

Chapter 1

Computational Models for Environmental Systems

Francesca Pianosi

Abstract Mathematical models are increasingly used to represent, understand, predict and manage environmental systems. In recent years, increasing data availability and computing power have induced dramatic changes in the way environmental models are developed and used. Models are being constructed at increasingly high resolution and complexity, while computer-based simulations allow for unprecedented uses of environmental models and data. Nonetheless, the complexity of environmental systems and the uncertainty in environmental data still pose a number of challenges in the construction, validation and use of environmental models. In this chapter, the relation between observations, modelling and numerical computing, and their implications in the environmental domain will be discussed through several examples. Without pretending to be exhaustive, the selection of examples mainly aim at highlighting the variety of contexts, application domains and modelling purposes affected by new computing technology.

Keywords Environmental modelling · Physically-based and data-driven modelling · Environmental uncertainty

1.1 Introduction

Mathematical models are increasingly used to represent, understand, predict and manage environmental systems, including the atmospheric system, water resources, ecosystems, agro-forestry systems, etc. [1]. The application domains and purposes of environmental models are manifold [2]. In some contexts, models are used to test

F. Pianosi (✉)

Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Milan, Italy
e-mail: francesca.pianosi@polimi.it

Department of Civil Engineering, University of Bristol, Bristol, UK
e-mail: francesca.pianosi@bristol.ac.uk

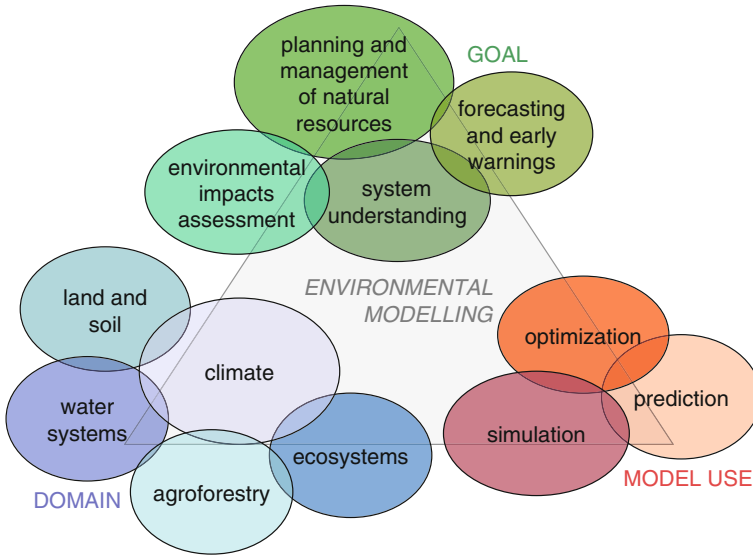


Fig. 1.1 Application domain, use and goal of environmental models

hypothesis, validate theories and ultimately improve our understanding of natural processes. Alternatively, models may be used for forecasting future system conditions or for assessing their response to human interventions via model simulation and what-if analysis. In many applications, models are expected to inform and support decision-making processes, from early warning systems to the sustainable planning and management of natural resources (Fig. 1.1).

Ever growing computing power on the one hand, and increasing data availability on the other, are revolutionizing the way environmental models are constructed and used. Advances in pervasive sensor networks [3] and remote sensing techniques [4] can provide environmental data at local and global scale at increasingly higher temporal and spatial resolution. Cheaper and more easily accessible computing facilities allow for processing this huge data flow and including them into more and more sophisticated models and innovative applications [5]. In this chapter the contribution of numerical computing to environmental modelling will be discussed through several examples. Without pretending to be exhaustive, the selection of examples mainly aim at highlighting the variety of contexts, application domains and modelling purposes affected by new computing technology.

The example in Sect. 1.2 comes from the field of ecology. The use of mathematical models to describe population dynamics has a long history and started much before the advent of computers.¹ The equations appearing in these models can sometimes

¹ The first quantitative population growth model can be traced back to the *Essay on the Principle of Population* by Thomas Robert Malthus in 1798. Another classical topic of ecology is the prey-predator dynamics, that was first modeled by Alfred J. Lotka and Vito Volterra in the 1920s.

be solved by pencil and paper, at least under specific conditions, for instance at equilibrium. However, computer-based numerical integration has allowed for solving ecological models virtually at any conditions, and this in turn has dramatically enlarged our possibility of testing models and their underlying theories, as shown in this example.

