to, existing and future RAMP low-cost sensor data (pghaqmap.com), sensors.AFRICA's low-cost PM sensors (sensors.africa), air quality data available through the OpenAQ API (openaq.org), the PurpleAir network of low-cost PM sensors (purpleair.com), and satellite data accessible through the GEOSS Platform. In interfacing these sources, the project will make use of GEOSS technical standards to ensure compatibility.

Goals: To develop a system to automatically query various databases for freely available air quality data from worldwide sources. This system should be inherently adaptable and expandable to include new data sources as they become available, with data from all sources being translated to a common format and integrated into a common multisource database.

Task 2: Bayesian data fusion and learning

Initially, uncertainties will be assigned to different data sources based on a literature review of uncertainty estimates, as well as prior assessments conducted with RAMP sensors and sensors.AFRICA's sensors. Uncertainties will be divided into "bias" and "noise", representing systematic offsets to which the sources are susceptible (including how offsets might change over time) and instantaneous differences between measurements due to inherent sensor noise. These prior estimates of uncertainty will be updated over time through Bayesian learning techniques. For example, the database of Task 1 will be periodically queried to identify collocated measurements, which can be compared to each other to revise uncertainty estimates.

Goals: To develop, test, and implement a system for quantifying the uncertainties in data from disparate sources using Bayesian techniques. This system must operate alongside the multisource database developed in Task 1, using information provided by this database to update its uncertainty estimates while feeding these estimates back to the database for use with later tasks.

Task 3: Spatial interpolation of data

Simple spatial interpolation between measurement locations based on relative distances is often insufficient to capture gradients in an urban environment (e.g. Schneider et al., 2017). A probabilistic prior model making use of factors such as land use may therefore be necessary to provide the necessary spatial fidelity. In specific regions, high resolution satellite or model simulation data may be available to substitute for these land use-based estimates. These two methods may also be combined; land use regression models can be calibrated using high resolution multisource data where available, and then generalized to other regions where data are sparser. Thus, this task will determine generally applicable methods of utilizing the dense multisource data from Task 1 to learn spatial patterns and relationships between air quality and geographic or land use factors. These patterns will form the priors for Bayesian inference of spatial pollution, to be combined with local multisource data to create the posterior pollution and uncertainty maps.

Goals: To develop, test, and implement a system for performing spatial interpolation of air pollution information using multisource data as compiled in Task 1. This system must respect the relative uncertainties of different sources quantified in Task 2 and must recognize spatial patterns and features of air pollution and replicate these where applicable as part of its interpolations.

Task 4: Temporal forecasting of data

Short term air quality forecasting will use common recurring patterns (e.g. diurnals) of air quality in different areas from the data of Task 1. These patterns (the prior) will be combined with recent air quality data (reflecting transient pollution events) via Bayesian inference to create the posterior forecast. As in Task 3, the appropriate methods for doing this remain an open research question. Therefore, as part of this task, different methods of performing this integration will be investigated and tested. Finally, residual uncertainty in forecasts will be quantified (e.g. we expect uncertainty to increase as prediction lead time increases, up to the limit of longterm aleatory uncertainty).

Goals: To develop, test, and implement a system for performing temporal forecasting of air pollution information using multisource data as compiled in Task 1, respecting the relative uncertainties of different data sources as quantified in Task 2. This system will make use of recurring temporal pollution patterns and extrapolate these forwards in time where applicable.

Phase 2: Implementation

Task 5: Visualization of data interpolation and forecasting

The data and Bayesian inference tools described previously must have their results presented in a suitable manner. Therefore, in this task, a web based tool for querying the multisource dataset of Task 1, as well as accessing the inferences of Task 3 and Task 4, will be created, building upon the existing system of sensors.AFRICA (map.aq.sensors.africa). Options will be provided for data screening, e.g. only presenting data from certain sources, or presenting multisource data without interpolation or forecasts. Options for quantitative or qualitative presentations will also be given, e.g. a quantitative presentation including uncertainties versus a qualitative presentation of "clean" versus "polluted" areas. Different timescales will be considered, including near real time information, near term forecasting, and historical information on seasonal or annual averages.

Goals: To present the results of products generated in Task 3 and Task 4, including residual uncertainties, in a clear visually interpretable manner, e.g. using maps. Where desired, methods for converting quantitative into qualitative information will be used. This task will require input from social scientists concerning the best way to convey these various data to different audiences.

