Instrumenting the Earth: Next-Generation Sensor Networks and Environmental Science

MICHAEL LEHNING
NICHOLAS DAWES
MATHIAS BAVAY
WSL Institute for
Snow and Avalanche
Research SLF

MARC PARLANGE
École Polytechnique
Fédérale de Lausanne

SUMAN NATH
FENG ZHAO
Microsoft Research

Increasing environmental challenges worldwide and a growing awareness of global climate change indicate an urgent need for environmental scientists to conduct science in a new and better way. Existing large-scale environmental monitoring systems, with their coarse spatiotemporal resolution, are not only expensive, but they are incapable of revealing the complex interactions between atmospheric and land surface components with enough precision to generate accurate environmental system models.

This is especially the case in mountainous regions with highly complex surfaces—the source of much of the world's fresh water and weather patterns. The amount of data required to understand and model these interactions is so massive (terabytes, and increasing) that no off-the-shelf solution allows scientists to easily manage and analyze it. This has led to rapidly growing global collaboration among environmental scientists and computer scientists to approach these problems systematically and to develop sensing and database solutions that will enable environmental scientists to conduct their next-generation experiments.

NEXT-GENERATION ENVIRONMENTAL SCIENCE

The next generation of environmental science, as shown in Figure 1, is motivated by the following observations by the atmospheric science community: First, the most prominent challenge in weather and climate prediction is represented by land-atmosphere interaction processes. Second, the average effect of a patchy surface on the atmosphere can be very different from an effect that is calculated by averaging a particular surface property such as temperature or moisture [1-3]—particularly in the mountains, where surface variability is typically very high.

Figure 2 shows an example of this—a highly complex mountain surface with bare rocks, debris-covered permafrost, patchy snow cover, sparse trees, and shallow and deep soils with varying vegetation. All of these surface features can occur within a single kilometer—a resolution that is typically not reached by weather forecast models of even the latest generation. Existing models of weather prediction and climate change still operate using a grid resolution, which is far too coarse (multiple kilometers) to explicitly and correctly map the surface heterogeneity in the mountains (and elsewhere). This can lead to severe errors in understanding and prediction.

FIGURE 1.
A typical data source context for next-generation environmental science, with a heterogeneous sensor deployment that includes (1) mobile stations, (2) high-resolution conventional weather stations, (3) full-size snow/weather stations, (4) external weather stations, (5) satellite imagery, (6) weather radar, (7) mobile weather radar, (8) stream observations, (9) citizen-supplied observations, (10) ground LIDAR, (11) aerial LIDAR, (12) nitrogen/methane measures, (13) snow hydrology and avalanche probes, (14) seismic probes, (15) distributed optical fiber temperature sensing, (16) water quality sampling, (17) stream gauging stations, (18) rapid mass movements research, (19) runoff stations, and (20) soil research.

In next-generation environmental science, data resolution will be addressed using densely deployed (typically wireless) sensor networks. Recent developments in wireless sensing have made it possible to instrument and sense the physical world with high resolution and fidelity over an extended period of time. Wireless connections enable reliable collection of data from remote sensors to send to laboratories for processing, analyzing, and archiving. Such high-resolution sensing enables scientists to understand more precisely the variability and dynamics of environmental parameters. Wireless sensing also provides scientists with safe and convenient visibility of *in situ* sensor deploy-

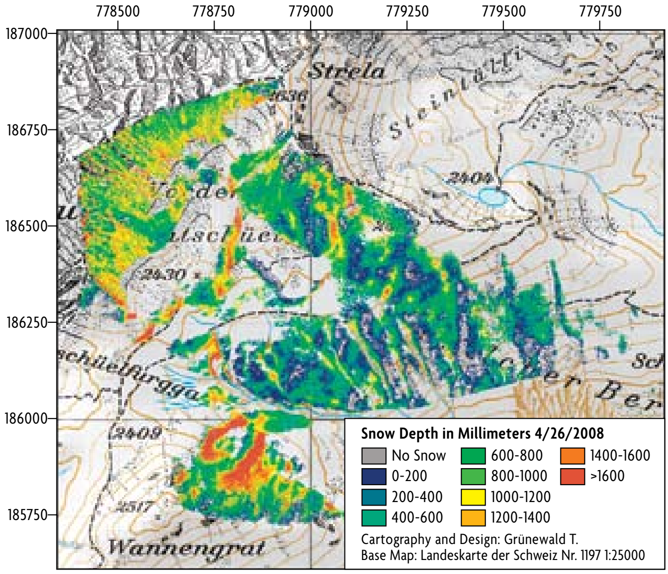

FIGURE 2.
Terrestrial laser scan for snow distribution in the Swiss Alps showing typical patchy snow cover.

ments and allows them to enable, debug, and test the deployments from the laboratory. This helps minimize site visits, which can be costly, time consuming, and even dangerous.