The example in Sect. 1.3 deals with water systems. Here again the mathematical models used to describe rivers and reservoirs translate simple hydraulic principles that have been known for decades, while the reservoir operation model is based on the Bellman optimality principle [6] dating back to the late 1950s. However, the implementation of these simulation and optimization models at the resolution scale used in this example would have been impossible up to few years ago because of limited computing power.

These examples illustrate how numerical computing has unlocked new uses of existing models with dramatic impacts from both the scientific and engineering perspective. Besides model use, numerical computing has also significantly impacted model construction. This is true in many different instances but is especially remarkable in the context of the so called *empirical models*, which simply could not exist without the computer programs that are used for their construction. This is discussed in Sect. 1.4.

Still, the improvement in our modelling capacity made possible by increasing availability of data and computing power is likely to confront limits in the environmental domain. The complexity of the investigated processes and the uncertainty in the data are so high that computers are not seemingly to resolve all the conflicting issues inherent in the construction and use of environmental models. This is discussed in Sect. 1.5.

1.2 Numerical Computing and New Understanding of Environmental Systems: Scientific Perspective

The first example comes from the domain of quantitative ecology. Ecological models aim at providing quantitative explanations and predictions of the relationships among living organisms and the environment, including infra-specific and inter-specific processes like population growth, stability and extinction, competition for available resources, predation, symbiosis, parasitism, etc. One issue that has puzzled scientists in the domain of aquatic ecology is the so called “paradox of the plankton”. The paradox consists in a mismatch between model predictions and real-world observations. Competition models predict that, at equilibrium, the number of coexisting species cannot exceed the number of the limiting resources they consume. However, for phytoplankton species the number of limiting resources (e.g., nitrogen, phosphorus, silicon, iron) is very low while empirical observations show that dozens of phytoplankton species coexist. A very interesting contribution to the debate comes

from [7], that provides a solution to the plankton paradox made possible by computer-based simulations.

The competition model is a set of differential equations that describe the rate of growth of the n phytoplankton species and of the k resources. Equations are derived by translating basic principles of mortality, fertility and competition into mathematical relations. By setting to zero all the differential equations, one can derive the equilibrium state of the system. This is the condition when population and resource densities are constant. Usually it is assumed that the ecosystem should spontaneously tend towards this condition and return to it when recovering from external disturbances. Given the relative simplicity of the model, the equilibrium solution can be derived analytically with pencil and paper. It says that the populations with nonzero density (i.e., that can coexist at equilibrium) are in number lower or equal to the number k of resources. The conclusion has been named the “principle of competitive exclusion” and, as anticipated, it is in contrast with empirical experience.

Alternative models have been proposed to explain the observed species diversity of planktonic communities. They include factors external to the phytoplankton dynamics, like selective predators, spatial heterogeneity, or time-varying weather conditions. By using computer-based simulation, instead, [7] demonstrates that the paradox can be resolved using the original competition model, provided that one looks at its simulated behaviour when the equilibrium is not reached. Specifically, numerical integration of the model equations show that: first, when there are at least three species and three resources, the equilibrium may not be reached and the simulated species densities may exhibit oscillations; second, the number of species that keep oscillating in time and do not go to extinction may be higher than the number of resources. As oscillations are not generated by other sources of external variability, this study demonstrates that competition theory is in fact sufficient to explain the coexistence of phytoplankton species.

1.3 Numerical Computing and New Understanding of Environmental Systems: Engineering Perspective

The second example comes from the domain of integrated water resources management. Mathematical models are widely used in this sector to support the operation and planning of water resources systems like rivers, lakes, wetlands, and human artifacts like dams, irrigation and drainage systems, urban supply networks, etc.

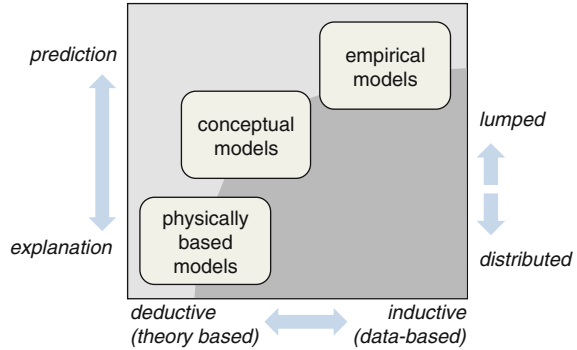
The example here reported is an application developed by the author and others, to the efficient use of regulated lakes and reservoirs. Reservoirs may enhance the economic, social and environmental value of watersheds by enabling water reallocation in space and time. However, often watersheds comprise multiple reservoirs that are operated independently one to another to meet different targets. The lack of coordination generates inefficiency and economic loss and induces conflicts among different water uses.