Task 6: Program integration and testing

Integration of all the previously described components of the project into a single cohesive program is encapsulated under this task. It must be ensured that the resulting program can run under a variety of different usage cases and on a variety of platforms, especially smartphones. Additional elements necessary for the project, such as SMS based alert systems and user feedback collection tools, will be developed in collaboration with CfA as part of this task.

Goals: To combine outputs of Tasks 1-5 in a system implemented using cloud computing services.

Task 7: Pilot deployment and user interaction studies

Once system development is completed under Task 6, a pilot deployment of the program will be conducted. This will represent a limited release of the system to selected users (e.g. air quality specialists including researchers, activists, educators, and decision makers worldwide) and to general users in a small geographical area (e.g. Kigali, Rwanda; Nairobi, Kenya; and Lagos, Nigeria). This task requires extensive coordination with local community groups and stakeholders to ensure participation of a representative group of anticipated product users, as well as to facilitate the gathering of anonymous user feedback though the website. KCRC and CfA will provide the necessary on-the-ground support at this stage. Engagement with key stakeholders of interest, such as local environmentalists, decision makers, educators, and researchers, is of particular importance during the pilot deployment to ensure their needs are being met.

Goals: To collect information on the performance of the complete program, especially on the ways users interact with the system and how useful its features are to different groups of users.

Phase 3: Usage

Task 8: System refinement and launch

Following Task 7, modifications of the system will be made. It is likely that many modifications will be required for the user interface system, to ensure that the information which is most useful to the general public is most readily accessible, and that the interface strikes an appropriate balance between ease of use and comprehensiveness in the data it provides. An additional round of pilot testing for the system may also be required. Once these refinements are substantially completed, the system will be released generally. At this stage, the finalized system's source code, including tools for interfacing with and combining data from existing free online data sources, will be made freely available to the general public, in accordance with GEOSS Data Sharing Principles.

Goals: To adapt and improve the existing program, based on the results of Task 7, to the point where it can be deployed for use by the general public.

<u>Task 9: System management and expansion in cooperation with local users and developers</u>

As the project's system is deployed, we will seek new opportunities to continue and expand the project's tools into the future. This effort will heavily focus on involving local users and software developers in the project, especially students who were involved in initial pilot deployments. It is especially important at this stage to involve educators who will make use of the system as a tool to educate students about air quality and other environmental health and justice issues. The involvement of computer engineering and information technology students at CMU-Africa and CfA contributors from a variety of backgrounds will foster a new generation of computer scientists who will support and expand the system in the future.

Goals: To facilitate the operation and expansion of the system through interactions with local users and developers, building local and community capacity to interact with and improve the system.

<u>Task 10: Ongoing improvement and expansion based on user and stakeholder input</u>

Beginning with the initial deployment of the system and continuing into the future, periodic reassessments of the system in terms of its functionality, how it is being used, and how it might be expanded should be conducted. Again, involvement of a broad range of stakeholders will be important to ensure these reevaluations are honest and beneficial. Through the efforts of Task 9, the pool of engaged active users who can provide beneficial inputs, as well as the pool of programmers and local entrepreneurs capable of expanding the system, should be continuously expanding. A system of accountability of the managers of the system to its users should also be established to ensure that the recommendations of these reevaluations are implemented.

Goals: To conduct periodic reassessment of the system and any new functionalities which might be added, based on feedback from users and engagement with community groups and stakeholders.

<u>Web Services Needs</u>

A preliminary analysis of the web services needs of this project was conducted using the Amazon Web Services simple monthly calculator tool (http://calculator.s3.amazonaws.com/index.html). This estimate used the template of a large web-based application as a starting point, and several specific additional capabilities were added. Ultimately, it was determined that a web services cost of approximately $2750 per month is likely required to support the needs of this project. This translates to a total cost of $99000 over three years. This estimate includes:

- At least 6 large servers for running data analysis and web based applications, allowing for redundancy and simultaneous operations of multiple instances for testing purposes.
- At least 2 Terabytes of data storage for multisource air quality data; this may need to be expanded depending on the number of different sources incorporated into the project.
- At least two large database instances for multisource air quality data storage and storage of additional data generated as a result of Bayeisan learning and inference processes.
- Access to a large computing cluster for computationally intensive Bayesian learning.

- Amazon simple notification services, processing at least 1000 mobile notifications per month (this may need to be increased based on results of the pilot deployment of Task 7).

