CLEAN AIR FUND

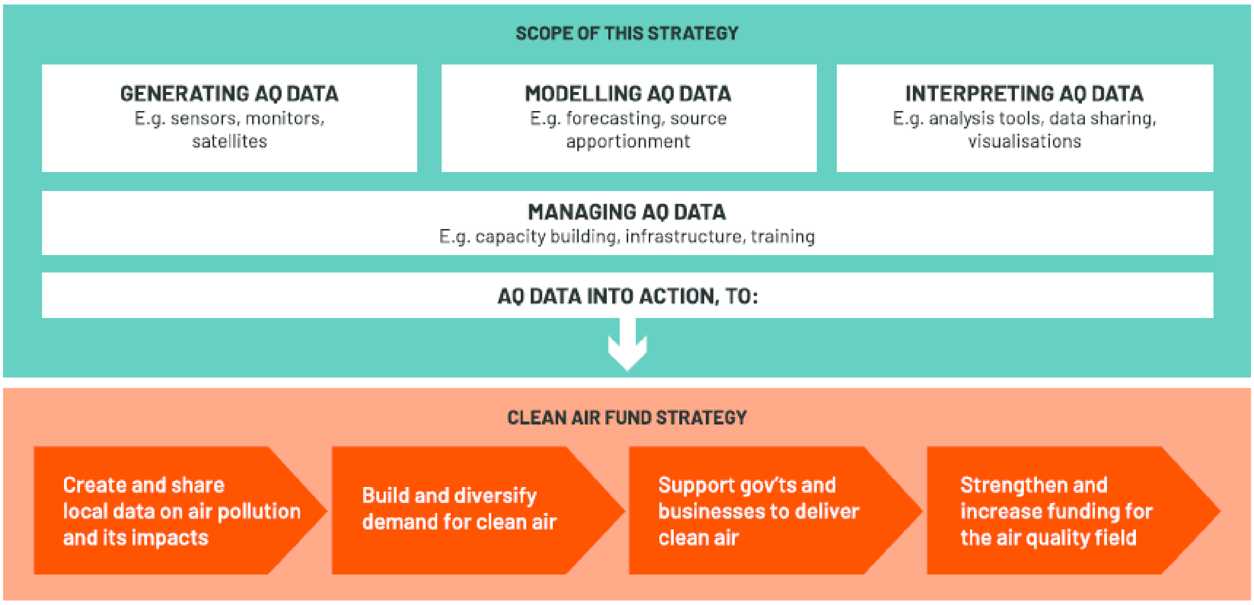
CLEAN AIR FUND AIR QUALITY DATA STRATEGY 2021 – 2024

ANNEX: ASSESSMENT OF THE CURRENT SITUATION

CHILDREN'S INVESTMENT FUND FOUNDATION
Bloomberg Philanthropies
IKEA Foundation
FIA FOUNDATION
Bernard van Leer FOUNDATION
Impact on Urban Health
OAK FOUNDATION
the ELMA foundation

Context

This document is the accompanying annex to the Clean Air Fund Data Strategy, viewable here at cleanairfund.org/publication/data-strategy. It provides the detailed review of the sector covering the full scope of air quality data, which informed the production of our strategy:



Our thanks to the many individuals that contributed and inputted to this document.

Contents

Air quality data: Current situation

Air quality data holds the key to understanding the lifecycle of air pollutants – and therefore to engaging with the problem and to forming solutions.	Air quality data provides the information required to understand, engage with, and develop solutions to poor air quality. The lifecycle of an air pollutant encompasses its emission, dispersion, reaction and removal. Air quality data in this context, can be broken down into two forms: 1) **Concentration data -** quantifies pollution levels (air quality) over some space and time. Concentration data is the amalgamation of individual pollutant particles at different stages of their lifecycles from a number of sources at a variety of scales. This information can tell us about the quality of the air we breathe but, alone, does not tell us about how and why the air quality is as it is. 2) **Emissions data** – quantifies the locations, time-dependencies and magnitudes of the sources of air pollution. This includes both primary information (e.g. traffic counts) and secondary information (e.g. source attribution of traffic emissions). Both forms of data are necessary to understand and combat poor air quality and the two are inherently linked. This section summarises the current state of the air quality data sector, structured around five key stages of air quality management: generation, modelling, interpretation, managing and actioning of data. The Clean Air Fund's overarching strategy focuses on the abatement of poor ambient air quality. The primary focus of this strategy is therefore on the first stage of this lifecycle: emissions (as opposed to removal; air purification, sequestration etc.).

1 Generating data

There is a wide set of tools available to monitor air quality. There is no 'right' method. Different monitor types support different applications, driven primarily by both <u>cost</u> (fig. 2.1) and <u>accuracy</u> (fig. 2.2)	Air quality monitoring is the deployment of sensors to provide a measure of air pollution levels. There are several available technologies that calculate pollutant concentrations over a range of spatial and temporal scales (fig. 2.1). Key factors affecting the application of these technologies are: • Life-cycle cost (hardware, software, installation, calibration, operation, and maintenance) • Performance (accuracy, precision, quality, and reliability) • Resolution (temporal and spatial) • Breadth (e.g. number of pollutants) • Usability (size, mobility, functionality)
Concentration data - pollutant concentrations vary spatially and temporally and monitoring techniques must account for this to produce usable data	**The spatial and temporal resolution of the monitoring technologies is a key attribute to consider**. A single data point from different monitoring technologies (see fig. 2.1) provides an average concentration over a certain space and over a certain time period. For example, a single stationary monitor provides an indication of the temporal change of concentrations at one given position. This provides useful information on what the air quality is in that position but, alone, little else. Pollutants derive from a number of sources and are spatially and temporally heterogenous at a range of scales. Two questions arise: 1) what is the spatial and temporal variation of the pollutant at the required scale? and 2) where is the pollution coming from (and where is it going)? The former can be obtained to some extent by denser networks of monitors (e.g. hyperlocal using small sensors) and satellites. The latter requires more data (e.g. emissions inventories and source attribution) and modelling (see section 1.2).

