**Boletín de Estadística e Investigación Operativa**
Vol. 25, No. 1, Febrero 2009, pp. 57-61

Opiniones sobre la profesión

# The role of statisticians in U.S. environmental regulation and policy

**Montserrat Fuentes**
North Carolina State University
✉ fuentes@stat.ncsu.edu

## Partnerships between academia and industry in the U.S.

Statisticians from academia in the U.S. are encouraged to participate in the transfer of new knowledge to industry (includes both industry and government). We (in academia) are able to guide our research based on the needs of the actual users of statistics, and we pursue and see industry as an important funding source. With our new developments in statistical theory we lead industrial problem solving into the future. This cooperation between academia and industry in the U.S. has been very successful, in part due to the fact that, we give as much credit to a joint publication addressing a significant real problem in a creative way, as we do to a theoretical paper addressing a relevant problem. We also believe that sending a PhD. graduate to a very good industrial position that requires a Ph.D. level training, is as valuable to our program and the profession as sending the student to an academic position.

In the U.S., statisticians from academia and industry very frequently exchanges roles, with faculty members taking one-semester to one-year "sabbaticals" from academia to contribute to industry, and industrial statisticians taking "sabbaticals" to work within universities. This two-way exchange of knowledge helps industry to inform academia of their needs for broadly educated statisticians, rather than menu-driven technicians. Apart from sabbatical opportunities, industry very often uses statisticians in academia for the solution of significant industrial problems, through research grants, contracts, and cooperative agreemments with academic institutions.

As a professor of statistics in a U.S. University, I have had the opportunity to work extensively in industry, where industry in my case refers to the U.S. Environmental Protection Agency (EPA).

## The birth of EPA

In July of 1970, the U.S. White House and Congress worked together to establish the EPA in response to the growing public demand for cleaner water, air, and land. Prior to the establishment of the EPA, the federal government was not structured to make a coordinated attack on the pollutants that harm human health and degrade the environment. EPA was assigned the task of repairing the damage that had been done to the natural environment, and establishing new criteria to guide Americans in making a cleaner and healthier environment a reality.

EPA leads the nation's environmental science, research, education, and assessment efforts. The mission of the EPA is to protect the environment and human health. When Congress writes an environmental law, EPA implements it by writing regulations. Often, EPA sets national standards that are enforced through their own regulations. EPA enforces the regulations, and helps companies to understand the requirements.

EPA employs 18,304 people across the country, including the headquarters offices in Washington, DC, 10 regional offices, and more than a dozen laboratories. The EPA staff are highly educated and technically trained; more than half are engineers, scientists, and policy analysts. In addition, a large number of employees are legal, public affairs, financial, information management and computer specialists. Approximately, 100 of the employees are statisticians, though there are many engineers with some statistical training.

From regulating auto emissions to banning the use of DDT; from protecting the ozone layer to cleaning up toxic waste; from increasing recycling to revitalizing inner-city brownfields, EPA's achievements have resulted in a cleaner air, purer water, and a much better protected land and ecosystem, and statistics has played a very relevant role, by supporting the research programs and assisting in the formation and development of environmental policy.

**What do statisticians do at EPA?**

Statistics could be defined as the scientific application of mathematical principles to the collection, analysis and presentation of numerical data. We (statisticians working on environmental protection) contribute to scientific inquiry, by applying our mathematical knowledge to the design of surveys and experiments (for e.g., national aquatic assessments); collection, processing, and analysis of data; prediction of pollution levels; studying correlations between pollution and health effects; and very importantly in the interpretation of the results.

Statisticians at EPA recently have been playing a very relevant role in the field of environmental health. Statisticians provide scientific expertise to assist EPA's air pollution program office on risk assessments and other program needs related to studies on the health effects of air pollutants. Statisticians also lead efforts in the evaluation of epidemiologic evidence on health effects of air pollutants, and contribute to the integration and synthesis of evidence on the health impacts of air pollutants.

There are many areas of statistical research at EPA, all within the context of environmental protection. The agency is currently very interested in the characterization of uncertainty in data and models. Since statistics could be considered the science of “uncertainty", statisticians at EPA are leading that effort. This interest in characterization of uncertainty has also led to more opportunities for collaborations between academic statisticians and EPA.

**Impact of statistics on policy and regulation**

Statisticians at EPA play a very important role in assisting in the formation of environmental policy. For instance, by providing a logical framework that shows “significant" statistical associations between environmental stressors and adverse human health outcomes. Recently, I have been a member of a committee of the National Academy of Sciences studying the link between exposure to ambient ozone and mortality. The association was studied using statistical models. Our report concluded that short-term exposure to smog, or ozone, is clearly linked to premature deaths, and that should be taken into account when measuring the health benefits of reducing air pollution. This is an important result, because it could lead to different regulation for ambient ozone. The findings contradict arguments made by some White House officials that the connection between smog and premature death has not been shown sufficiently, and that the number of saved lives should not be calculated in determining clean air benefits.