The lake Como watershed considered in [8] is a typical example. The water system develops along the river Adda in Northern Italy, with a topology common to many Alpine watersheds: a large storage capacity distributed in many hydropower reservoirs in the upper watershed region; a large regulated lake in the middle region; and multiple water consumption users, mostly farmers, in the lower region. Spring snowmelt is the most important contribution to the creation of the seasonal storage, which is reallocated over time according to two different strategies. The lake regulation exploits the accumulated volume in the summer to supply downstream irrigation, while hydropower operators keep their storages full up until the following winter when the demand for energy peaks and the production is more valuable. This results in the potential for conflict between farmers and hydropower companies, which is highest in particularly dry summers, when farmer associations claim that water shortages could be mitigated if the water retained by the hydropower companies were available.

In the study [8], mathematical models are used to compare the system performances under different institutional settings and thus assess the space for improving the overall system efficiency. First, the study simulates the hypothetical condition where a single super-operator has full access to the system data conditions and makes all the decisions simultaneously, balancing upstream (hydropower) and downstream (irrigation) interests. Simulation experiments show that under such a centralized approach there exists a win-win solution in which the irrigation deficit can be significantly reduced without economic loss in the hydropower production. In other terms, the study demonstrates that the main limitations to the current system performance do not stem from physical constraints (e.g., limited storing capacity) but they come from the institutional, legal and operational framework. Subsequently, the centralized operating policy is analyzed in order to gain insights into suitable strategies to foster cooperation among the involved agents. The analysis suggests a coordination mechanism based on constraining the minimum release of upstream hydropower reservoirs in particularly critical situations. Simulations show that this coordinated approach, although suboptimal with respect to the ideal centralized solution, still can significantly improve the historical operation.

The study combines simulation models of the physical components like reservoirs, hydropower plants and the river network, and optimization models that can mimic the decision-making process of reservoir operators. The application of the latter is particularly challenging from the computational standpoint. The optimization algorithm there used, stochastic dynamic programming, has computational complexity that grows exponentially with the number of system variables, so that the application to a multi-reservoir network like the one in [8] would have been computationally unaffordable only few years ago.

Fig. 1.2 Classification of models in the environmental domain



1.4 Numerical Computing and New Modelling Paradigms

Environmental models are often divided into three categories, *physically-based*, *conceptual* and *empirical*. The classification is based on the way models are constructed, i.e., by deduction from a scientific theory or by induction from the data, but also reflects the purpose they are constructed for, i.e., explanation or prediction (Fig. 1.2). Numerical computing has had a significant and sometimes predominant influence on the development and use of all the three model categories.

Physically-based models provide a detailed description of the processes occurring in the system. They consider both temporal and spatial variability and thus take the form of a set of partial differential equations with distributed parameters. Model identification is based on the translation of physical principles into the model equations, and parameter values are mainly decided based on field measurements or laboratory experiments. Historically, the progress in the development of such models was limited by our imperfect knowledge of environmental processes as well as by computer capability. Cheaper and more easily accessible computing power has allowed for simultaneously increasing the number of processes reproduced in the model, their spatial resolution and the length of simulation horizons. For instance, [9] reviews the advances of physically-based climate models boosted by the advent of supercomputers.

Conceptual models give a simplified description of the system functioning. They reproduce only the main processes and usually neglect the spatial dimension or give it a very simplified representation. The model variables represent spatial averages of the quantities of interest and their temporal dynamics is given by a set of differential equations with lumped parameters. Since parameters are not associated to measurable quantities their value must be guessed or inferred from data by minimizing the distance between model outputs and observations. The data-based parameter estimation exercise, or model calibration, can be carried out manually or by means of automatic procedures. Ever increasing availability of data and computing power has dramatically encouraged the use of automatic calibration procedures and nowadays

most of the software packages implementing environmental models also include specific routines for automatic calibration.