Fig. 2.1: Monitoring approaches by cost and application scale

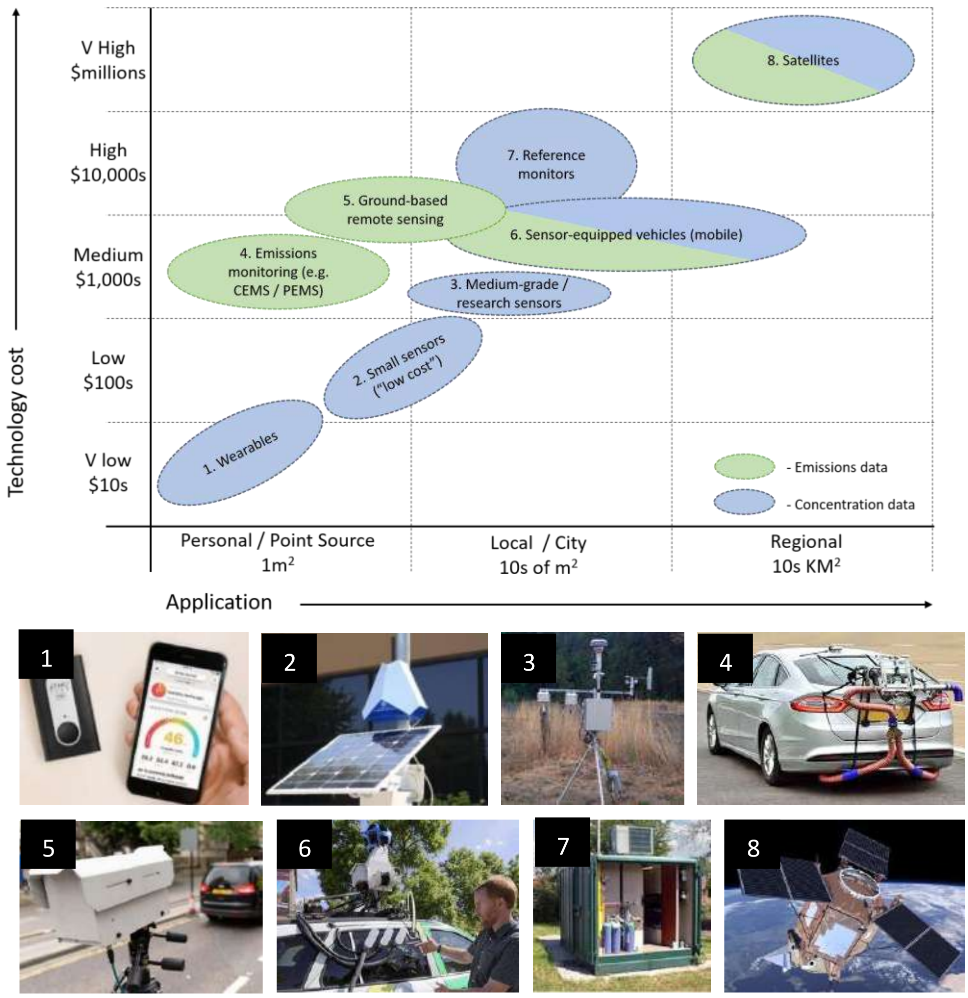
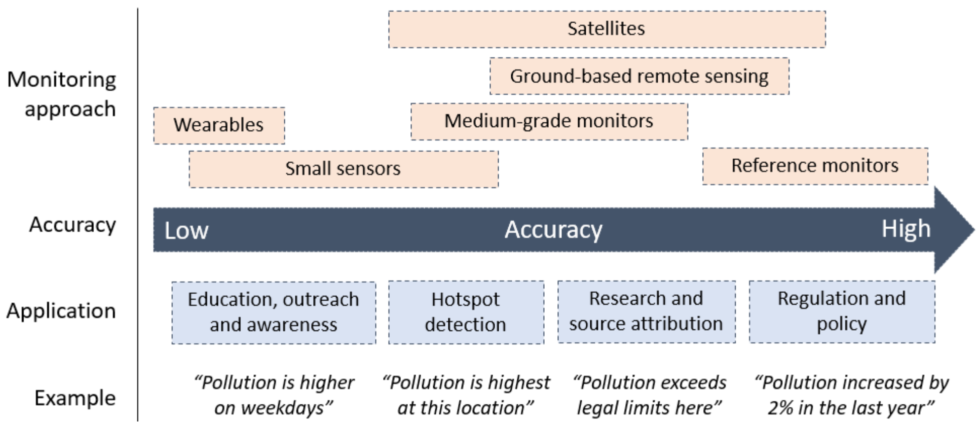


Fig. 2.2 – Indicative accuracy and appropriate applications of different monitoring technologies



Emissions data - Emissions inventories collate the latest data on the sources of air pollution

Emissions inventories are generally the end goal in terms of emissions-based data generation. These inventories span across sectors (from cooking to aviation) to provide a holistic view on the sources of air pollution over a given region. The breadth and scope of these datasets means that, by definition, require numerous, multi-sectorial forms of data generation.

Dozens of inventories have been developed at global and regional levels, built by research institutes, international organisations and academia[1]. These provide baselines of emissions activity and are applicable to air quality studies focussing across large regions, owing to their relatively

[1] See Annex 5 of Vital Strategies' *Accelerating City Progress on Clean Air* for a detailed list of available air pollutant emissions inventories.

	coarse resolution (typically >0.1° grid size, or approximately 11km^2). National and smaller scale inventories exist in some countries, such as the UK (National and London Atmospheric Emissions Inventories), but are lacking in many others. **Machine learning is an emerging technology that has the potential to alleviate some of the challenges of emissions data generation**. For example, earth observation using satellite images and traffic cameras in cities could be used to automate the identification of source types. The spatial and temporal resolution of these inventories is a key consideration.
Existing methods of air quality monitoring have failed to scale quickly enough. Vast numbers of people lack access to any locally relevant air quality data. *Fig. 2.3 – The distribution of government and research grade air quality data that is openly accessible. Each dot represents a data source*[i]*.*	**Until recently, air quality monitoring was conducted almost exclusively by governments, regulators, and researchers**, using networks of well-established and generally high accuracy monitors (i.e. research grade and reference grade monitors). Networks of reference grade monitors, coupled with technical expertise to analyse and model the measurements, is considered the 'gold-standard' as it provides the highest quality data which can be used for regulatory purposes, such as to demonstrate compliance with air quality standards. **Relying solely on reference monitoring is costly, requires deep expertise and is out of reach for many governments**. It is estimated that the global system of reference monitoring has cost in excess of $250 million in capital equipment alone (approximately 10,000 stations). This does not include ongoing quality control, dissemination, staff time, maintenance or analysis costs, which are likely to be much higher. Lower capacity countries that do operate reference monitors may not have the resources to maintain them, leading to poor quality or inconsistent data. **As a result, a significant proportion of the world's population are left without any air quality data**, particularly low-income countries, (fig. 2.3): - Almost half of the world's governments produce no publicly accessible reference-grade air quality data at all[ii], impacting a total population of 1.4 billion people. - If the governments of just 13 countries invested in air quality monitoring programmes it could bring data to 1 billion people. - Only 38% of governments share real-time[2] air quality data in some capacity, even if not in a fully open form. The remaining 62% of countries represent a total population of 2.1 billion people. - At least 30 governments generate real-time data but do not share it in a fully open manner. Making this data public would provide open real-time data to 4.4 billion people. 
New technologies bring significant potential to fill data gaps	Reference-grade sensors have an important role for monitoring air quality, but alone they cannot fill the growing demand for data. Filling data gaps will rely on the deployment of different technologies in combination with existing methods. Two technologies in particular – small sensors and satellites – are widely cited as potential game-changers for filling data gaps.
Small sensors have the potential to greatly increase the amount of local monitoring, but there are challenges	**Small sensors (often referred to as 'low cost sensors') are smaller, lighter and easier-to-use compared to reference methods**, and have seen a rapid rise in prominence in recent years. Pros of small sensors: - **Low capital costs:** ranging from $100s to a few $1000s per unit, which is considerably lower than research grade sensors ($5 – 15k per instrument) or reference monitors ($15k+ per instrument). The wide variation in the costs between small sensors is driven by the number of pollutants measured, the amount of meta-data provided (e.g. GPS location), data transfer capability (e.g.

