Next-generation sensor technology is challenging traditional notions of how environmental monitoring is performed and by whom. Much of this new monitoring technology is portable, inexpensive, and relatively easy to use, making it more accessible to environmental professionals in the private and public sector and even to average citizens.

While these new tools are unlikely to replace the federally-required monitor types and methods used to determine regulatory compliance (at least not anytime soon), they offer the possibility of equipping organizations with supplemental data needed to draw a more detailed picture of air and water quality issues within a specific jurisdiction.

But as with any new technology, there are some kinks to be worked out. A significant one facing distributed, or participatory, monitoring relates to data quality. How can environmental managers help ensure the data gathered by novices using next-generation sensors meet a minimum standard of quality?

**Data Quality**

The field of statistics deals with questions like this routinely and is a critical resource to anyone considering supplementing regulatory monitoring with next-generation sensors, or even a community monitoring pilot project. Some basic principles of statistics can show what makes data good enough to yield useful findings, and steps that can be taken to achieve that standard before any data collection takes place.
First, it’s important to understand what is meant by data quality (or more specifically, the criteria used to determine that data are of appropriate quality). Data quality can be defined differently depending on the purpose of the data collection and the objectives to be addressed from the data. As it relates to data, the term “quality” indicates the general level of acceptability of the data for their intended use, with “high quality” implying a high likelihood of acceptability. This level of acceptance is typically expressed through one or more measurable attributes, such as:

- **type** (verification that you are collecting the right data, using acceptable technologies and following correct sampling protocols);
- **precision** (the extent of random variability in your data);
- **bias** (any consistent error or distortion present in your data);
- **representativeness** (the extent to which your data are representing a desired characteristic); and
- **completeness** (how much of the planned data you successfully collected).

Suppose a particular organization needs air monitoring data for a particular purpose, but to meet its needs, the organization sets specific criteria on the type and representativeness of the data, and states that the data must have high precision and minimal or no inherent bias. If you can sufficiently demonstrate that your data achieve all of these criteria and are complete relative to your initial plan, then that organization will recognize your data as meeting their objectives for quality; otherwise, they likely will not be able to use your data for their specific purpose.

Intended use is key to determining whether your data have acceptable quality. Your data may be acceptable to use for one purpose, but perhaps not another. For example, suppose you aim to collect data to characterize the concentration of toxic metals in ambient air in your community, but you locate your sensors close to the fenceline of a coal-fired power plant in the outskirts of town, away from schools and residential/recreational areas, that you suspect is a source for the metals. Even if you could demonstrate that your sensors were sufficiently maintained and calibrated and the collected data were highly precise and complete, the data would be biased and not representative of concentrations present in the community, and thus the data would not achieve the quality objectives for your intended use.

The data gathered from a community monitoring project are more likely to be useful and credible if the data collection effort is planned out, with a goal of having the data achieve specific quality objectives, and with documenting information on the extent to which the data achieve these objectives.

Such information is called “metadata.” As participatory monitoring is still an emerging field, government agencies are still formulating guidance on the appropriate levels of quality and types of metadata to collect, which can differ from one sampling technology to another. For example, qualitative data on a citizen’s personal exposure to airborne chemicals (e.g., determining high vs. low concentration levels) or data collected to identify the presence of any isolated areas of unusually high concentrations can have lower quality acceptance levels than quantitative data to be used to supplement monitoring conducted for regulatory compliance.

### Key Steps to Improving Data Quality

Government agencies, such as the U.S. Environmental Protection Agency (EPA) and the National Institutes of Health, as well as state and local agencies, are recognizing the important role that next-generation sensors and community monitors can play in collecting and reporting information on environmental exposures and their impact on human or ecological health. Such information can provide these agencies with additional data at a cost savings, and the ongoing advancements in sampling technology are increasing the likelihood that these data will meet their acceptance criteria for certain initiatives (e.g., supplementing information from national monitoring networks, monitoring personal exposures).

