

COMMENT

PUBLISHING Lessons from 30 years of trying to improve peer review **p.31**



LANDSCAPE From 'Capability' Brown to the US National Park Service **p.34**

POLICY Fetal-tissue research is key to infant health, say subpoenaed researchers **p.37**

SOCIAL SCIENCE Physicists, biologists and chemists, why so snooty? **p.37**

KEVIN FRAYER/GETTY



The greater availability of data on air quality has gripped the public, especially in heavily polluted cities such as Beijing.

Validate personal air-pollution sensors

Alastair Lewis and Peter Edwards call on researchers to test the accuracy of low-cost monitoring devices before regulators are flooded with questionable air-quality data.

The public is increasingly aware of the health and economic costs of air pollution. Poor air quality is linked to over three million deaths each year, and 96% of people in large cities are exposed to pollutant levels that are above recommended limits¹. The costs of urban air pollution amount to 2% of gross domestic product in developed countries and 5% in developing countries (see go.nature.com/28qv0ka).

Media attention and the increasing availability of data are reinvigorating efforts

in many countries to tackle air pollution, driven as much by local and national politics as by science.

In response, start-up companies are rushing to produce cheap air-monitoring sensors, costing hundreds rather than tens of thousands of pounds. Such devices bridge gaps between sparse government measurements and individuals' wishes to track their personal exposures². In a wealthy city, a single official monitoring station might represent 100,000 people; in emerging economies,

one instrument covers millions of citizens.

Although personal sensors have not yet achieved their market potential, applications are promising. Portable sensors are becoming a mainstay of health research by showing people's exposure to environmental factors ranging from noise to particulate matter^{3,4}. Live pollution data can be integrated into traffic-management systems to track the impacts of policies such as low-emissions zones. Affordable air-quality devices are being produced for developing countries. For example, the ▶

► United Nations Environment Programme launched a device in 2015 at a modest cost (around US\$1,500) to measure particulates, sulfur and nitrogen oxides as part of a government pilot scheme in Kenya.

All this excitement presumes that these low-cost air-pollution sensors are fit for purpose. For regulatory applications, governments and scientists use the most accurate, but expensive, detectors. And although the interpretation of the data is a subject of lively debate, the quality of readings is rarely questioned. By contrast, few of these low-cost devices have been rigorously tested and most researchers view the buzz as being beyond the serious business of academia.

The research and regulatory communities are behind the curve. The penetration of these devices into the public domain, generating large volumes of untested and questionable data available to all, is inevitable and will increasingly become a headache for those who are responsible for managing air quality. And opportunities beckon. Atmospheric chemists must engage so that these technologies can realize their huge potential.

COMPLEX BLEND

Measuring atmospheric pollutants is challenging. Most gaseous pollutants, such as nitrogen dioxide (NO₂) or ozone, occur at parts-per-billion levels in air and are blended with thousands of other compounds. Unburnt fuel, for example, contributes many different hydrocarbons to the urban atmospheric mix. Added to this are large and changeable amounts of water vapour and carbon dioxide, at temperatures anywhere between -30 °C and 50 °C. This is difficult analytical chemistry at the best of times.

Atmospheric chemistry research has long been a hotbed of invention for detection technologies and analysis methods. Ideas emerged mainly from universities, institutes and a few research-led companies, such as Aerodyne Research and Picarro in the United States and Ionicon in Austria. The fruits of this labour have been tested by peer review; there are entire journals devoted to atmospheric instruments. Fresh technologies must establish credentials. The best ones are absorbed by a few early-adopter research groups. Over perhaps a decade, successful methods find their way into research use; a rare few make it into regulatory networks. Along the way come dozens of papers, international evaluations, comparison exercises, reference materials and best-practice guides.

By contrast, most of the latest air-pollution sensors are developed by small- and medium-sized enterprises, backed by venture capital and crowdsourced funding. Many devices adapt off-the-shelf technologies. Peer review and academic evaluation may be bypassed. The public are the early adopters; research chemists and physicists are largely on the



A sensor used to measure air quality in Kenya.

sidelines. Academics' funding is threatened by this commercial acceleration, because these devices mean that incremental research developments — such as the miniaturization of high-quality detectors, often based on optical absorption, particle counting or mass spectrometry — are less attractive to granting agencies. Many of the processes used for cheap sensors, such as chemical interactions between gases and surfaces, are less well understood.