[2] Real time data is important as it promotes accountability to government policy and enables reactive responses to protect populations from high pollution, such as temporary traffic measures or public information notices.

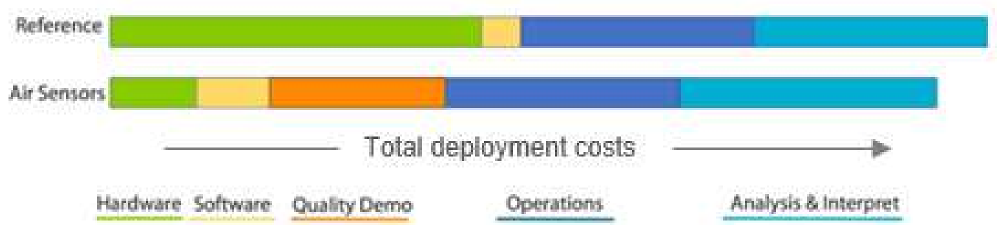
with accuracy and reliability

WiFi, Bluetooth, SIM) and the level of sophistication for controlling environmental variables (e.g. temperature and humidity).
- **Low power requirements**: can often be solar or battery powered, making them suitable for areas with unreliable electricity or for quick installation without accessing mains power.
- **Suitable for a range of monitoring approaches**: their small size and weight (generally a few kilos or less) allow them to be used in fixed, mobile or personal monitoring settings.
- **Rapidly developing** with performance continually improving resulting from a highly competitive market that fosters fast evolution in the technology[iv].

Cons of small sensors:

- **Unreliable accuracy and precision**[3]: the quality of data from small sensors can be inconsistent, both in accuracy and precision, with even the same product often providing different measurements in the same location, particularly in poorer quality units.
- **Reliability:** instruments degrade over time (rapidly in harsh environments) requiring ongoing maintenance and recalibration. Harsh conditions also compound sensor performance (particularly in high humidity, dusty areas or very high temperatures). The life expectancy of a small air quality sensor is highly variable depending on the environment in which it is installed. Even in favourable environments sensors are likely to fail after approximately two years of continued use.
- **Variable product quality**: there are currently no requirements for companies to publish performance data or demonstrate compliance with standards, resulting in a highly inconsistent market in terms of product performance.
- **Inconsistent performance across pollutants**: small sensors are generally more accurate at measuring particulate matter compared to gaseous pollutants, as the latter depends more significantly on pre-calibration (before use) and regular recalibrations (during use) to maintain accuracy[v].
- **Overpromised performance**: Small sensors can mistakenly be seen as 'plug and play' monitoring solutions, without requiring quality assurance or quality control (QA/QC) such as calibration, ongoing maintenance or data treatment. This may be true for some applications with low accuracy requirements, such as citizen science or education, but QA/QC becomes more important as both the numbers of monitors in a network increases (due to the need to provide comparability between individual sensors) and as the application of the data broadens from education towards decision making (see fig. 2.2).
- **Underestimated costs**: As QA/QC requirements grow, so too do the costs and expertise needed to successfully deploy and operate small sensors. Whilst hardware costs are much lower than regulatory monitoring, operating costs can mean total network costs can become comparable to maintaining reference grade sensors (fig. 2.4)[vi].

Fig. 2.4 – Indicative breakdown of deployment costs for reference and small sensors

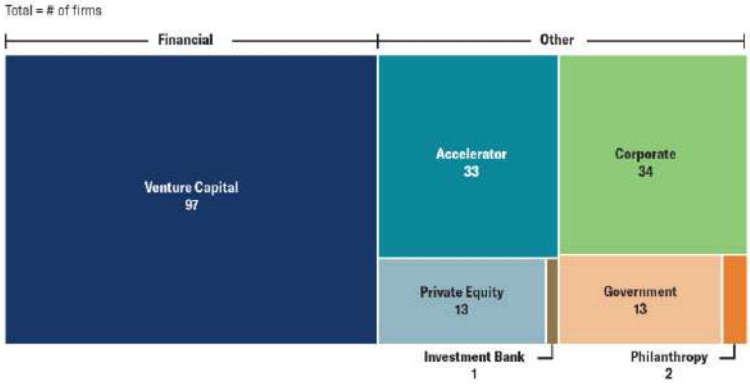


Applications for small sensors:

- **Governments and regulators** widely use small sensors to supplement regulatory networks for identifying hot-spots and optimising placement of higher cost monitors.
- **NGOs and citizens** use them for communication, education and advocacy purposes.

The market for small sensors is growing rapidly. At its height in 2019, a new small sensor manufacturer was entering the market at a rate of approximately one a week[vii]. Between 2015 and 2018, investment in air quality monitoring technology grew by 17.6% per annum, driven by venture

[3] Accuracy is how close a measurement is to its true value. Precision is how repeatable a measurement is.

Fig. 2.5 – A breakdown of global investors in air quality monitoring[iii]	capital (fig. 2.5), which suggests most of the investment went into developing new technologies like small sensors. 

