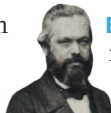


# COMMENT

**CLIMATE** Al Gore's second film joins distant dots in our warming world **p.400**



**ECONOMICS** *Das Kapital* reappraised for these tumultuous times **p.401**

**REPRODUCIBILITY** Enforce deposition and sharing of research materials **p.403**

**CONSERVATION** The many ways to judge a species' value to an ecosystem **p.403**

COURTESY OF UK NATIONAL PHYSICAL LABORATORY



Ground-based measurements of light bouncing off deserts can be used to calibrate satellite observations of reflectivity and improve climate modelling.

## Metrology is key to reproducing results

Scientists of all stripes must work with measurement experts so that studies can be compared, urge **Martyn Sené, Ian Gilmore and Jan-Theodoor Janssen.**

**I**magine you are a policymaker who needs to know how much carbon is stored in the South American forest. On-the-ground data in this area are slim. So when you come across two recently published maps of surface biomass, both made using the exact same satellite data, you think it's your lucky day. Unfortunately, these

maps differ in their estimates of biomass by about 20% across the continent, and by even more on a local level. Which map, if either, can you trust<sup>1</sup>?

Many column inches have been dedicated to discussing this 'reproducibility crisis' in scientific research. Researchers are rarely incentivized to try to replicate

results, and when they do, those results often don't match<sup>2</sup>.

Little attention has been paid in these discussions to how metrology can help. Metrology is the science of measurement: practitioners develop internationally agreed reference points so that measures — of anything from length or mass to radiation ▶



► doses or gene activity — can be compared to standards with a known uncertainty. Metrologists (like us) also work with scientists making measurements to develop and disseminate best practice. Greater attention to those standards and best practices, and the development of new ones, is needed to help researchers reproduce results.

Today's scholarship is increasingly multidisciplinary and fast-moving, bringing together scientists with widely differing expertise, using different technical languages and techniques. This can lead to measurements being made without the ability or opportunity to validate them properly.