The range of devices is wide. The cheapest, costing a few dollars each, use technologies that have been repurposed from hazard detectors, such as metal-oxide sensors that measure oxidizable gases. For tens to hundreds of dollars, electrochemical or photoionization detection can notionally observe particular compounds or classes. In the \$150–1,500 band come miniaturized instruments, such as optical particle counters that can fit in your palm. In general, reducing cost inevitably reduces specificity or sensitivity, or both.

KEEP TESTING

Most commercial sensors target parameters that governments need to track, such as levels of particulate matter (PM) and NO₂. To do a thorough job requires calibration of the target compound and all other possible interferences that might be present. City authorities and the public lack the technical means of checking these themselves, so must take the quality of the measurements on trust from the supplier. The US Environmental Protection Agency has created a technical framework for testing sensors in public use, benchmarking them against the most accurate monitors. But manufacturers might not engage with this process unless they are required to.

The literature on real-world sensor performance is thin. Anecdotally, we have heard that leading research labs have tested

commercial sensors and found them wanting. But because papers reporting negative results have low priority, only a few studies have been published (see, for example, refs 5 and 6). These reveal stability and sensitivity issues, and show that the sensors react to other air pollutants and longer-lived gases such as CO₂ and hydrogen. They are also influenced by meteorological conditions such as humidity, temperature and wind speed.

Simple sensors perform best when pollution levels are high and when the compound of interest swamps others — for example, sensors for nitric oxide (NO) and NO₂ seem to work well in locations that have heavy traffic and high pollution levels, where concentrations of these gases approach the parts-per-million level. In more typical conditions, sensors might respond to other atmospheric species as well. Calibrations of cheap sensors performed in the lab and in the field can differ markedly³, and most relationships observed in the field only apply to that location and for a limited time.

Our research shows that the biggest headaches are caused by interfering chemicals, such as CO₂ and H₂, and by the irreproducibility of measurements. Our real-world test of 20 identical ozone sensors on a roof found a difference of a factor of 6 between the highest and lowest measurements⁷. In other words, the variability of the responses was greater than that of the actual atmosphere. We tested a mid-priced electrochemical sensor for NO₂ in real conditions for an atmospheric concentration of 40 micrograms per cubic metre (the European air-quality limit value). We found that roughly half of the signal from the sensor was from NO₂, and that the rest came from the sensor's response to ambient CO₂. The device was detecting changes in air pollution minute by minute, but not only changes in NO₂.

FAIR USE

Does it matter that a sensor reports an indicative value or trend? It depends on how they are sold and used. Some cheap devices are advertised as being simply for raising awareness of pollution, and it might be expecting too much for them to report accurate values. Others claim to give pollutant measures that can be compared against conventional monitors or official model forecasts.

Until there is agreement on what degree of sensor accuracy is acceptable, we urge caution. Their fitness for purpose should be demonstrated, particularly where they will have a role in decision-making — whether it is at a city, community or personal level. Although we do not wish to stifle innovation, sensors that claim to be able to measure ambient pollution levels could be required to undergo an independent testing regime, as is the case for instruments that are used in regulatory measurements. Some definition of measurement

uncertainty is needed, as is standard practice in other fields — even bathroom scales come with uncertainties printed on them. A mark should signify that the sensor meets a minimum quality standard

If such a stamp of approval sounds bureaucratic, think of how the data might be used. People with asthma might use their local sensor data to make personal decisions on medication; an air-pollution sensor is not meant as a medical device, but its real-world application could make it function like one. Privately owned sensor data could trigger legal actions in areas that apparently exceed local air-quality standards. The economic and socially disruptive costs of closing roads or banning cars based on live sensor data would be huge.

NEXT STEPS

The academic air-pollution community must do the hard yards in the lab and field on calibration and testing. It must also find ways to overcome some measurement challenges. Researchers should take the lead on evaluating sensor performance, creating better devices and designing research applications that are suited to the quantified capabilities of sensors.

More creativity is needed in experimental design. If the long-term performance of sensors is a problem, as is likely, then we need

to design shorter-term experiments that can be performed reliably. For example, a fine-scale but qualitative measure of pollution might help to simulate the turbulent flows of pollution in street canyons or tree canopies over a few days. There might be experiments in which a fast-responding bulk sensor — one that measures the sum of many organic compounds, for example — might be able to track rapid temporal changes that add context to a slower but more quantitative instrument, such as a gas chromatograph or diffusion tube. Statistical and machine-learning methods might be developed to enable better extraction of signals from a mix of pollutants⁸.