Satellites have the potential to provide global-scale monitoring but at the expense of granular spatial resolution	**Satellites have been monitoring air pollution since the early 2000s, but data volume and quality has improved considerably in recent years with the launch of new satellites.**
	Pros of satellite monitoring:
	- **Geographic coverage:** ability to monitor nearly the entire planet, collecting data about previously unmonitored areas.
	- **Consistency of measurement:** ability for the same sensor to make measurements everywhere, providing comparability of measurements across the globe.
	- **Rapidly developing:** several launches are planned by the US, EU and South Korean space agencies which will continue to improve data quality and spatial and temporal resolution[viii].
	Cons of satellite monitoring:
	- **Coarse spatial resolution:** often at the kilometre+ level which limits the application of satellite data to assessing pollution at regional scales. Satellite derived data products are available at sub-kilometre level for NO_2, but applications sit largely within academia. Satellites capable of higher resolution are still a few years from being launched and the full applications of its data are yet to be demonstrated.
	- **Costly and complex:** instruments and data products have high development and computational costs, requiring considerable data processing to produce a usable dataset.
	- **Unable to directly monitor ground-level pollution or exposure:** a satellite can only measure the concentration of pollutants across the whole 'column' of atmosphere between the ground and the satellite sensor, and therefore need to be combined with models to estimate the concentration at ground-level (the level at which people breathe the air).
	- **Unreliable in certain conditions:** Cannot measure at night (for the current generation of satellites) and unreliable over cloudy, snowy or dusty conditions, or across desert and mountainous terrain.
	Applications for air quality satellite monitoring:
	- Satellite data is largely used by governments, researchers or technical NGOs to:
	    o **Provide estimates of pollution concentration at a regional or global level**, which has provided the evidence needed to assess the global burden of disease from air pollution and especially in low- and middle-income countries that lack local monitoring.
	    o **Provide comparisons and trends** of air quality levels over large regions.
	    o **Provide information about transboundary pollution.**
	    o **Evaluate air quality models.**
	The market for satellite monitoring of air quality is difficult to quantify as satellite data is used often in secondary applications (i.e. data is accessed for free and integrated with other data sources to produce a product). However, the number of satellite instruments being launched continues to grow, demonstrating that demand for data products is also increasing.

CLEAN AIR FUND

Technologies to monitor point sources of emissions are important for policy assessment and enforcement.	**Large point sources of pollution like power stations or factories often need to comply with regulated emission limits.** The technology employed to measure emissions at their source (rather than concentrations in ambient air) is known as Continuous Monitoring Systems (CMS). Expanding the use of CMS technology is largely a political challenge: a country can relatively easily mandate that industries bear the costs of installing and maintaining equipment to monitor their emissions. The challenge for many countries is enforcement (in India, for example[ix]). **Vehicle emissions – point sources which at city-wide scales are important contributors to pollution exposure – can be directly measured through two complementary methods:** - **Ground-based remote sensing:** which employs similar techniques to satellite monitoring at the roadside to identify pollution levels from vehicle exhaust plumes. A single measurement campaign using this approach can characterise pollution levels from a fleet of thousands of vehicles. - **Portable Emissions Measurement System (PEMS):** similar technology to a CMS but attached to a vehicle to measure emissions under a range of different driving conditions (see picture 4 in Fig 2.2). This approach provides an accurate characterisation of emissions from a particular vehicle model. Both techniques were instrumental in exposing the Dieselgate scandal, a highly influential moment in modern air pollution policy and its enforcement.
As the adoption of new monitoring techniques accelerates, so too does the need for guidance and support to ensure consistent monitoring that is fit-for-purpose	**Traditional monitoring techniques have established guidance:** The use of reference grade sensors is based on well-developed protocols and operating procedures, owing to their use for regulatory compliance. For example, the EPA maintains a detailed list of training materials and QA/QC procedures[x]. In most countries, reference-grade monitors and the data they produce must conform to certain performance standards or certification before being accepted, ensuring consistent data quality. **Newer monitoring technologies have much less detailed or organised guidance:** Information is increasing – governments[xi], NGOs[xii] and international organisations[xiii] have produced a range of supporting materials, especially for the deployment of small sensors – but currently there is a fragmented and incomplete approach to maintenance of the sensors and use of the data (e.g. guidance on QA/QC, data management, data corrections). **There is currently no certification, standardisation or kite mark for small sensors, although several schemes are in development**. Both the Government of India and the US EPA are developing performance standards criteria due for publication in 2020. The European Union has a working group for small sensor standardisation. Depending on how standards are implemented, it could potentially set a minimum performance standard for small sensors reaching the market. **There are several evaluation programmes which test sensor performance in the lab or field to help inform users of the performance of different products**[4]. However, there are issues: there is no set approach to these evaluations making it difficult to compare between studies; most evaluations take place in western conditions so the results may not be applicable elsewhere; and the evaluations tend not to keep up with the rate of new sensors entering the market. This makes it difficult for end-users to establish which monitor is the most appropriate for the intended application, resulting in the need to conduct bespoke evaluations for individual deployments, which further erodes their cost advantage.
Innovation in data processing is rapidly becoming proprietary to private companies, resulting in a lack of transparency.	**It is normal practice for raw air quality measurements to go through post-measurement 'corrections'** to remove the effect of cross-interferences and environmental influence like varying temperature and humidity during the measurement period. As the air sensing market has commercialised, the algorithms employed for post-processing have become both increasingly sophisticated and proprietary to the manufacturer. These algorithms may be 'trained' to certain environmental conditions, meaning a sensor that performs well in one environment may perform

[4] Examples include AQ-SPEC, an EU assessment study and from the US EPA.

	poorly in another. A lack of transparency in post-processing techniques makes it difficult to assess the empirical accuracy of a sensor and the resulting trustworthiness of the data. **There is a field of research in improving the fundamental science of monitoring.** Emerging techniques, for example in spectroscopy, offer the opportunity for a more detailed understanding of pollutant chemistry to understand in more detail how pollution affects health. These methods are not commercial or intended for widespread use, but for mainly health impacts research.

Case Study: Five outputs from Breathe London[5] that will help other cities to deploy lower cost small sensors

Problem	Breathe London solution	Benefit
1. **Existing calibration methods are time and resource intensive**, requiring each sensor to be individually co-located with a reference station	**The 'Gold Sensor' technique**: to precisely calibrate a small sub-set of sensors ('gold sensors') and cycle these around the network to transfer their calibration	**Increased speed and efficiency in deploying and maintaining a network:** - Sensors are calibrated without needing to co-locate every sensor - Allows larger deployments in regions with limited access to reference monitors
2. **The 'Gold Sensor' technique is robust, but still requires manual intervention**	**A new 'Network Based Calibration' system:** using advanced statistics to enable the whole network to calibrate itself without the need for individual co-locations.	**Reduced cost and manpower in operating networks with high accuracy:** - Allows calibration of a large network of sensors without extensive co-location (co-location still needed for at least one sensor) - Highest benefit likely in high pollution areas
3. **A lack of established methods to understand device uncertainty:** essential to inform appropriate uses of the data	**Using machine learning tools to quantify measurement uncertainty:** comparing data from the lower cost network and nearby reference sensors to quantify uncertainty and distinguish between local and regional sources	**Improved performance:** - Remotely identifying malfunctioning devices - Determining levels of air pollution from the regional background
4. **Difficulty in identifying the sources of pollution** from lower cost sensors	**Using a new technique for source apportionment:** Comparing CO_2 and NO_x to more rapidly distinguish sources & evaluate trends	**Assessing policy:** - Using low cost sensors to derive an understanding of the impact of specific policies and key source drivers of pollution
5. **Stationary monitors can't be deployed everywhere in a city;** mobile monitoring can help fill monitoring gaps	**A mobile monitoring methodology and QAQC process:** to ensure representative coverage and acceptable uncertainty	**More complete spatial coverage of pollution across a city:** - Better understanding of pollution in a city and opportunities for citizen engagement