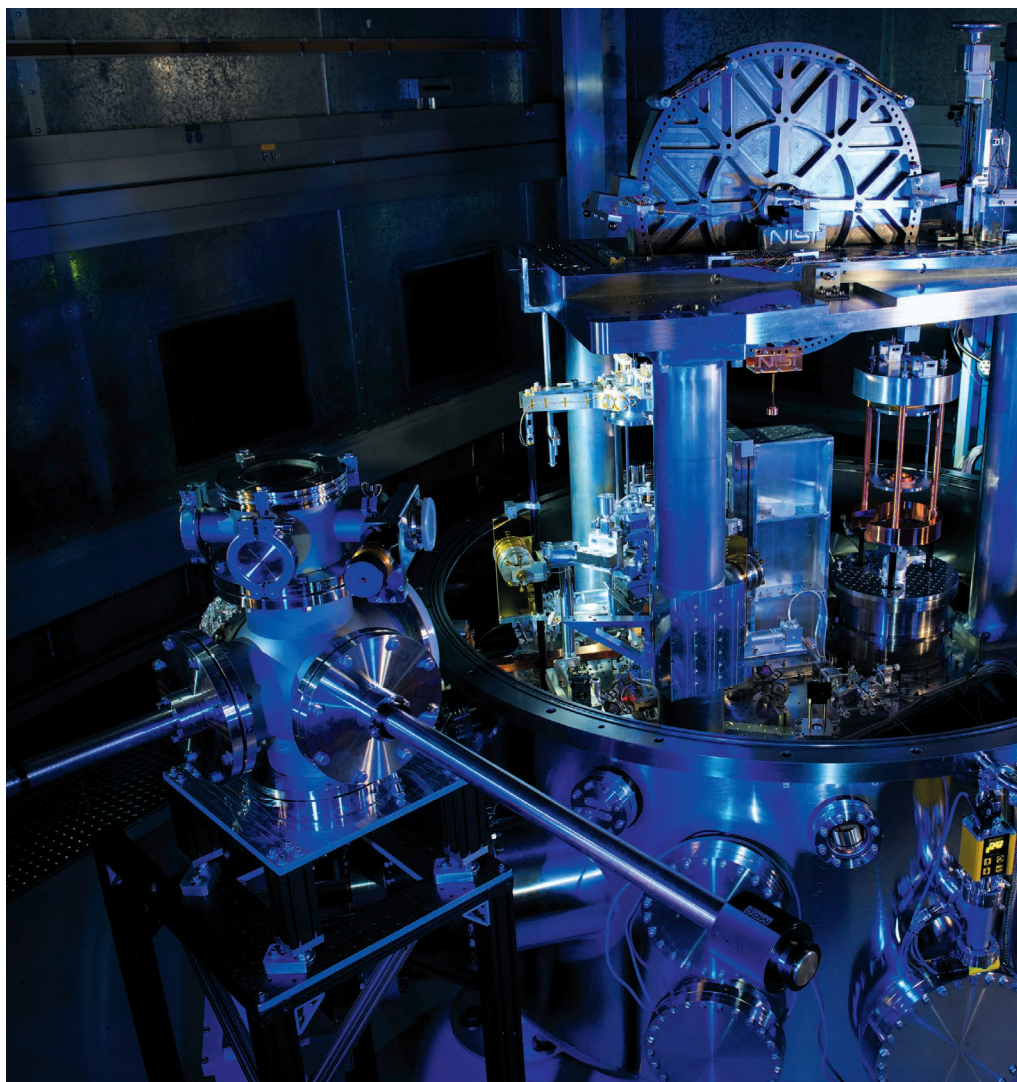
Measurement technology is becoming more powerful and complex. Software often stands between the raw data and the user: numbers are processed and data sets are combined automatically. Tracking and quantifying the uncertainty of the final result can get lost amid all this data crunching. Researchers often treat such tools as a 'black box' that spits out answers they take on trust, and find it harder to have an intuitive feel for when the answers are wrong.

A renewed focus on how data are collected, annotated and analysed could help to fill in this piece of the reproducibility puzzle<sup>3</sup>. In the South American forest example, differences in instrument calibration, uncertainty in ground-based reference data and differences in modelling methods created the mismatch between the maps<sup>1</sup>. A serious investigation into exactly how and why results differ can turn up systematic errors, or at least quantify the measurement uncertainty. Without work like this, there is no way that maps such as these can ever match.

### INTERNATIONAL EFFORT

Our institution, the National Physical Laboratory (NPL) in Teddington, UK, is one of dozens of National Metrology Institutes around the world that sit at the heart of the international measurement system. This system provides the framework, facilities and expertise to enable measurements to be reproduced with confidence, and with quantified uncertainty, across the globe.

The benefits of good metrology have been reaped for centuries. In the 1800s, a coherent, agreed system for measuring length and mass helped countries to have confidence in how much they were buying and selling in global trade and in the accuracy of their maps. Prototypes of the metre and the kilogram were made and locked in a vault in France so that no one could dispute their true values. The Industrial Revolution took off because people agreed on common manufacturing standards such as the type of screw thread used. Some two centuries later, the Global Positioning System relies on satellites that carry highly synchronized atomic clocks providing precise measures of time.



This Kibble balance precisely measures Planck's constant, which is helping to redefine the kilogram.

Although it was Albert Einstein who said that the speed of light was constant, it was the metrology community that measured that speed and set the agreed number.

Today, the International Bureau of Weights and Measures (BIPM), based in Paris, coordinates a robust metrological framework for all seven base units of measurement, from the metre to the kelvin. Advances continue to be made. In November 2018, for example, the definition of the kilogram and some other units are set to change, completing a long-term project to link all units to fundamental, unchanging properties of the Universe that researchers have measured to great precision (in the case of the kilogram, relating it to Planck's constant). If researchers are properly trained to use best metrological practice, following clear procedures and calibrating their measurements against standards that are directly linked to the agreed base units, we can all have confidence in the results.