For instance, EPA’s Volunteer Monitoring Program1 provides materials and hosts conferences...
to educate citizen scientists about the methods of volunteer monitoring. EPA has also been working on evaluating next-generation air sensors so environmental managers have more valid information about the quality and accuracy of these new, smaller, and less expensive tools that are increasingly being used by members of the community to gather data.2

To make sure the data are credible enough to be used for analysis and decision-making, below are a few basic tips for environmental managers who are considering incorporating mobile sensors or community members in their data collection efforts.

**Do Appropriate Planning**

Nothing substitutes for sufficient planning in terms of setting your data collection on the proper track for collecting the right types and amounts of data to meet your needs without wasting valuable resources and time. EPA, for example, recommends the use of the Data Quality Objectives process as one type of planning tool, followed by documenting all quality-related procedures to be implemented within a Quality Assurance Project Plan (QAPP; http://www.epa.gov/quality). This planning and documentation process can take many forms and does not need to be elaborate or overly complex or time-consuming—it simply helps you to “do your homework” prior to collecting data of the quality that you will need.

**Follow an Approved Sampling Protocol**

If you aim to provide data to a government agency for a specific use (such as assessing air pollutant exposures in the community), you must follow all government-approved sampling protocols established for that purpose, including the use of an approved sampling technology. For example, the protocol may require that certain quality control activities be performed to help assess bias and precision in a given set of sample measurements (e.g., blanks, spiked samples, duplicate sampling). You also may need to adhere to a checklist of activities, such as verifying that the technology is in working order, the sampler has been appropriately calibrated, and so on. This checklist would be part of your QAPP.

Regardless of the intended use of the monitoring data, it is important to follow standard practices to ensure accurate results. For example, New York State’s Department of Environmental Conservation’s Community Air Screen Program3 uses EPA-approved sampling and analysis methods. In 2013, EPA Region 2 issued Guidance for the Development of QAPPs for Citizen Science Projects, which is available at http://www.epa.gov/region2/citizenscience.

**Document Conditions Associated with Sample Collection**

As part of the collection of metadata, you should document any conditions that may influence the outcome of your data collection, and therefore, the quality of your data (e.g., the extent of bias). The sampling protocol may indicate some information you should record, such as wind speed and direction (which you can get from local media sources if you don’t use a weather station as part of your process). This information is especially important when characterizing trends in data over time. You should also make note of any unusual or unexpected occurrences during the sampling process that may impact the measurement, such as malfunctioning or damaged equipment, or

---

**In the Next Issue...**

**GHG Performance Standards**

The U.S. Environmental Protection Agency (EPA) has proposed three rules for new, existing, and modified/reconstructed sources to regulate greenhouse gas (GHG) emissions by power generators. The February issue will present a variety of stakeholder views on these important proposed regulations, which EPA is expected to finalize by mid-2015.

Also look for...
- Asian Connections
- IT Insight
- YP Perspective
inability to collect a sample for the recommended time period. This latter information will be critical to data analysts who may need to determine whether to exclude the affected data values. Noting when deviations from the protocol occurred is also important.

New York State’s Department of Environmental Conservation also follows such protocols in its community monitoring program. Its volunteer reporting materials require that monitors “mark the exact location where the sample was collected…and to consider including photos of the sampling site.”4 Citizen scientists are also encouraged to record any observations that might be relevant, such as the amount of traffic, nearby industrial facilities, or heavy winds.

Recognize When Other Data Are Needed
To address all of your questions and achieve your goals, you may need to collect different types of data. For example, to better determine if the pollutant levels that you’re measuring in the environment of a localized area are leading to an unhealthy dose to individuals in that area, you may also need to collect information on the typical activity patterns of those individuals, including how often they are out of doors and performing certain activities. This may be as simple as having them complete a questionnaire or diary, but this type of objective does suggest that you should utilize personal air monitors for collecting data on pollutant levels for individuals. A statistician may be helpful in bringing the different types of data together and optimizing your ability to make conclusions from all of this information.

References