2 Modelling data

Modelling extends our understanding of air quality beyond what monitoring can provide by 'filling the gaps' between measurements, predicting future scenarios and by helping to identify sources.	**Modelling is needed to understand how pollution will vary across space and time, to identify the sources of pollution and to estimate which policies to reduce pollution will be most effective.** It supplements the data generated from monitoring to provide more detailed information on the pollutant life cycle (capturing the emission, dispersion, reaction and removal of pollutants). Models have several applications, such as: - **Developing our understanding** of the state of air quality across scales (from global- to street-level), thus providing a platform from which to develop effective solutions. - **Developing policy** by informing which sources to target, forecasting and mapping future air quality levels in order to inform and evaluate pollution control policies. - **Underpinning investment decisions** by reviewing the impact of different projects on air quality. - **Assessing compliance** with air quality standards and undertaking impact assessments. - **Informing health research** by mapping pollution levels between measurement points to estimate population exposure, the basis for understanding and predicting the impact of local pollution on health.

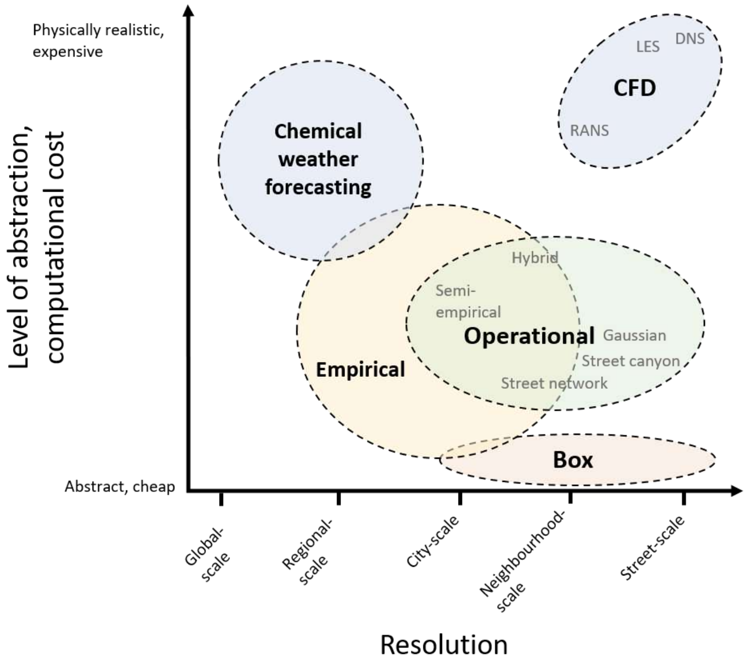
[5] Breathe London is deploying lower-cost small sensors to baseline policy and support campaigning activities in London.

Modelling combines data about what we do know to make predictions about what we don't know.	Modelling transforms available data into information on where pollution is coming from and where it is likely to be in the future. It is possible to provide the link between emissions and concentration data types through two main types of models: - **Predictive models** use emissions data as an input to output concentration data in space and time. Such models can be used to understand the fundamentals of pollution dispersion, forecast real-time air quality, predict future scenarios and undertake bottom-up source apportionment. Numerous modelling approaches exist that vary in scale, accuracy, computational cost and level of abstraction. - **Receptor models** take samples of air (concentration data) as an input and derive the relevant sources (emissions data) as an output. This type of modelling is also called top-down source apportionment.
A range of different modelling approaches exist that vary in scale, resolution and computational cost.	Predictive air quality models combine available data (emissions, meteorology and topology) with numerical representations of real-world processes that define the lifecycle of pollutants. **The two main challenges for air quality modelling are**: 1) **The range of scales associated with pollution dispersion** (from the turbulent wake behind a moving vehicle to global-scale weather systems) demand very high resolutions as well as very large domains and long runtimes. The cost of computational modelling typically increases exponentially with these model characteristics. 2) **The physical and chemical processes affecting air quality are complex**, non-linear and cannot be solved analytically. These processes must be: a. Solved numerically: which means throwing computational power at the problem to obtain results from the governing equations at the highest possible resolutions. The higher the resolution, the more accurate the result but the greater the computational cost. b. Simplified via parametrisations or statistical relationships: these techniques simplify the numerical representation of the problem in a way that significantly reduces computational costs. The disadvantage of these techniques is in how this affects the accuracy and uncertainty associated with the model results. **These challenges result in an inherent trade-off between scale, resolution, and computational cost** (fig 2.6). Supercomputers are generally needed for models that use numerical solutions. Even with this additional computational power, those that simulate at global- or regional scales, e.g. chemical weather forecasting (CWF) models[xiv], are limited to resolutions in the tens of kilometres and those that want to capture street-by-street scale air quality in cities (e.g. computational fluid dynamics models) are limited to neighbourhood-scale sized domains[xv]. **Operational models simplify the problem in a way that allows them to span further across these scales**. Gaussian, street-network and street-canyon models can be applied to predict pollution concentrations at high resolutions across city- and regional scales. **Empirical and statistical models focus on the causal relationship between emissions and concentrations** and therefore bypass the challenges associated with the complexity of the physical and chemical processes that define it. This technique minimises the computational cost associated with running simulations, but it can be computationally intensive to initially derive these relationships. Such models vary from simple matrix modelling systems to complex machine learning algorithms. **Hybrid models exist that combine these approaches, often to achieve better resolutions over larger scales**. Model nesting, where smaller scale models are set within larger scale ones, can be used to obtain greater resolutions over a specific areas, e.g. a city.

<table>
<tr>
<td>

Fig 2.6: Approximate resolutions, costs and levels of abstraction for a range of predictive modelling approaches.

</td>
<td>



</td>
</tr>
<tr>
<td>

Predictive models vary significantly in accuracy and uncertainty. Model evaluations against monitoring data are crucial to gaining trust in modelling results.

</td>
<td>

The accuracy and uncertainty of models does not necessarily improve as the amount of computational power and resolution increases . Model performance varies in different ways for different model types. For example, empirical modelling accuracy and uncertainty is dependant on the model's 'training' and on the application of the model whereas with CWF models you would expect better performance with higher resolutions and improved numerical methods.