Other Sources of Support

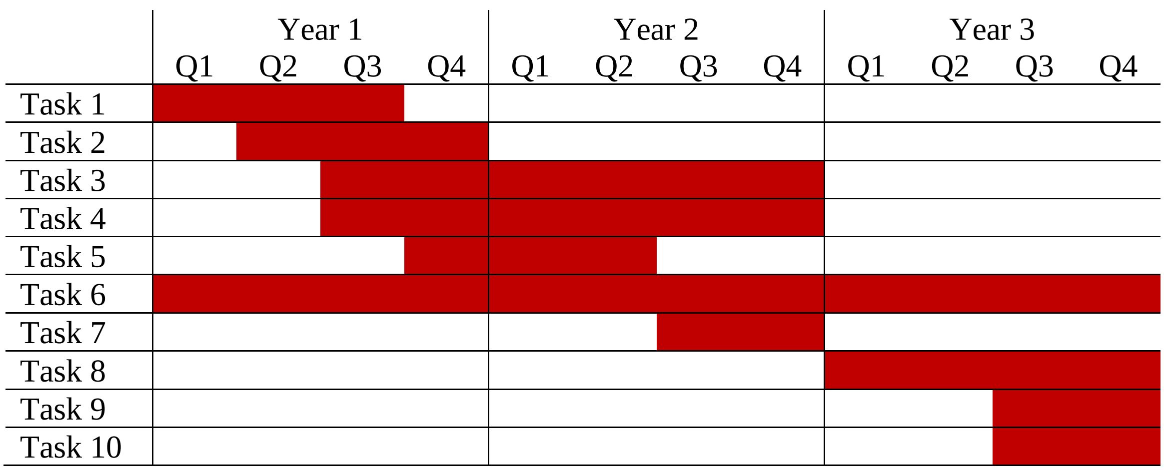
Other needs of this project, including coordination activities and meetings with local stakeholders, will be supported through the "Make Air Quality Great Again" project (http://osu-efluve.u-pec.fr/presentation/actualites/programme-make-our-planet-great-again-l-upec-encore-a-l-honneur-avec-l-accueil-d-un-laureat-a-l-osu-efluve-841917.kjsp). This project is supported by the French National Research Agency (Project number ANR-18-MPGA-0011) as part of the "Make Our Planet Great Again (MOPGA)" program. This grant supports PI Subramanian through 2022 and Malings through 2020, and can be used to support additional activities of the project as needed.

CfA uses civic (open source) technology and open data to empower citizens since 2012. CfA is a registered not for profit organization in South Africa with subsidiary organizations in Kenya and Nigeria. We work as a federation of country-based Code organizations in Kenya, Morocco, Nigeria, Sierra Leone, South Africa, Tanzania, Uganda with additional affiliate networks in 10 other African countries. CfA was established with a seed grant by the Bill & Melinda Gates Foundation (BMGF) in 2012, and has since won renewed grants from BMGF, as well as other major impact investors/donors in the "tech for good" space based on the success and impact of our work. sensors.AFRICA is a CfA initiative which uses low-cost air quality sensors to monitor and collect data in real-time, working with community organizations in a number of African countries.

Project Timeline

Table 1 presents the proposed project timeline. Initial system development (Tasks 1 through 5) will begin in year 1, continuing on as required. An initial pilot deployment (Task 7) is scheduled for year 2. The result of this pilot deployment will inform final redesign and launch of the system (Task 8) during year 3, when project sustainment activities (Tasks 9 and 10) will also begin.

Table 1: Proposed project timeline



Anticipated Outcomes

Besides the development of a system for multisource air quality data assimilation, interpolation, prediction, and presentation, as outlined above, several additional outcomes are anticipated.

Research relevant outcomes

The proposed research into Bayesian learning methods and data fusion will have applications well beyond the current project. For air quality researchers, the developed techniques and the platform for multisource air quality data integration will be widely applicable globally. Insights gained into spatial and temporal patterns of air pollution via Bayesian learning from a rich multisource dataset will provide numerous opportunities for air quality research and discovery. For the larger scientific community, innovations in Bayesian modeling of spatiotemporal phenomena and multisource data assimilation and integration developed as a result of this project will be widely disseminated in journals and at conferences (IEEE, American Geophysical Union and European Geosciences Union meetings, Africa Innovation Summit), such that these innovations can be applied to other realms of research.

Policy relevant outcomes

UN 2030 Agenda for Sustainable Development: The primary goal of this project, to provide a system for disseminating multisource information about air pollution, directly supports Goals 3 and 4 of the UN 2030 Agenda for Sustainable Development. Namely, it supports health and wellbeing by providing timely, actionable information on local air quality. It also supports education by providing a broadly accessible resource of air quality information which educators can use to engage with students on issues of environmental quality and justice, as well as directly supporting the education of students who will be involved in the project. This project is also aligned with Goal 9 (providing freely accessible digital infrastructure for air quality data) and Goal 13 (the air quality parameters included in this project include several greenhouses gases).