However, dense sensor deployments in harsh, remote environments remain challenging for several reasons. First, the whole process of sensing, computation, and communication must be extremely energy efficient so that sensors can remain operational for an extended period of time using small batteries, solar panels, or other environmental energy. Second, sensors and their communication links must be fairly robust to ensure reliable data acquisition in harsh outdoor environments. Third, invalid sensor data due to system failures or environmental impacts must be identified and treated accordingly (e.g., flagged or even filtered from the dataset). Although recent research (including the Swiss Experiment and Life Under Your Feet) partially addresses these issues, further research is needed to address them in many production systems.

MANAGING AND EXPLORING MASSIVE VOLUMES OF SENSOR DATA

High-resolution environmental sensing introduces severe data management challenges for scientists. These include reliably archiving large volumes (many terabytes) of data, sharing such data with users within access control policies, and maintaining sufficient context and provenance of sensor data using correct metadata [4].

Environmental scientists can use commercial database tools to address many of the data management and exploratory challenges associated with such a massive influx of data. For example, Microsoft's SenseWeb project [5] provides an infrastructure, including an underlying Microsoft SQL Server database, for archiving massive amounts of sensor data that might be compressed and distributed over multiple computers. SenseWeb also maintains suitable data indexes and enables efficient query processing to help users quickly explore the dataset to find features for detailed analysis [5-7]. But even with these capabilities, SenseWeb hits just the tip of the iceberg of the challenging data management tasks facing environmental scientists. Additional tools are necessary to efficiently integrate sensor data with relevant context and provide data provenance. Querying such data in a unified framework remains challenging. More research is also needed to deal with uncertain data that comes from noisy sensors and to handle the constant data flow from distributed locations.

To better understand environmental phenomena, scientists need to derive and apply various models to transform sensor data into scientific and other practical results. Database technology can help scientists to easily integrate observational data from diverse sources, possibly distributed over the Internet, with model assessments and forecasts—a procedure known as *data assimilation.* Sophisticated data mining techniques can allow scientists to easily explore spatiotemporal patterns of data (both interactively as well as in batch on archived data). Modeling techniques can provide correct and timely prediction of phenomena such as flooding events, landslides, or avalanche cycles, which can be highly useful for intervention and damage prevention, even with just a few hours of lead time. This very short-term forecasting is called *nowcasting* in meteorology.

Scientists in the Swiss Experiment project[1] have made progress in useful data assimilation and nowcasting. One case study in this project applies advanced sensors and models to forecasting alpine natural hazards [8]. A refined nowcast relies on the operational weather forecast to define the target area of a potential storm that

---

[1] www.swiss-experiment.ch

would affect a small-scale region (a few square kilometers) in the mountains. The operational weather forecast should allow sufficient time to install local mobile stations (such as SensorScope stations[2]) and remote sensing devices at the target area and to set up high-resolution hazard models. In the long term, specialized weather forecast models will be developed to allow much more precise local simulation.

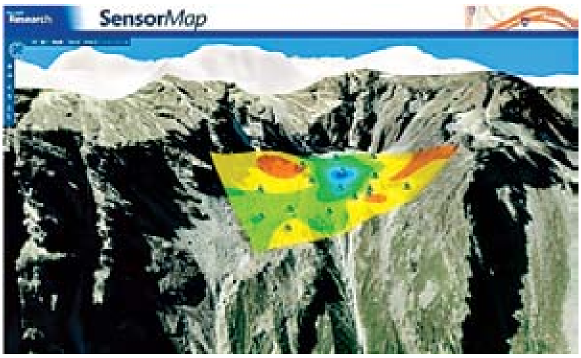

FIGURE 3.
SensorMap showing temperature distribution overlaid on 3-D mountain terrain.

To increase the public's environmental awareness and to support decision and policy makers, useful findings from scientific experiments must be presented and disseminated in a practical fashion. For example, SenseWeb provides a Web-based front end called SensorMap[3] that presents real-time and historical environmental factors in an easy-to-understand visual interface. It overlays spatial visualizations (such as icons showing current air pollution at a location or images showing distribution of snowfalls) over a browsable geographic map, plays the visualizations of selected environmental datasets as a movie on top of a geographic map, and shows important trends in historic environmental data as well as useful summaries of real-time environmental data. (See Figure 3.) At present, such platforms support only a limited set of visualizations, and many challenges remain to be solved to support the more advanced visualizations required by diverse audiences.

WORLDWIDE ENVIRONMENTAL MONITORING

We have described the next-generation environmental monitoring system as isolated—focused on a particular region of interest such as a mountain range, ice field, or forest. This is how such environmental systems are starting to be deployed. However, we foresee far more extensive monitoring systems that can allow scientists to share data with one another and combine and correlate data from millions of

---

[2] www.swiss-experiment.ch/index.php/SensorScope:Home
[3] www.sensormap.org

sensors all over the world to gain an even better understanding of global environmental patterns.