Irrelevant to the modelling approach, the key to gaining trust in the output data is to conduct rigorous model evaluations. Comparing model concentration outputs with field and experimental data provides an indication of how well a model is able to predict air quality in the real world. Best practice model evaluation techniques and metrics should be applied and model performance should be presented openly where possible[xvi].

Model inter-comparisons are also a good tool to assess the functionality of different predictive models. For example, Defra has compared numerous predictive models in order to ascertain what models can be trusted to provide policy relevant information.

</td>
</tr>
<tr>
<td>

Data assimilation optimally combines monitoring data with predictive models to improve model performance

</td>
<td>

Data assimilation is a technique where monitoring data is integrated into the model during run time, typically to help adjust initial conditions. By grounding the predictive model results with real-world data, data assimilation can help to improve model performance. It is another useful tool to increase trust in model results, especially those that also cover regions with no monitoring data for comparison.

Combining the data from different monitoring techniques with predictive models in this way can help fill data gaps, improve accuracy and assess monitoring and modelling performance simultaneously if applied correctly.

</td>
</tr>
<tr>
<td>

All predictive modelling approaches can produce 'good-enough' data depending on the stakeholder, scale and application.

</td>
<td>

All modelling approaches have a role to play in the generation of 'good-enough' data depending on the stakeholder, scale and application. Computationally intensive models that require high levels of technical capacity are generally limited to research settings. Generally, these models present the best performing tools to predict future air quality scenarios and to conduct bottom-up source apportionment [xvii].

The learnings from computationally intensive models are increasingly being used to create easier-to-use and less computationally intensive modelling tools. Advancements in operational and empirical modelling techniques are providing models that can be used as effective local-level air quality management tools. Lower model performance, as long as it is acknowledged and

</td>
</tr>
</table>

understood, can be acceptable in terms of facilitating tools that are able to provide air quality forecasts and other tools in data poor regions.

Source apportionment studies tell us where the pollution is coming from – and therefore provides the information to combat air pollution at its source.

Source apportionment is key to identifying the sources of pollution and therefore developing effective solutions. Studies are based around two approaches:

- **Bottom-up**: predictive modelling is performed in a way that allows for the resulting concentrations to be characterised by a breakdown of their source components. This provides vital data on what sources are contributing the problem and where and when this is most significant (fig. 2.8).
- **Top-down modelling**: involves collecting samples of air and analysing the chemical composition of the samples to predict the contribution of different sources to the levels of air pollution at a particular place (also called receptor modelling).

The bottom-up approach is cheaper, faster and more scalable. As a result, it tends to be the most common approach. However, its accuracy depends on the availability and quality of emissions data (inventories, emissions factors, activity data). High uncertainty in the input data will result in high uncertainty in the output (see case study). The accuracy of such studies also depends on the predictive model that is used.

The top-down approach tends to be costly and requires significant equipment and expertise, but provides empirical data which is more accurate, especially where inventory data is poor quality, which is common in low-income settings[xviii].

In an ideal world, both techniques should be deployed to increase confidence in the data.

Case study: Uncertainties in bottom-up source apportionment for Delhi

Delhi is a vast city with severe air pollution problems. Four bottom-up source apportionment studies for PM2.5 were conducted by different groups between 2016 and 2018. The results vary widely (fig. 2.7)[xix]. For example, the estimated contribution of transport to PM2.5 emissions in Delhi varies between studies from a minimum of 17% to a maximum of 39%, industry between 3% and 28%, and road dust between 13% and 38%.

The differences are due to the different methodologies used: each study sampled different seasons, baseline years, geographical extents, and with different emission factors. None are necessarily 'wrong'; but this does demonstrate the complexity in assessing the sources of pollution. It shows that asking a simple question – "where does pollution come from" – can be difficult to answer in an irrefutable and objective way.

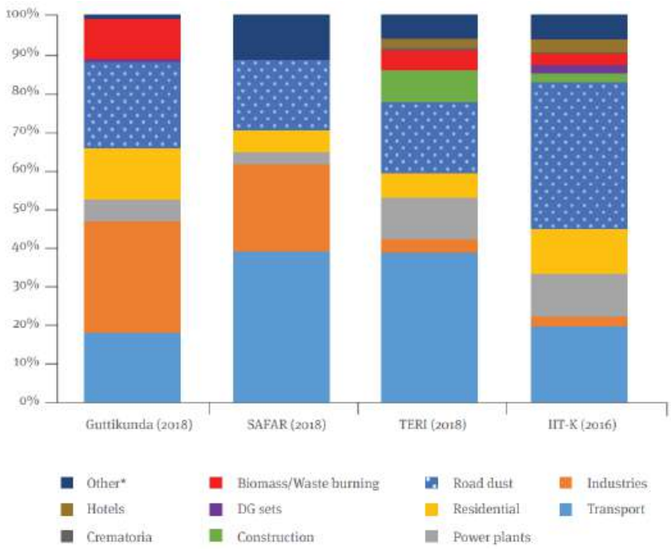


Fig. 2.7: Sectoral contribution to PM2.5 (%) in four bottom-up studies employed in Delhi in a three-year period. Source: CEEW

There is a large bank of resources and tools to support air quality modelling.

There is a growing pool of open-source information to support bottom-up modelling:

- **Comprehensive ('full form') models:** A range of predictive models are freely available, built on by decades of model development research. These provide the computational element of modelling (fig 2.8). An example of an open-source dispersion model is CAMx – which is used by Urban Emissions in India.
- **Simplified ('reduced form') models:** Where input data is patchy or scarce, reduced-form models can provide a starting point for assessing sources and the impact of policy measures. Multiple reduced-form models have been developed and are regularly applied in largely lower-income settings, such as LEAP-IBC and C40's *Pathways* climate action planning tool.

The biggest barriers to scaling air quality modelling are the quality of local data and the local capacity to apply the tools that already exist.

As modelling capability improves, the biggest challenge is making it locally relevant. This requires that:

1. The availability of local data such as emissions inventories improve
2. Technical support is provided to integrate local data with existing tools and open-source models, and
3. The data product provided is actionable: i.e. provided in a form that is relevant to policymakers and layman audiences.

Making models locally relevant is hard to do at scale: it requires deep partnership with the local government to assess available data, identify how data gaps can be filled, and apply model findings in a targeted way.

3 Interpreting data

Unlocking the full value of monitoring and modelling outputs requires "data" to become "information"

Maximising the use of data requires it to be open and accessible, and communicated in a way that is made actionable to technical and non-technical audiences:

- **Technical audiences** (such as regulators and researchers) typically require access to fully transparent and interoperable – also known as "programmatic" - level data. This means the data is:
    - Shared in physical units like micrograms per cubic metre, not transformed into an index like AQI (fig. 2.9).
    - Provided with meta data, for example information about where the data was collected and at what time
    - Provided in a timely way (near real-time).
    - Provided in an accessible form (e.g. through an API or in a machine-readable form, not as a downloadable spreadsheet or PDF).