Statisticians at EPA also provide input to policy development, by facilitating the undertaking of a statistical inferential approach to monitoring design. An environmental monitoring program has to be adequate in its quality and quantity of data, so that the EPA environmental objectives can be assessed. However, monitoring programs can fail to return satisfactory information for policymakers, because future statistical needs were not anticipated, potential confounding factors were not considered, or sampling protocols did not specify suitable randomization. A statistical framework for monitoring design should result in data which are more general in their interpretation. Another key intermediate role exists for statisticians in providing a logical framework for using monitoring data to test hypotheses about fulfillment of the EPA environmental objectives.

Despite the relevant role that statistics can play in the formation and development of environmental policy. Statistical models can not be used to determine compliance with air quality standards. The statements of U.S. environmental air quality policy are intended to be assessed using only information from monitoring programs. Therefore, at a location where we do not have a monitoring station, lack of compliance could not established based on a statistical interpolation of nearby monitoring data. Instead, a monitoring station should be placed at that location. What would be the reason for not using the most natural statistical approach to determine compliance? Well, EPA is very often sued, and it is very difficult to defend successfully in court, a result based on a statistical interpolation model. Because there is not a "perfect" statistical model, and different models could provide different results.

It is, in fact, very difficult to defend statements in court based on statistical analyses. I was recently working for the U.S. Department of Justice, in a case against some companies that were emitting very high levels of pollution. The government was using EPA deterministic air quality models to determine the levels of particulate matter (PM), a pollutant that is being regulated in the U.S., that were potentially obtained due to the emissions from the companies being sued by the government. Statisticians from both sides were presenting their cases. The main point of the other side was that the EPA deterministic air quality models are not perfect, they have their own uncertainties, and if one takes into account the magnitude of their uncertainty. Then, there was not evidence that the obtained levels of PM were "significantly" larger than they would be without the emissions from the companies under consideration. But there is not a unique way to characterize uncertainty, so each side was presenting different characterizations that led to different results.

This difficulty for the agency to defend in court their potentially statistical-based environmental regulations, is one of the reasons why statistics is not as widely used as it should be in the context of environmental policy and regulation. However, slowly this is changing, and statistics is starting to make its way through, and mainly due to the unavoidable need of uncertainty analysis, since lack of characterization of uncertainty can lead to faulty assessments, some of which have serious consequences. However, as I mentioned before, it is not easy to determine which ones are the "right" or "defendable" uncertainty analyses.

**Statistical software at EPA**

The main statistical software that statisticians at EPA have always used in SAS. However, in the last few years, R is starting to be very widely used at EPA, and not only by the statisticians. Other popular softwares in academia, such as WinBUGS, are not yet as widely used at EPA. Mainly, because the characterization of uncertainty has been done for the most part using simulation Monte Carlo approaches, rather than hierarchical fully Bayesian frameworks. However, this more complex Bayesian frameworks introduced by some of us, are now being adopted by EPA to produce maps of pollution levels across the country, to determine if the pollution levels are increasing or decreasing.

Most of the programming at EPA is done using FORTRAN, since that is the programming language used for most of the EPA deterministic and stochastic models, in particular for air quality models and some of the exposure models.

A statistician at EPA is expected to have high level computing skills.

There are many challenges that statisticians have to face when working in multidisciplinary research teams on the interface between science and policy.

Persuasion becomes vital when working in a team, especially if you are a statistician, as some may not see statistics as having an important role. However, having the opportunity to motive my research by important scientific problems on environmental protection, and seeing the impact that is having on our society, the ecosystem, and human well-being, is extremely rewarding. While much has been accomplished, many more challenges and exciting projects remain ahead.

### About the author

**Montserrat Fuentes** is a full professor of Statistics (with tenure) at North Carolina State University, and a visiting scientist at the U.S. Environmental Protection Agency (EPA). Fuentes received her B.S. in Mathematics from the University of Valladolid (Spain), and her Ph.D. in Statistics from the University of Chicago (1998). She is a member-elect of the International Statistical Institute. Fuentes is a member of the scientific advisory board of the U.S. EPA, and a member of the scientific advisory committee of Health Canada. She is in the board of directors of the International Environmetrics Society, and in the board of trustees of the National Institute of Statistical Sciences. She is a member of the Biostatistical Methods Research Design study section of the National Institutes for Health. She has also worked for the U.S. Department of Justice as an expert witness, and she was a member of a committee of the National Academies working on the impact of ozone on mortality (2007-2008). Fuentes has been elected for the IMS council. She is also currently the council of sections representative of ASA. She received the Abdel El-Shaarawi Young Research's Award in recognition of outstanding contributions to environmetric research (2003). Fuentes was named an ASA fellow (2008) for outstanding contributions to research in spatial statistics.