Paris Agreement: The UN and WHO emphasize the link between air quality and climate change, stating that “climate change is intricately tied to air quality” (Sarmad, 2018). Thus, by providing a global platform for access to multisource air quality data, this project will provide an important tool for assessing climate change drivers and impacts as well. This will be especially important for developing nations, which do not have access to dense networks for collecting local air quality data and therefore must rely on other sources of information to track greenhouse gases.

Sendai Framework for Disaster Risk Reduction: The Sendai Framework for Disaster Risk Reduction includes “focusing on monitoring, assessing and understanding disaster risk and sharing such information”; this project proposes a platform for freely disseminating information on the risks of air pollution. The framework also proposes to “substantially increase the availability of and access to multihazard early warning systems and disaster risk information and assessments to people by 2030”; this project proposes a form of local air pollution hazard information system, combined with near term forecasts for early warning. The frameworks also advocates accounting for the needs of different classes of users of hazard data, as well as promoting real time access to space based and in situ data and utilizing advances in information and communications technology to improve data analysis and dissemination. This project develops such a system for data analysis and dissemination and will incorporate feedback from different users in the development of its data visualization tools.

AfriGEOSS: Through coordination with GEOSS, this project will contribute to the goals of AfriGEOSS by increasing accessibility for earth observation data in Africa and by combining it with other data sources. Also, coordinated in situ observations networks in Africa, especially those involving low-cost sensors for previously unmonitored or undermonitored regions, will be facilitated by this program, as their data will also be integrated into the proposed system.

References

Abel, D.W., Holloway, T., Harkey, M., Meier, P., Ahl, D., Limaye, V.S., Patz, J.A., 2018. Air-quality-related health impacts from climate change and from adaptation of cooling demand for buildings in the eastern United States: An interdisciplinary modeling study. PLOS Medicine 15, e1002599. https://doi.org/10.1371/journal.pmed.1002599

Apte, J.S., Marshall, J.D., Cohen, A.J., Brauer, M., 2015. Addressing Global Mortality from Ambient PM2.5. Environmental Science & Technology 49, 8057–8066. https://doi.org/10.1021/acs.est.5b01236

AQ-SPEC, 2017. PurpleAir PA-II Sensor Evaluation Report, Air Quality Sensor Performance Evaluation Center. South Coast Air Quality Management District.

Bell, M.L., Ebisu, K., Belanger, K., 2007. Ambient Air Pollution and Low Birth Weight in Connecticut and Massachusetts. Environmental Health Perspectives 115, 1118–1124. https://doi.org/10.1289/ehp.9759

Berger, J.O., 1993. Statistical decision theory and Bayesian analysis, 2nd ed, Springer series in statistics. Springer-Verlag, New York.

Bergerson, J., Muehleisen, R., 2015. Bayesian Large Model Calibration Using Simulation and Measured Data for Improved Predictions. SAE International Journal of Passenger Cars - Mechanical Systems 8. https://doi.org/10.4271/2015-01-0481

Bond, T.C., Doherty, S.J., Fahey, D.W., Forster, P.M., Berntsen, T., B. J. DeAngelo, M. G. Flanner, S. Ghan, B. Kaercher, D. Koch, S. Kinne, Y. Kondo, P. K. Quinn, M. C. Sarofim, M. G. Schultz, M. Schulz, C. Venkataraman, H. Zhang, S. Zhang, N. Bellouin, S. K. Guttikunda, P. K. Hopke, M. Z. Jacobson, J. W. Kaiser, Z. Klimont, U. Lohmann, J. P. Schwarz, D. Shindell, T. Storelvmo, S. G. Warren, C. S. Zender, 2013. Bounding the role of black carbon in the climate system: A scientific assessment. Journal of Geophysical Research-Atmospheres 118, 5380–5552.