- **Non-technical audiences** (such as policymakers, the public, media and NGOs): typically require easily interpretable data that is communicated effectively and graphically. Data that shows an individual is exposed to 50 micrograms per cubic metre of PM2.5 is useless information to a layperson. Actionable information would be data that tells a layperson when air quality is unhealthy and suggestions for how they can mitigate their exposure. The US EPA established the AirNow International programme to develop an Air Quality Index and support consistent, layperson-friendly and comparable communication of air quality levels around the world:

Physical and Indexed air quality for PM2.5

Physical Units	
WHO guideline	**Annual mean (µg/m3)**
WHO AQG	10
WHO IT-3	15
WHO IT-2	25
WHO IT-1	35

Air Quality Index	
AQI Category	**Index Value**
Good	0 - 50
Moderate	51 - 100
Unhealthy for sensitive groups	101 - 150
Unhealthy	151 - 200
Very unhealthy	201 - 300
Hazardous	301+

**AQG = Air Quality Guideline. IT = Intertim Target*

Fig. 2.9 – comparing physical units and indexed units. Both are useful communication tools, but for different audiences.

Open data and transparency is the bedrock of interpreting data. There are several efforts to aggregate and share data to unlock their full potential

The fundamental requirement in turning data into information is ensuring that it is open. At least 30 national governments generate real-time data, but do not yet share them in a fully open manner (i.e. programmatically). Making these existing datasets fully open would bring full data access to 4.4 billion people (China, India, Indonesia and Brazil all fall within this category)[xx].

There are several reasons why governments and regulators cannot or choose not to make data fully open:

- **Technical**: lack of real-time data and limited IT infrastructure and personnel to develop a shareable data platform.

	- **Political**: necessary permissions between government departments to publish data and concerns over downstream data misuse. - **Financial**: a low-priority investment when the capacity to even produce data is limited. **There are multiple efforts to aggregate and harmonise data and make it accessible**, including OpenAQ, the WHO Outdoor Air Pollution Database and the recently launched 'Urban Air Action Platform' from UN Habitat and IQAir, which integrates data from the AirVisual sensor. Each varies in its approach and goals, but together have significantly improved the accessibility of air quality data in recent years. **It is not just data that benefits from transparency**. The codes, algorithms and assumptions to process and synthesise air quality data are in high demand. This demand has created a competitive market, which disincentives transparency and results in commercial IP protections. The WHO convened the Global Platform for Air Quality and Health in 2014 in an effort to improve data quality and encourage transparency, enhance cooperation, identify suspected errors and provide possible solutions. This was created following the publication of the first integrated estimate of global air pollution by North American researchers in 2013. The WHO identified likely errors in the data, but could not confirm them without the scientists willingness to open up the black box of assumptions and algorithms used. The Platform is convened every 2-3 years but is targeted at academic and research applications, rather than the commercial market. Innovation is still increasingly becoming proprietary, despite these efforts (see section 3.1).
No common set of standards and protocols exist for communicating sensor data.	**As the number of sensors and manufacturers has increased, so too has the diversity of approaches used to describe and exchange data.** For example, there is no standardised way for data parameters like time of day, averaging periods, parameter names or metadata like type of instrument used. This makes it more difficult to compare data across separate deployments, limiting the usability and applications of data, even if the data is made openly accessible. The US EPA conducted a review of data standards for continuous monitoring data in 2017. This reviewed the state of play and made several recommendations, but the initiative was a one-off review and not sustained[xxi].
4 Managing data	
Improving air quality data is not just a technical challenge, there are human and legal dimensions	**Sustainable provision of actionable air quality data relies on how monitoring is funded and managed**. For example, those with responsibility for air quality also need[xxii]: - **Long-term resources, usually secured through a legal mandate**: the responsible government or regulatory agency works best when there is a clear legal mandate for air quality that ensures ownership of the issue and commitment of funding. - **Cross-government coordination**: Responsibility for air quality is often spread across different authorities such as Ministries of Environment, Transport, Energy, Meteorology and sub-national governments. The lack of clear roles and responsibilities can result in a fragmented, inefficient or inadequate approach to the collection and use of air quality data. - **Political will to ensure long-term funding**: air quality management infrastructure is not built overnight. It can take decades of sustained funding to build and operate an effective air quality data network. This requires sustained commitment beyond the term of a political office.
Foundations can provide seed funding but government or local organisation buy in is required to ensure sustainability	**Ensuring sustained impact requires the recipient government, or local campaigners, to be invested from the start**. A lack of political will is likely to result in a lack of long-term impact. Donors that want to support air quality data projects can address this by requiring co-funding from local governments before a project begins. For example, ensuring the government is responsible for procurement and ongoing maintenance of hardware, whilst donor funding builds capacity or develops data management and analysis tools. For awareness-raising activities, the local community should be involved in the collection of data to maximise engagement[xxiii].

There is a deep-rooted lack of capacity and expertise for generating, analysing and using air quality data in most countries	**The lack of skilled staff and the retention of staff is commonly highlighted as a major barrier to sustainable, long term monitoring in low- and middle-income countries[xxiv].** In many countries, a small team or sometimes even an individual is responsible for all air quality management within a city, including air emission inventories, deploying and maintaining monitors and enforcing regulations. **There is no 'one size fits all' approach to air quality management.** Guidance documents, reports, and city blueprints are important to share learnings but alone cannot fill capacity gaps. Local contexts need to be considered, and staff training provided.

5 Actioning data

Making data actionable requires an application-driven approach to air quality management.	**Air quality data in isolation provides little information beyond simply that regulations are being complied with**. When mixed with other data, like weather, climate, population, economic indicators and energy, air quality data can illuminate problems and solutions. **The definition of "actionable data" depends on the application**. For citizen action, data needs to be personal and easily interpretable. For policy action or research, it needs to be robust and consistent. The first step in developing an impactful investment into air quality data is to first define what the end-goal is.

Case study: One Tweet sparks a profound change to China's environmental policy

Fig. 2.10: The US Embassy Tweet. Values more than 500 AQI are now referred to as "Beyond index".

In 2008, the US Embassy in Beijing installed a reference air quality monitor on its rooftop. It was programmed to automatically tweet pollution data each hour. In November 2010, levels exceeded 500 on the AQI scale. The Twitter bot tweeted that this level of pollution was "crazy bad", an inside joke among the programmers who had used this terminology because they thought such levels would never be seen. The undiplomatic language caused the Tweet to go viral and helped to accelerate investment by the Chinese government in a high-density network of its own. By 2013, around 500 PM2.5 stations had been installed in 70 cities. Five years after that, in 2018, PM2.5 levels in Beijing had fallen by 35%[xxv].

