Environmental Modelling & Software xxx (2015) 1–9

Contents lists available at ScienceDirect

Environmental Modelling & Software

journal homepage: www.elsevier.com/locate/envsoft

Integrating modelling and smart sensors for environmental and human health[☆]

Stefan Reis [a, j, *], Edmund Seto [b], Amanda Northcross [c], Nigel W.T. Quinn [d], Matteo Convertino [e], Rod L. Jones [f], Holger R. Maier [g], Uwe Schlink [h], Susanne Steinle [k], Massimo Vieno [a], Michael C. Wimberly [i]

[a] Natural Environment Research Council, Centre for Ecology & Hydrology, Bush Estate, Penicuik EH26 0QB, United Kingdom
[b] Department of Occupational & Environmental Health Sciences, School of Public Health, University of Washington, 1959 NE Pacific Street, Seattle, WA 98195, USA
[c] Department of Environmental and Occupational Health, George Washington University, 950 New Hampshire Ave. NW, Washington, DC 20052, USA
[d] HydroEcological Engineering Advanced Decision Support, Berkeley National Laboratory, 1 Cyclotron Road Bld 64-209, Berkeley, CA 94720, USA
[e] HumNat Lab, School of Public Health, Division of Environmental Health Sciences & Health Informatics Program, Institute on the Environment and Institute for Engineering in Medicine, University of Minnesota, Twin-Cities USA
[f] Department of Chemistry, Cambridge University, United Kingdom
[g] School of Civil, Environmental and Mining Engineering, The University of Adelaide, Adelaide SA 5005, Australia
[h] UFZ Helmholtz Centre for Environmental Research, Permoserstraße 15, 04318 Leipzig, Germany
[i] Geospatial Sciences Center of Excellence, South Dakota State University, Brookings, SD 57007-3510, USA
[j] University of Exeter Medical School, Knowledge Spa, Truro, TR1 3HD, United Kingdom
[k] Institute of Occupational Medicine, Research Avenue North, Riccarton, Edinburgh, EH14 4AP, United Kingdom

ARTICLE INFO

Article history:
Received 22 February 2015
Received in revised form
9 June 2015
Accepted 10 June 2015
Available online xxx

Keywords:
Integrated modelling
Environmental sensors
Population health
Environmental health
Big data

ABSTRACT

Sensors are becoming ubiquitous in everyday life, generating data at an unprecedented rate and scale. However, models that assess impacts of human activities on environmental and human health, have typically been developed in contexts where data scarcity is the norm. Models are essential tools to understand processes, identify relationships, associations and causality, formalize stakeholder mental models, and to quantify the effects of prevention and interventions. They can help to explain data, as well as inform the deployment and location of sensors by identifying hotspots and areas of interest where data collection may achieve the best results. We identify a paradigm shift in how the integration of models and sensors can contribute to harnessing 'Big Data' and, more importantly, make the vital step from 'Big Data' to 'Big Information'. In this paper, we illustrate current developments and identify key research needs using human and environmental health challenges as an example.

1. Introduction

1.1. Background

Models have become widely used and indispensable tools to assess effects of environmental factors on human and ecosystem health. Applications include, but are not limited to, the modelling of environmental processes, such as the emission, dispersion and environmental fate of pollutants in atmospheric (e.g., Vieno et al., 2010, 2014), terrestrial and aquatic environments (e.g., Wu et al., 2014a,b; Perelman and Ostfeld, 2013), the quantification of human exposures to these pollutants (e.g., McKone, 1993; MacIntosh et al., 1995), the risks and public health burdens from exposures to environmental pollutants (e.g., Lim et al., 2012; Schlink et al., 2010), the dynamics of biomarkers in relation to drugs and pathogens, and the efficacy of efforts to control the consequences of these processes on human health (e.g., May et al., 2008; Wu et al., 2014b), and the quantification of stakeholder mental models for optimal decision making (Wood et al., 2012; Voinov et al., 2014; Boschetti, 2015). Models have important uses in examining the accidental or natural release of chemicals, radionuclides, volcanic

☆ Thematic Issue on Modelling Health Risks.

* Corresponding author. Natural Environment Research Council, Centre for Ecology & Hydrology, Bush Estate, Penicuik EH26 0QB, United Kingdom. Tel.: +44 131 445 8507.
E-mail address: srei@ceh.ac.uk (S. Reis).

http://dx.doi.org/10.1016/j.envsoft.2015.06.003

ash, or pathogens in the environment. Generally, both physical process-based and statistical models are calibrated and validated against observed environmental data, which have traditionally been obtained from few, typically sparsely distributed routine monitoring stations, or from costly short-term field measurement studies. In both cases, the spatial and temporal performance of models is evaluated against relatively few directly measured data points.