Brauer, M., Amann, M., Burnett, R.T., Cohen, A., Dentener, F., Ezzati, M., Henderson, S.B., Krzyzanowski, M., Martin, R.V., Van Dingenen, R., van Donkelaar, A., Thurston, G.D., 2012. Exposure Assessment for Estimation of the Global Burden of Disease Attributable to Outdoor Air Pollution. Environmental Science & Technology 46, 652–660. https://doi.org/10.1021/es2025752

Brook, R.D., Rajagopalan, S., Pope, C.A., Brook, J.R., Bhatnagar, A., Diez-Roux, A.V., Holguin, F., Hong, Y., Luepker, R.V., Mittleman, M.A., Peters, A., Siscovick, D., Smith, S.C., Whitsel, L., Kaufman, J.D., on behalf of the American Heart Association Council on Epidemiology and Prevention, Council on the Kidney in Cardiovascular Disease, and Council on Nutrition, Physical Activity and Metabolism, 2010. Particulate Matter Air Pollution and Cardiovascular Disease: An Update to the Scientific Statement From the American Heart Association. Circulation 121, 2331–2378. https://doi.org/10.1161/CIR.0b013e3181dbece1

Castanedo, F., 2013. A Review of Data Fusion Techniques. The Scientific World Journal 2013, 1–19. https://doi.org/10.1155/2013/704504

Cressie, N.A.C., Wikle, C.K., 2011. Statistics for spatio-temporal data, Wiley series in probability and statistics. Wiley, Hoboken, N.J.

Cross, E.S., Williams, L.R., Lewis, D.K., Magoon, G.R., Onasch, T.B., Kaminsky, M.L., Worsnop, D.R., Jayne, J.T., 2017. Use of electrochemical sensors for measurement of air pollution: correcting interference response and validating measurements. Atmospheric Measurement Techniques 10, 3575–3588. https://doi.org/10.5194/amt-10-3575-2017

Cuesta, J., Eremenko, M., Liu, X., Dufour, G., Cai, Z., Hopfner, M., von Clarmann, T., Sellitto, P., Foret, G., Gaubert, M., Beekmann, M., Orphal, J., Chance, K., Spurr, R., Flaud, J.-M., 2013. Satellite observation of lowermost tropospheric ozone by multispectral synergism of IASI thermal infrared and GOME-2 ultraviolet measurements. Atmos. Chem. Phys. 13, 9675–9693.

DeWitt, H.L., Gasore, J., 2018. The Rwanda Climate Observatory: Developing climate science in East Africa. Clean Air Journal 28. https://doi.org/10.17159/2410-972x/2018/v28n2a15

Di, Q., Wang, Yan, Zanobetti, A., Wang, Yun, Koutrakis, P., Choirat, C., Dominici, F., Schwartz, J.D., 2017. Air Pollution and Mortality in the Medicare Population. New England Journal of Medicine 376, 2513–2522. https://doi.org/10.1056/NEJMoa1702747

Engel-Cox, J.A., Holloman, C.H., Coutant, B.W., Hoff, R.M., 2004. Qualitative and quantitative evaluation of MODIS satellite sensor data for regional and urban scale air quality. Atmospheric Environment 38, 2495–2509. https://doi.org/10.1016/j.atmosenv.2004.01.039

English, P.B., Olmedo, L., Bejarano, E., Lugo, H., Murillo, E., Seto, E., Wong, M., King, G., Wilkie, A., Meltzer, D., Carvlin, G., Jerrett, M., Northcross, A., 2017. The Imperial County Community Air Monitoring Network: A Model for Community-based Environmental Monitoring for Public Health Action. Environ Health Perspect 125, 074501. https://doi.org/10.1289/EHP1772

Fountoukis, C., Megaritis, A.G., Skyllakou, K., Charalampidis, P.E., van der Gon, H.A.D.C., Crippa, M., Prevot, A.S.H., Fachinger, F., Wiedensohler, A., Pilinis, C., Pandis, S.N., 2016. Simulating the formation of carbonaceous aerosol in a European Megacity (Paris) during the MEGAPOLI summer and winter campaigns. Atmos. Chem. Phys. 26, 3727–3741.

Hagan, D.H., Isaacman-VanWertz, G., Franklin, J.P., Wallace, L.M.M., Kocar, B.D., Heald, C.L., Kroll, J.H., 2018. Calibration and assessment of electrochemical air quality sensors by co-location with regulatory-grade instruments. Atmospheric Measurement Techniques 11, 315–328. https://doi.org/10.5194/amt-11-315-2018

Han, W., Tong, L., Chen, Y., Li, R., Yan, B., Liu, X., 2018. Estimation of High-Resolution Daily Ground-Level PM2.5 Concentration in Beijing 2013–2017 Using 1 km MAIAC AOT Data. Applied Sciences 8, 2624. https://doi.org/10.3390/app8122624