In 2015, the US government installed monitors and tweeted levels at its other diplomatic missions. The World Air Quality Index project (AQICN.org) started as a way of sharing Embassy data[xxvi].

The progress China made on tackling its pollution was obviously not entirely down to that single tweet. However, it undoubtedly sped progress up, and it prompted the government to review its own approach to air quality monitoring.

Case study: Citizen collected data puts air quality on the political map in Bulgaria

In April 2017, Bulgaria, a country of 7 million people, had less than a dozen monitoring stations. A small group of air quality enthusiasts established AirBG.Info to address the lack of monitoring. Their volunteer and community-based model used DIY Particulate Matter sensors and a simple website to aggregate the data onto maps, using an approach and API built by Sensor Community (previously called Luftdaten). Today, the platform has more than 1,000 data points around Bulgaria, providing data in an open format. It is regularly used by media and institutions, and in 2019 led to AirBG.Info organising the first public debate on air quality with the five candidates for the Mayor of Sofia.

Fig. 2.11: The rudimentary citizen sensor used by AirBG.Info

6 References

[i] OpenAQ (2020) Distribution of government and research grade air quality data that is openly accessible. Available at: https://openaq.org/#/map?_k=hwxcqd

[ii] OpenAQ (2020) Open Air Quality Data: The Global State of Play. Available at: https://openaq.org/assets/files/2020_OpenData_StateofPlay.pdf

[iii] Stax (2019) Whitepaper: A Breathe of Fresh Air. A roadmap for investing in air quality sensing technologies to address climate, health, and air pollution. Available at: https://www.stax.com/news/FreshAir

[iv] Tim Dye (2020) Personal communication from a talk given to a small sensor conference with UC Davis

[v] Idrees and Zheng (2020) Low cost air pollution monitoring systems: A review of protocols and enabling technologies. *Journal of Industrial Information Integration.* 17. Available at: https://doi.org/10.1016/j.jii.2019.100123

[vi] Tim Dye (2020) Modified from a talk given to a small sensor conference with UC Davis

[vii] Tim Dye (2020) Personal communication.

[viii] NASA (2020) Air quality satellite constellation begins taking shape. Available at: https://www.nasa.gov/feature/langley/air-quality-satellite-constellation-begins-taking-shape

[ix] Ghosh (2016) Reforming the liability regime for air pollution in India. *Environmental Law & Practice Review.* (pp 125-146).

[x] EPA (2020) Ambient Monitoring Technology Information Center (AMTIC). Available at: https://www.epa.gov/amtic

[xi] EPA (2014) How to use air sensors: Air sensor guidebook. Available at: https://www.epa.gov/air-sensor-toolbox/how-use-air-sensors-air-sensor-guidebook

[xii] EDF (2019) Making the Invisible Visible: A guide for mapping hyperlocal air pollution to drive clean air action. Available at: https://www.edf.org/media/new-edf-guide-mapping-hyperlocal-air-quality-drive-clean-air-action

[xiii] World Bank (2017) Filling the gaps: Improving measurement of ambient air quality in low and middle-income countries. Available at: http://pubdocs.worldbank.org/en/425951511369561703/Filling-the-Gaps-White-Paper-Discussion-Draft-November-2017.pdf

[xiv] Kukkonen et al. (2012) A review of operational, regional-scale, chemical weather forecasting models in Europe. *Atmospheric Chemistry and Physics.* 12, 1-87. Available at: https://acp.copernicus.org/articles/12/1/2012/acp-12-1-2012.pdf

[xv] Thunis et al. (2016) Overview of current regional and local scale air quality modelling practices: Assessment and planning tools in the EU. *Environmental Science & Policy*. 65, 13-21. Available at: https://www.sciencedirect.com/science/article/pii/S1462901116300648

[xvi] Chang, J. C., & Hanna, S. R. (2004). Air quality model performance evaluation. Meteorology and Atmospheric Physics, 87(1), 167-196. Available at: https://link.springer.com/article/10.1007/s00703-003-0070-7

[xvii] European Environment Agency (2020) Guidance report on preliminary assessment under EC air quality directives. Available at: https://www.eea.europa.eu/publications/TEC11a/page011.html

[xviii] Vital Strategies (2020) Accelerating City Progress on Clean Air. Available at: https://www.vitalstrategies.org/resources/accelerating-city-progress-on-clean-air-innovation-and-action-guide/

[xix] Jalan and Dholakia (2019) What is pollution Delhi's air? Understanding uncertainties in emissions inventories. Available at: https://www.ceew.in/sites/default/files/CEEW_What_is_Polluting_Delhi_Air_Issue_Brief_PDF_12Apr19.pdf

[xx] OpenAQ (2020) Open Air Quality Data: The Global State of Play. Available at: https://openaq.org/assets/files/2020_OpenData_StateofPlay.pdf

[xxi] EPA (2017) Data standards for continuous monitoring data. Available at: https://e-enterprisefortheenvironment.net/our-projects/data-standards-for-continuous-monitoring-data/

[xxii] World Bank (2017) Filling the gaps: Improving measurement of ambient air quality in low and middle-income countries. Available at: http://pubdocs.worldbank.org/en/425951511369561703/Filling-the-Gaps-White-Paper-Discussion-Draft-November-2017.pdf and Vital Strategies (2020) Accelerating City Progress on Clean Air. Available at: https://www.vitalstrategies.org/resources/accelerating-city-progress-on-clean-air-innovation-and-action-guide/

[xxiii] Carlos Dora (2020) Personal communication.

[xxiv] World Bank (2017) Filling the gaps: Improving measurement of ambient air quality in low and middle-income countries. Available at: http://pubdocs.worldbank.org/en/425951511369561703/Filling-the-Gaps-White-Paper-Discussion-Draft-November-2017.pdf

[xxv] UN Environment (2019) A review of 20 Years' Air Pollution Control in Beijing. Available at: https://wedocs.unep.org/bitstream/handle/20.500.11822/27645/airPolCh_EN.pdf?sequence=1&isAllowed=y
[xxvi] Wired (2015) Opinion: How the US Embassy tweeted to clear Beijing's air. Available at: https://www.wired.com/2015/03/opinion-us-embassy-beijing-tweeted-clear-air/

Clean Air is a Human Right.

90% of the world's population – approximately 6.8 billion people – live in places where the air they breathe is damaging their health.

The issue is getting more urgent.

We believe in a world where everyone can breathe clean air.

Help us make it happen.

Interested in joining forces?

www.cleanairfund.org
@cleanairfund

CLEAN
AIR
FUND