As increasing computing power has made model calibration techniques faster and faster, data-based inference can be used not only to derive parameter estimates but also the model structure itself, by repetition and comparison of calibration results under different hypothesized model structures. The models so obtained are called *empirical*, or data-driven or black-box. The latter term highlights the purely predictive nature of these models, which relate system inputs and outputs without any attempt at reproducing the inner system processes. The use of empirical models in the environmental domain has often been questioned because they are deemed to be hardly understood and trusted by their expected users. Nonetheless, they have gained more and more attention because they are usually cheaper to develop and easier to use, requiring less data and computing power at both stages, while providing comparatively good performances at least for operational purposes. Note that for empirical models, numerical computing is not simply a factor influencing the accuracy, usability or range of application of the model, but rather it is the essential reason for their very existence.

Information and communications technology is likely to further shape the way models are constructed, used and verified in even more complex and unpredictable fashions. Mobile devices and online applications can be used to give people an active role in the collection and verification of data that are lately used to feed models, and to disseminate and verify the information derived by models. For instance, [10] reviews a number of recent applications in the water domain, where mobile phones are used for gathering user-recorded water level data, for providing model-based advices to farmers, for disseminating flood forecasts, etc. Social computing will allow for new ways of eliciting distributed knowledge and expertise that can be used to verify and complement with analytic knowledge embodied in mathematical models [11]. Although these applications are still at a prototype stage, their development is gaining increasing interest from researchers and entrepreneurs, and in the next future they may challenge the traditional boundaries between expert-based and mathematical models, as well as between physically-based, conceptual and empirical models.

1.5 Environmental Models, Data and Uncertainty

The discussion in the previous section seems to suggest that increasing computing power can only lead to expand our modelling capability. However, this may be not so obvious, due to the high level of system complexity and uncertainty in the data. In environmental systems many different physical, biological, chemical processes overlap and influence each other, variables are strongly heterogenous in time and space, controlled experiments are often impossible, observations are highly uncertain, even the very definition of the system boundaries is critical and sometimes arbitrary since interfaces between climate, water, soils, vegetation, etc. are not clearcut and reciprocal influences may be negligible or not depending on the scale of interest. Finally,

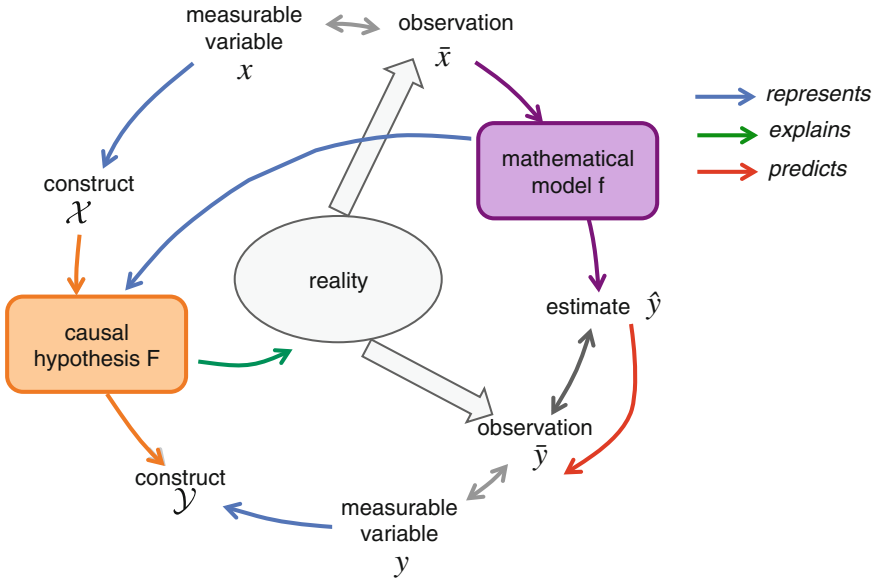


Fig. 1.3 Processes, variables and observations in the construction of models and theories

the influence of human behavior on the earth systems has become so broad and deep in the last decades that scientists have coined the term *Anthropocene* to refer to the geological era we are living in [12], and consequently, in many instances, natural processes cannot be investigated independently from socio-economic ones [13], which adds up new dimensions of complexity to the problem of modelling environmental systems.