Heft-Neal, S., Burney, J., Bendavid, E., Burke, M., 2018. Robust relationship between air quality and infant mortality in Africa. Nature 559, 254–258. https://doi.org/10.1038/s41586-018-0263-3

Hoek, G., Beelen, R., de Hoogh, K., Vienneau, D., Gulliver, J., Fischer, P., Briggs, D., 2008. A review of land-use regression models to assess spatial variation of outdoor air pollution. Atmospheric Environment 42, 7561–7578. https://doi.org/10.1016/j.atmosenv.2008.05.057

Jerrett, M., Arain, A., Kanaroglou, P., Beckerman, B., Potoglou, D., Sahsuvaroglu, T., Morrison, J., Giovis, C., 2005a. A review and evaluation of intraurban air pollution exposure models. Journal of Exposure Science & Environmental Epidemiology 15, 185–204. https://doi.org/10.1038/sj.jea.7500388

Jerrett, M., Burnett, R.T., Ma, R., Pope, C.A., Krewski, D., Newbold, K.B., Thurston, G., Shi, Y., Finkelstein, N., Calle, E.E., Thun, M.J., 2005b. Spatial Analysis of Air Pollution and Mortality in Los Angeles. Epidemiology 16, 727–736. https://doi.org/10.1097/01.ede.0000181630.15826.7d

Karner, A.A., Eisinger, D.S., Niemeier, D.A., 2010. Near-Roadway Air Quality: Synthesizing the Findings from Real-World Data. Environmental Science & Technology 44, 5334–5344. https://doi.org/10.1021/es100008x

Liousse, C., Assamoi, E., Criqui, P., Granier, C., Rosset, R., 2014. Explosive growth in African combustion emissions from 2005 to 2030. Environmental Research Letters 9, 035003. https://doi.org/10.1088/1748-9326/9/3/035003

Liousse, C., Guillaume, B., Grégoire, J.M., Mallet, M., Galy, C., Pont, V., Akpo, A., Bedou, M., Castéra, P., Dungall, L., Gardrat, E., Granier, C., Konaré, A., Malavelle, F., Mariscal, A., Mieville, A., Rosset, R., Serça, D., Solmon, F., Tummon, F., Assamoi, E., Yoboué, V., Van Velthoven, P., 2010. Updated African biomass burning emission inventories in the framework of the AMMA-IDAF program, with an evaluation of combustion aerosols. Atmospheric Chemistry and Physics 10, 9631–9646. https://doi.org/10.5194/acp-10-9631-2010

Loh, M., Sarigiannis, D., Gotti, A., Karakitsios, S., Pronk, A., Kuijpers, E., Annesi-Maesano, I., Baiz, N., Madureira, J., Oliveira Fernandes, E., Jerrett, M., Cherrie, J., 2017. How Sensors Might Help Define the External Exposome. International Journal of Environmental Research and Public Health 14, 434. https://doi.org/10.3390/ijerph14040434

Maddox, W., Garipov, T., Izmailov, P., Vetrov, D., Wilson, A.G., 2019. A Simple Baseline for Bayesian Uncertainty in Deep Learning. arXiv:1902.02476 [cs, stat].

Malings, C., Pozzi, M., Klima, K., Bergés, M., Bou-Zeid, E., Ramamurthy, P., 2018. Surface heat assessment for developed environments: Optimizing urban temperature monitoring. Building and Environment 141, 143–154. https://doi.org/10.1016/j.buildenv.2018.05.059

Malings, C., Pozzi, M., Klima, K., Bergés, M., Bou-Zeid, E., Ramamurthy, P., 2017. Surface heat assessment for developed environments: Probabilistic urban temperature modeling. Computers, Environment and Urban Systems 66, 53–64. https://doi.org/10.1016/j.compenvurbsys.2017.07.006

Malings, C., Tanzer, R., Hauryliuk, A., Kumar, S.P.N., Zimmerman, N., Kara, L.B., Presto, A.A., R Subramanian, 2019a. Development of a general calibration model and long-term performance evaluation of low-cost sensors for air pollutant gas monitoring. Atmospheric Measurement Techniques 12, 903–920. https://doi.org/10.5194/amt-12-903-2019

Malings, C., Tanzer, R., Hauryliuk, A., Saha, P.K., Robinson, A.L., Presto, A.A., Subramanian, R., 2019b. Fine particle mass monitoring with low-cost sensors: Corrections and long-term performance evaluation. Aerosol Science and Technology 1–15. https://doi.org/10.1080/02786826.2019.1623863

Marshall, J.D., Nethery, E., Brauer, M., 2008. Within-urban variability in ambient air pollution: Comparison of estimation methods. Atmospheric Environment 42, 1359–1369. https://doi.org/10.1016/j.atmosenv.2007.08.012

Masson, N., Piedrahita, R., Hannigan, M., 2015. Quantification Method for Electrolytic Sensors in Long-Term Monitoring of Ambient Air Quality. Sensors 15, 27283–27302. https://doi.org/10.3390/s151027283

Matshidiso Moeti, 2018. Cleaning up Africa's air would pay for itself in economic gains: Pollution is dragging down the continent's GDP and harming its children. Financial Times.