The system can work extremely well, even for highly complex projects producing vast

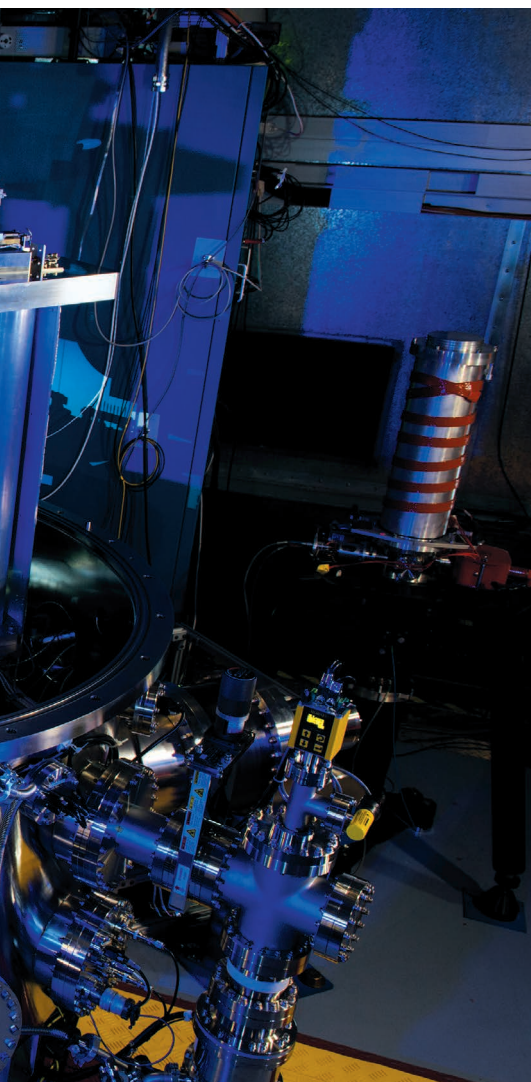
amounts of data from a range of instrumentation. The detections of the Higgs boson in 2012 and of gravitational waves in 2016, for example, were made with such attention to detail that they produced quantitative results that the world can have confidence in.

### PROBLEM AREAS

An increasing number of research areas lack a metrological framework, however. This is particularly the case in fields such as biology and environmental science, which do not share the long history of metrological practice found in physics and engineering. Defining measurement units in the life sciences is an intrinsically tricky task. Every electron can be counted on to have the same mass and charge, whereas living things have a wide range of natural variability, making it hard to develop and define standards. Before we start to tackle such variability, we need to ensure the measurements we do know how to make, and the tools we can characterize, are on a firm foundation.

One problem area is radiotherapy — the





practice of using ionizing radiation to kill cancers or to otherwise affect cells. Although there is strict regulation about how to measure dose delivery for patients in clinical settings, similar regulation does not exist for research labs studying the cellular impacts of radiation. A 2013 report by the US National Institute of Standards and Technology (NIST) found that, in a year's worth of studies in the journal *Radiation Research*, only 7% cited written dosimetry standards or guides. The NIST survey concluded that radiobiological measurement is "frequently inadequate, thus undermining the reliability and reproducibility of the findings"<sup>4</sup>. This creates a barrier to the translation of preclinical studies into clinical practice, and unnecessarily increases the number of animals used in studies.

In response to this, work is under way in the United States at a number of centres to standardize dosimetry. In the United Kingdom, NPL is leading the way on services specifically aimed at preclinical studies.

Earth observation has problems too. Light bouncing off the planet's surface, for

instance, can be reasonably well calibrated by looking at reflections from polar regions and deserts. Although this effect has been studied sufficiently to allow for consistent, reproducible results from different satellites, the information is not reliable enough to use in climate-change studies or measures such as forest coverage.