Conversely, the capabilities and availability of cheaper, more sensitive and sophisticated sensors for gases, particulates, water quality, noise and other environmental measurements have improved and are enabling researchers to collect data in unprecedented spatial, temporal and contextual detail (Stocker et al., 2014). These sensors range from bespoke devices designed for specific applications, to those found on more mainstream personal devices, such as smartphones. In some cases, people may act as environmental sensors by reporting what they see, hear and feel by participating in the crowdsourcing of environmental conditions (Salathé et al., 2012). By leveraging widely available computing, networking and sensor technologies, many new sensor systems are relatively low-cost compared with technologies used in established monitoring networks. Low-cost sensing has the potential to broaden the scale of environmental measurements, both through improving the feasibility of larger scale monitoring networks and by empowering non-traditional researchers, such as community groups, environmental justice organizations and citizen scientists to participate in collecting environmental, biological and clinical data. Hence, new sensors may potentially solve the limitations of traditional environmental monitoring by improving data collection in currently under-monitored areas, including urban areas with large spatio–temporal variations in pollutant concentrations and exposures, as well as rural areas and developing countries where few conventional monitoring sites may be available. One challenge of ubiquitous sensing is a potential explosion of data collected by multiple groups for different purposes, with differing accuracy, precision and hence data quality. Advances in data science and data fusion are vital to enable researchers to make best use of the vast amounts of additional, heterogeneous measurement data. Environmental models will potentially play an important role in integrating these data as inputs to refine and quantify important environmental relationships and processes (Banzhaf et al., 2014; Galelli et al., 2014). Models may also benefit from having new data to use as calibration, validation, and assimilation points to improve the outputs of increasingly complex and downscaled models. Documenting, understanding and implementing quality assurance and quality control processes that are responsive to heterogeneous sensor data will be critical if they are to be used for modelling. Modellers are not only users of sensor data, but can also help to inform the sensor community by identifying existing modelling uncertainties, sensitivities, and constraints that could benefit from improved empirical data, so as to guide what, when and where sensors should measure. Ultimately, data from both sensors and models provides evidence to policy decision-makers, hence the role of stakeholders and their interaction with the scientific community is a vital area for discussion in this context.

1.2. Approach

This paper presents the potential benefits and opportunities available to the modelling community through improved adoption and integration of sensor technologies. For the purpose of this paper, we use the term 'data' to specifically identify raw and unprocessed observations specifically, and 'information' to illustrate data that has undergone validation, quality assurance/quality control (QA/QC) and (objective-based) interpretation to be used for decision making. Finally, as 'Big Data' does not have a concise and generally accepted, scientific definition to date (the moving target presented by defining a volume of data that is pushing the boundaries of current processing capabilities), we adopt the widely used definition by Doug Laney and applied by industry (e.g. SAS, 2015), which stipulates 'Big Data' as being determined by the three Vs, *volume*, *velocity* and *variety*. These three aspects are important when monitoring a wide variety of data and are therefore highly relevant to the purposes of this paper.

We discuss cases in which models may benefit from large datasets emerging from new sensor networks, particularly in terms of increased model accuracy through better calibration/validation and global uncertainty/sensitivity analyses (Saltelli et al., 2010), while also benefiting groups designing, deploying, and analysing data from sensor networks. Fig. 1 presents a conceptual framework in which both the sensing and modelling communities play integral roles in information science, with this science ultimately operating within and informing policy. Critically, missing from this conceptual diagram are the details of data management, processing and flows.

The environmental monitoring community produces data that are subject to QA/QC, which then could be used on their own as empirical data related to environmental processes. However, data could also flow to the modelling community as inputs and calibration and validation points for modelling. The combination of quality data and a validated conceptual model that incorporates state of the science understanding of environmental and disease processes can be explored via simulation, scenario, and global sensitivity and uncertainty analyses to produce information relevant for policy and planning. In this framework, we acknowledge that all measurement data are subject to error, and can benefit from QA/QC to filter the data for errors and anomalies leading to the use of models for data synthesis. Models can also vary in complexity

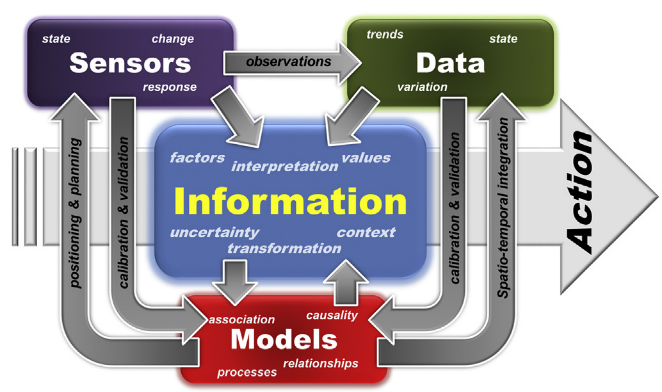

Fig. 1. A conceptual model for sensor-model integration illustrating the complex system required for the development of evidence and data based action (e.g. policy development and implementation). The central role of information (factors, interpretation, values, uncertainty, transformation and context) is highlighted. Here, *information* is also depicted as input to the modelling stage, e.g., to reduce the size of 'Big Data' by extracting only data with high information value for the question being asked (Shannon and Weaver, 1949; Lazer et al., 2014; Galelli et al., 2014; Convertino et al., 2014, 2015). Information in general and the policy questions to be assessed in particular include value judgements (Voinov et al., 2014). This can affect the interpretation of data, for instance by identifying priorities and setting the context for analyses. A robust science-policy interface (Reis et al., 2012) can establish trust in data and information generated by sensors and models. This is essential, as transparency and traceability of data flows and processing methods are key requirements to assess the quality of data. Such science-policy interfaces need to reflect stakeholders' conceptual and mental models (alternatives, preferences, utility, and drivers) embedded in decision science frameworks, integrating those (mainly) qualitative models with (quantitative) biophysical models and decisions (see Wood et al., 2012; Boschetti, 2015 and Section 7).

Please cite this article in press as: Reis, S., et al., Integrating modelling and smart sensors for environmental and human health, Environmental Modelling & Software (2015), http://dx.doi.org/10.1016/j.envsoft.2015.06.003