However, academics should not become gatekeepers or validation bodies. This is a job for manufacturers and regulators, who need to define how and where sensors can and cannot be used effectively.

Governments must provide advice now to potential ‘professional users’, such as in cities and regional environmental agencies. For sensors that might be used for public policy, health studies or any type of infrastructure control, independent testing and verification is essential, as is already being done through

“Manufacturers and regulators need to define how and where sensors can be used.”

long-standing environment-agency committees and national air-pollution schemes. Even sensors that are designed for entertainment or awareness-raising need appropriate labelling to define their capabilities.

Well designed sensor experiments, that acknowledge the limitations of the technologies as well as the strengths, have the potential to simultaneously advance basic science, monitor air pollution — and bring the public along. ■

Alastair Lewis is a science director at the National Centre for Atmospheric Science in Leeds, UK, and professor of atmospheric chemistry at the University of York, UK.

Peter Edwards is a research fellow in the Wolfson Atmospheric Chemistry Laboratories at the University of York, UK. e-mails: ally.lewis@ncas.ac.uk; pete.edwards@york.ac.uk

1. Lim, S. S. *et al. Lancet* **380**, 2224–2260 (2012).
2. Kumar, P. *et al. Environ. Int.* **75**, 199–205 (2015).
3. Piedrahita, R. *et al. Atmos. Meas. Tech.* **7**, 3325–3336 (2014).
4. Nieuwenhuijsen, M. J. *et al. Environ. Sci. Technol.* **49**, 2977–2982 (2015).
5. Mead, M. I. *et al. Atmos. Environ.* **70**, 186–203 (2013).
6. Kamionka, M., Breuil, P. & Pijolat, C. *Sens. Actuators B Chem.* **118**, 323–327 (2006).
7. Lewis, A. C. *et al. Faraday Discuss.* <http://dx.doi.org/10.1039/C5FD00201J> (2015).
8. De Vito, S., Piga, M., Martinotto, L. & Di Francia, G. *Sens. Actuators B Chem.* **143**, 182–191 (2009).

Make peer review scientific

Thirty years on from the first congress on peer review, **Drummond Rennie** reflects on the improvements brought about by research into the process — and calls for more.

Peer review is touted as a demonstration of the self-critical nature of science. But it is a human system. Everybody involved brings prejudices, misunderstandings and gaps in knowledge, so no one should be surprised that peer review is often biased and inefficient. It is occasionally corrupt, sometimes a charade, an open temptation to plagiarists. Even with the best of intentions, how and whether peer review identifies high-quality science is unknown. It is, in short, unscientific.

A long time ago, scientists moved from alchemy to chemistry, from astrology to astronomy. But our reverence for peer review still often borders on mysticism. For the past three decades, I have advocated for research to improve peer review and thus the quality of the scientific literature. Here are some reflections on that winding, rocky path, and some thoughts about the road ahead.

I trained as a physician, studying the pathophysiology of exposure to high altitudes. In 1977, I became deputy editor of *The New England Journal of Medicine* (*NEJM*), working with what I assumed was a smoothly oiled peer-review system. I found myself driving an enormous machine whose operation was sometimes interrupted by startling hiccups. The first big one occurred a year after I arrived. An author who had submitted a paper to our journal accused one of our reviewers, who worked at a competing lab, of plagiarizing parts of her paper. She sent us a manuscript that her lab chief had been sent to assess for another journal, one that I could see had been typed on the same typewriter that the reviewer had used to write his review. I was told to sort it out.

This was more than a decade before a formal definition of research misconduct and systems for its investigation were established. Several careers fell apart. That

of the actual plagiarist, and also that of his chief, our reviewer, who was the senior co-author of the manuscript that contained the plagiarism. Tragically, our innocent submitting author also gave up research when her accusations were rebuffed, and she was bullied and demeaned for her persistence and integrity.

This slow-motion catastrophe angered me. How common was such incompetence, confusion and corruption? Did peer review root it out — or just lob it down the road? A few years later, revelations of fabricated data in scores of papers by US cardiologist John Darsee, in *NEJM* and other journals, showed that peer review was usually helpless in detecting gross fraud. More recently, the cases of Dutch psychologist Diederik Stapel and US-based cancer researcher Anil Potti underline how easily false data continue to get through the system. Even if peer review could not detect outright