Take, for example, four different satellites monitoring leaf area index — a measure proportional to the percentage of ground covered by green, photosynthetically active leaves. All four satellites (CYCLOPES, GEOLAND, GLOBCARBON and MODIS) have a resolution of 1 kilometre. Yet their monthly data over two years vary wildly from each other, sometimes by more than a factor of seven. The reasons for this are likely to be complex: the satellites pass at slightly different times, for instance, so the property being measured might be changing. But there are also differences in how the satellites are calibrated and the data analysed. Researchers are working to build a rigorous system of long-term validation and inter-comparison studies, to tease out systematic uncertainties and create more reliable data.

Another example of such work is the Versailles Project on Advanced Materials and Standards (VAMAS). Established in 1982, it was designed to develop international best practices and standards for making and measuring new materials. It is serving the community well for the measurement of ultra-flat 2D materials such as graphene — metrologists have refined techniques for gauging purity and thickness down to the atomic level.

Today, the National Metrology Institutes are leading efforts to standardize many biological measures, such as quantifying small amounts of protein in complex serums.

Such community efforts are incredibly important, yet they remain much less glamorous than discovery research.

#### WAYS FORWARD

So what can be done? One simple step would be for funding bodies to involve more metrologists in project selection and assessment. This would encourage the funding of replication studies, help to ensure that financed studies use good metrological practice and set studies up to allow for future attempts at replication. Grants should assess 'pathways to reproducibility' along with 'pathways to impact'.

Funding bodies often require that the raw data behind research are captured and made available. This requirement should be extended to include information on the quality of the data. It must be clear how, and how thoroughly, researchers worked to ensure their measures were linked to an internationally recognized standard and to quantify uncertainty. If this information is

consistently stored alongside data, it will make it much easier to track uncertainties as data sets are processed and combined.

Some organizations are taking steps in this direction. The Australian Terrestrial Ecosystem Research Network (TERN), for example, has a framework and best-practice guide for collecting this sort of metadata. NPL is taking a leading role in the development of quality-control systems for Earth observation data sets being submitted to the European Copernicus Climate Change Service (C3S). This will ensure that all of the data in the C3S data store are fully traceable and well documented.

Quantifying uncertainty in complex problems is almost becoming a field in itself. The metrology community needs to step up to this challenge, in particu-

**“Grants should assess ‘pathways to reproducibility’ along with ‘pathways to impact’.”**

lar by engaging more statisticians, data experts and researchers from problematic areas such as cell biology. Metrology should be woven into scientific training, at

all levels, to forge a dedication to precision measurement throughout science.

In the meantime, researchers should take full advantage of their National Metrology Institutes. It's surprising how many scientists have never heard of us. Labs having trouble reproducing their measurements can simply give us a call: we work in collaboration to provide advice, and to improve or develop new techniques. We measure almost every physical and chemical parameter imaginable, from time with an accuracy better than one second in the lifetime of the Universe, to the amount and localization of drug uptake in single cells. Speaking to us can often save time and improve the precision of results.

The task ahead is a challenging one that cannot be tackled by the metrology community alone. But it does require the mindset of metrologists: an attention to detail and a dedication to global comparability. ■

*Martyn Sené is deputy chief executive at NPL with a background in nuclear physics.*

*Ian Gilmore is a senior NPL fellow leading the National Centre of Excellence in Mass Spectrometry Imaging and is head of science.*

*Jan-Theodoor Janssen is research director at NPL with a background in solid-state quantum technology.*  
e-mail: jt.janssen@npl.co.uk

1. Mitchard, E. T. A. *et al. Carbon Balance Manag.* **8**, 10 (2013).
2. Baker, M. *Nature* **533**, 452–454 (2016).
3. Plant, A. L., Locascio, L. E., May, W. E. & Gallagher, P. D. *Nature Methods* **11**, 895–898 (2014).
4. Desrosiers, M. *et al. J. Res. Natl Inst. Stand. Technol.* **118**, 403–418 (2013).