Such a global-scale sensor deployment would introduce unprecedented benefits and challenges. As sensor datasets grow larger, traditional data management techniques (such as loading data into a SQL database and then querying it) will clearly prove inadequate. To avoid moving massive amounts of data around, computations will need to be distributed and pushed as close to data sources as possible [7]. To reduce the storage and communication footprint, datasets will have to be compressed without loss of fidelity. To support data analysis with reasonable latencies, computation should preferably be done over compressed data [9]. Scientific analysis will also most likely require additional metadata, such as sensor specifications, experiment setups, data provenance, and other contextual information. Data from heterogeneous sources will have to be integrated in a unified data management and exploration framework [10].

Obviously, computer science tools can enable this next-generation environmental science only if they are actually used by domain scientists. To expedite adoption by domain scientists, such tools must be intuitive, easy to use, and robust. Moreover, they cannot be “one-size-fits-all” tools for all domains; rather, they should be domain-specific custom tools—or at least custom variants of generic tools. Developing these tools will involve identifying the important problems that domain scientists are trying to answer, analyzing the design trade-offs, and focusing on important features. While such application engineering approaches are common for non-science applications, they tend not to be a priority in science applications. This must change.

CONCLUSION

The close collaboration between environmental science and computer science is providing a new and better way to conduct scientific research through high-resolution and high-fidelity data acquisition, simplified large-scale data management, powerful data modeling and mining, and effective data sharing and visualization. In this paper, we have outlined several challenges to realizing the vision of next-generation environmental science. Some significant progress has been made in this context—such as in the Swiss Experiment and SenseWeb, in which an advanced, integrated environmental data infrastructure is being used by a variety of large environmental research projects, for environmental education, and by individual scientists. Meanwhile, dramatic progress is being made in complementary fields such as basic sensor technology. Our expectation is that all of these advances in instrumenting the Earth will help us realize the dreams of next-generation environmental science—allowing scientists, government, and the public to better understand and live safely in their environment.

REFERENCES

[1] M. Bavay, M. Lehning, T. Jonas, and H. Löwe, “Simulations of future snow cover and discharge in Alpine headwater catchments,” *Hydrol. Processes,* vol. 22, pp. 95–108, 2009, doi: 10.1002/hyp.7195.

[2] M. Lehning, H. Löwe, M. Ryser, and N. Raderschall, “Inhomogeneous precipitation distribution and snow transport in steep terrain,” *Water Resour. Res.*, vol. 44, 2008, doi: 10.1029/2007WR006545.

[3] N. Raderschall, M. Lehning, and C. Schär, “Fine scale modelling of the boundary layer wind field over steep topography,” *Water Resour. Res.*, vol. 44, 2008, doi: 10.1029/2007WR006544.

[4] N. Dawes, A. K. Kumar, S. Michel, K. Aberer, and M. Lehning, “Sensor Metadata Management and Its Application in Collaborative Environmental Research,” presented at the 4th IEEE Int. Conf. e-Science, 2008.

[5] A. Kansal, S. Nath, J. Liu, and F. Zhao, “SenseWeb: An Infrastructure for Shared Sensing,” *IEEE MultiMedia,* vol. 14, no. 4, pp. 8–13, Oct. 2007, doi: 10.1109/MMUL.2007.82.

[6] Y. Ahmad and S. Nath, “COLR-Tree: Communication Efficient Spatio-Temporal Index for a Sensor Data Web Portal,” presented at the Int. Conf. Data Engineering, 2008, doi: 10.1.1.65.6941.

[7] A. Deshpande, S. Nath, P. B. Gibbons, and S. Seshan, “Cache-and-Query for Wide Area Sensor Databases,” Proc. 22nd ACM SIGMOD Int. Conf. Management of Data Principles of Database Systems, 2003, doi: 10.1145/872757.872818.

[8] M. Lehning and C. Wilhelm, “Integral Risk Management and Physical Modelling for Mountainous Natural Hazards,” in *Extreme Events in Nature and Society,* S. Albeverio, V. Jentsch, and H. Kantz, Eds. Springer, 2005.

[9] G. Reeves, J. Liu, S. Nath, and F. Zhao, “Managing Massive Time Series Streams with MultiScale Compressed Trickles,” *Proc. 35th Int. Conf. Very Large Data Bases,* 2009.

[10] S. Nath, J. Liu, and F. Zhao, “Challenges in Building a Portal for Sensors World-Wide,” presented at the First Workshop on World-Sensor-Web, 2006, doi: 10.1109/MPRV.2007.27.