Mead, M.I., Popoola, O.A.M., Stewart, G.B., Landshoff, P., Calleja, M., Hayes, M., Baldovi, J.J., McLeod, M.W., Hodgson, T.F., Dicks, J., Lewis, A., Cohen, J., Baron, R., Saffell, J.R., Jones, R.L., 2013. The use of electrochemical sensors for monitoring urban air quality in

low-cost, high-density networks. Atmospheric Environment 70, 186–203. https://doi.org/10.1016/j.atmosenv.2012.11.060
Menut, L., Bessagnet, B., Khvorostyanov, D., Beekmann, M., Blond, N., Colette, A., Coll, I., Curci, G., Foret, G., Hodzic, A., Mailler, S., Meleux, F., Monge, J.-L., Pison, I., Siour, G., Turquety, S., Valari, M., Vautard, R., Vivanco, M.G., 2013. CHIMERE 2013: a model for regional atmospheric composition modelling. Geoscientific Model Development 6, 981–1028. https://doi.org/10.5194/gmd-6-981-2013
Naidja, L., Ali-Khodja, H., Khardi, S., 2018. Sources and levels of particulate matter in North African and Sub-Saharan cities: a literature review. Environmental Science and Pollution Research 25, 12303–12328. https://doi.org/10.1007/s11356-018-1715-x
Petkova, E.P., Jack, D.W., Volavka-Close, N.H., Kinney, P.L., 2013. Particulate matter pollution in African cities. Air Quality, Atmosphere & Health 6, 603–614. https://doi.org/10.1007/s11869-013-0199-6
Pope, C.A., Burnett, R.T., Thun, M.J., Calle, E.E., Krewski, D., Ito, K., Thurston, G.D., 2002. Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. JAMA 287, 1132–1141.
Popoola, O.A.M., Stewart, G.B., Mead, M.I., Jones, R.L., 2016. Development of a baseline-temperature correction methodology for electrochemical sensors and its implications for long-term stability. Atmospheric Environment 147, 330–343. https://doi.org/10.1016/j.atmosenv.2016.10.024
REMA, 2018. Inventory of Sources of Air Pollution in Rwanda: Determination of Future Trends and Development of a National Air Quality Control Strategy (No. 382754 | 3 | E). Rwanda Environment Management Authority.
Ren, Z., Zhu, J., Gao, Y., Yin, Q., Hu, M., Dai, L., Deng, C., Yi, L., Deng, K., Wang, Y., Li, X., Wang, J., 2018. Maternal exposure to ambient PM10 during pregnancy increases the risk of congenital heart defects: Evidence from machine learning models. Science of The Total Environment 630, 1–10. https://doi.org/10.1016/j.scitotenv.2018.02.181
Sadighi, K., Coffey, E., Polidori, A., Feenstra, B., Lv, Q., Henze, D.K., Hannigan, M., 2018. Intra-urban spatial variability of surface ozone in Riverside, CA: viability and validation of low-cost sensors. Atmospheric Measurement Techniques 11, 1777–1792. https://doi.org/10.5194/amt-11-1777-2018
Sally Liu, L.-J., Tsai, M.-Y., Keidel, D., Gemperli, A., Ineichen, A., Hazenkamp-von Arx, M., Bayer-Oglesby, L., Rochat, T., Künzli, N., Ackermann-Liebrich, U., Straehl, P., Schwartz, J., Schindler, C., 2012. Long-term exposure models for traffic related NO2 across geographically diverse areas over separate years. Atmospheric Environment 46, 460–471. https://doi.org/10.1016/j.atmosenv.2011.09.021
Schneider, P., Castell, N., Vogt, M., Dauge, F.R., Lahoz, W.A., Bartonova, A., 2017. Mapping urban air quality in near real-time using observations from low-cost sensors and model information. Environment International 106, 234–247. https://doi.org/10.1016/j.envint.2017.05.005
Schwartz, J., Dockery, D.W., Neas, L.M., 1996. Is daily mortality associated specifically with fine particles? J Air Waste Manag Assoc 46, 927–939.
Silva, R.A., West, J.J., Lamarque, J.-F., Shindell, D.T., Collins, W.J., Faluvegi, G., Folberth, G.A., Horowitz, L.W., Nagashima, T., Naik, V., Rumbold, S.T., Sudo, K., Takemura, T., Bergmann, D., Cameron-Smith, P., Doherty, R.M., Josse, B., MacKenzie, I.A., Stevenson,