In such a context, the fit-to-data, which is the guiding principle of model validation, can be critical. To discuss this topic, consider the formalization in Fig. 1.3, which is a personal elaboration of the one given in [14]. A mathematical model f is expected to provide a relation between an input variable x and an output variable y . In *predictive modelling*, model f is used to produce output predictions. In *explanatory modelling*, model f is used to test the causal hypothesis F . In other words, f is obtained by translation of F into mathematical equations and tested against data, so that if model f fits the data, then the underlying theory F is deemed to explain reality.

Because of measurement errors, the measured input \bar{x} that feeds the model differs from the real input x , and the model estimate \hat{y} must be evaluated against the observed output \bar{y} instead of the actual output y . Still, if measurement errors are small, x and \bar{x} , or y and \hat{y} can be confounded, as we usually do for instance when we say that we predict a *variable* while we actually predict its *measurement*. Also, in formulating the causal hypothesis F , inputs and outputs are usually given in terms of “theoretical constructs” \mathcal{X} and \mathcal{Y} (see again [14] and reference herein), which later are associated to some measurable variables x and y . For instance, in hydrology, theoretical constructs may be the rainfall over a catchment (input) and the river flow

(output), while measurable variables are rainfall and flow at some specific sites where climate and hydrological stations are located. So again there is a slight shifting in meaning when we say that we use models to reproduce “natural processes” because actually we use models to reproduce variables (y) that we consider representative of processes (\mathcal{Y}). And in practice we end up in representing data (\bar{y}) instead of variables, because data is the only thing we possess.

The confusion is acceptable when the choice of observable variables is robust and observation errors are small, so that observations \bar{y} and \bar{x} are really representative of processes \mathcal{X} and \mathcal{Y} . However, this is often not the case in environmental modelling, where system complexity makes the choice of model variables far from obvious, and measurement errors are usually large and sometimes even in the same order of magnitude as the measured variable.²

In such a context, can we conclude that, say, the predictive model f_1 is preferable over the predictive model f_2 , based on the fact that the predictions \hat{y}_1 by f_1 are closer to observations \bar{y} , when measurement errors ($y - \bar{y}$) may be the same order of magnitude as the difference between model predictions ($\hat{y}_1 - \hat{y}_2$)? Or, should we reject theory F because the associated model f does not fit well the data, when we know that those data contain large errors? And maybe the very choice of the model variables x and y as representative of constructs \mathcal{X} and \mathcal{Y} is doubtful?

In practice, such controversies are resolved by environmental modellers based on their judgement and personal experience. Often, models f are trusted because the underlying theories F . In other words, the explanatory process is reversed and, instead of looking at the model’s fit to data in order to confirm the theory, it is the theory that is used to motivate the credibility of the model, even against observations. In fact, a low fit-to-data of model f is often taken as an evidence of poor data quality rather than of a weak underlying theory. Sometimes this is justified based on previous successful applications of the theory to other case studies. Still, the issue about which should be trusted more, theory or data, is open and underlies the long-standing controversy between supporters and opponents of the empirical modelling approach [18].

According to Beven [19], new computing technology will contribute to unravel the matter. As more and more data and computing power become available and virtually every place can be represented by mathematical models, emphasis will

² Uncertainty stems from multiple sources including inaccuracy of measuring devices, errors introduced by preprocessing, problems of sampling. The quantification of such uncertainty has become an important research field. Some examples from the water systems domain are given in [15–17]. Mittelbach et al. [17] provides an example of problems arising in the calibration of measuring devices. They investigate the accuracy of soil moisture sensors and show that in field conditions none of the evaluated sensors has a level of performance consistent with the respective manufacturer specifications developed in laboratory conditions. Baldassarre and Montanari [16] investigates preprocessing errors and define a methodology for quantifying the uncertainty of river flow data, which are obtained from river stage data by application of rating curves. They show that in their case study, overall errors may be as large as 42.8% (at the 95% confidence level) with an average value of 25.6%. An example of sampling difficulties is given in [15] when discussing river quality modelling, where monitoring programs have sampling frequency and location designed for regulation purposes and often inadequate to capture the system dynamics as required for successful model identification.