D.S., Zeng, G., 2017. Future global mortality from changes in air pollution attributable to climate change. Nature Climate Change 7, 647.
Snyder, E.G., Watkins, T.H., Solomon, P.A., Thoma, E.D., Williams, R.W., Hagler, G.S.W., Shelow, D., Hindin, D.A., Kilaru, V.J., Preuss, P.W., 2013. The Changing Paradigm of Air Pollution Monitoring. Environmental Science & Technology 47, 11369–11377. https://doi.org/10.1021/es4022602
Spinelle, L., Gerboles, M., Villani, M.G., Aleixandre, M., Bonavitacola, F., 2015. Field calibration of a cluster of low-cost available sensors for air quality monitoring. Part A: Ozone and nitrogen dioxide. Sensors and Actuators B: Chemical 215, 249–257. https://doi.org/10.1016/j.snb.2015.03.031
Subramanian, R., Ellis, A., Torres-Delgado, E., Tanzer, R., Malings, C., Rivera, F., Morales, M., Baumgardner, D., Presto, A., Mayol-Bracero, O.L., 2018. Air Quality in Puerto Rico in the Aftermath of Hurricane Maria: A Case Study on the Use of Lower Cost Air Quality Monitors. ACS Earth and Space Chemistry 2, 1179–1186. https://doi.org/10.1021/acsearthspacechem.8b00079
Tan, Y., Lipsky, E.M., Saleh, R., Robinson, A.L., Presto, A.A., 2014. Characterizing the Spatial Variation of Air Pollutants and the Contributions of High Emitting Vehicles in Pittsburgh, PA. Environmental Science & Technology 48, 14186–14194. https://doi.org/10.1021/es5034074
Turner, M.C., Nieuwenhuijsen, M., Anderson, K., Balshaw, D., Cui, Y., Dunton, G., Hoppin, J.A., Koutrakis, P., Jerrett, M., 2017. Assessing the Exposome with External Measures: Commentary on the State of the Science and Research Recommendations. Annual Review of Public Health 38, 215–239. https://doi.org/10.1146/annurev-publhealth-082516-012802
UNEP, 2016. Global Assessment of Sand and Dust Storms. United Nations Environment Programme ; World Meteorological Organization (WMO) ; United Nations Convention to Combat Desertification.
van Donkelaar, A., Martin, R.V., Brauer, M., Boys, B.L., 2015. Use of Satellite Observations for Long-Term Exposure Assessment of Global Concentrations of Fine Particulate Matter. Environmental Health Perspectives 123, 135–143. https://doi.org/10.1289/ehp.1408646
Waltz, E., Llinas, J., 1990. Multisensor data fusion, Artech House radar library. Artech House, Boston.
WHO, 2018a. Ambient (outdoor) air quality and health. World Health Organization.
WHO, 2018b. WHO ambient (outdoor) air quality database Summary results, update 2018. Public Health, Social and Environmental Determinants of Health Department, World Health Organization.
WHO, 2017. Global Health Observatory (GHO) data: Exposure to ambient air pollution. World Health Organization.
WHO, 2016. Ambient air pollution: a global assessment of burden and disease. World Health Organization.
WHO, 2006. WHO Air quality guidelinesfor particulate matter, ozone, nitrogendioxide and sulfur dioxide: Summary of risk assessment (No. WHO/SDE/PHE/OEH/06.02). World Health Organization.
World Bank, 2016. The Cost of Air Pollution: Strengthening the Economic Case for Action. The World Bank and Institute for Health Metrics and Evaluation, University of Washington, Seattle.

Zimmerman, N., Presto, A.A., Kumar, S.P.N., Gu, J., Hauryliuk, A., Robinson, E.S., Robinson, A.L., R Subramanian, 2018. A machine learning calibration model using random forests to improve sensor performance for lower-cost air quality monitoring. Atmospheric Measurement Techniques 11, 291–313. https://doi.org/10.5194/amt-11-291-2018