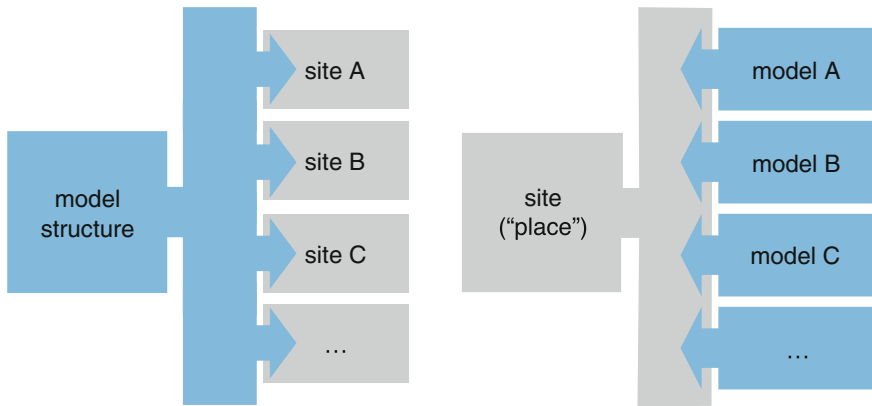


Fig. 1.4 From learning about processes to learning about places

change from a process of learning about model structures (theories) to a process of learning about the specific features of particular places. In other words, if in the past environmental modellers used to search for generalized model structures that could be efficiently adjusted (calibrated) and applied to multiple sites, exchanging accuracy for generality, in the next future the abundance of data and computing power will make it possible to investigate each specific site “from scratch”, testing different model structures and designing the ad hoc combination that guarantees the best possible performance for a specific site and a specific purpose (Fig. 1.4). In the words of [19], “a new generation of environmental models [will appear] that are geared towards the management of specific places rather than general process representation”.

1.6 Conclusions

Computer technology is changing the way environmental models are constructed and used in many respects. Existing models are being used in new ways to perform simulation/optimization experiments in a wider range of conditions or at more realistic resolution steps, providing new insights on system functioning and thus contributing to advance our scientific knowledge as well as supporting better management of natural resources. New models are being built with increasing complexity of structures while empirical models are becoming a viable alternative to traditional physically-based models. Still, because of the complexity of the investigated systems, the multiplicity of interacting components, and the high level of uncertainty in environmental data, the identification and validation of environmental models can be regarded as a “wicked problem” that rarely has a true-or-false solution. The choice of the most appropriate model structure or parameterization is not univocal and it may vary with the scale and purpose of the modelling exercise and change over time as new knowledge becomes available. Human expertise is likely to keep playing a

crucial role in the construction of environmental models and the interpretation of their results. By multiplying the mechanisms for information gathering and dissemination, information and communications technology itself will possibly contribute to reinforcing the integration of human and computer intelligence in the environmental modelling domain.

References

1. Wainwright J, Mulligan M (2004) *Environmental modelling: finding simplicity in complexity*. Wiley, Chichester
2. Jakeman A, Leitch R, Norton P (2006) Ten iterative steps in development and evaluation of environmental models. *Environ Model Softw* 21:602–614
3. Hart J, Martinez K (2006) Environmental sensor networks: a revolution in the earth system science? *Earth Sci Rev* 78(3–4):177–191
4. Butler D (2007) Earth monitoring: the planetary panopticon. *Nature* 450:778–781
5. Lux T, Sydow A (2005) Special issue on environmental modelling. *ERCIM News* 61
6. Bellman R (1959) *Dynamic programming*. Princeton University Press, Princeton
7. Huisman J, Weissing F (1999) Biodiversity of plankton by species oscillations and chaos. *Nature* 402(25):407–410
8. Anghileri D, Castelletti A, Pianosi F, Soncini-Sessa R, Weber E (2013) Optimizing watershed management by coordinated operation of storing facilities. *J Water Res (PI-ASCE In press)*, doi: 10.1061/(ASCE)WR.1943-5452.0000313
9. Washington W, Buja L, Craig A (2008) The computational future for climate and earth system models: on the path to petaflop and beyond. *Philos Trans R Soc A* 367:833–846
10. Jonoski A, Alfonso L, Almoradie A, Popescu I, van Andel S, Vojinovic Z (2012) Mobile phone applications in the water domain. *Environ Eng Manag J* 11(5):919–930
11. Fraternali P, Castelletti A, Soncini-Sessa R, Ruiz CV, Rizzoli A (2012) Putting humans in the loop: social computing for water resources management. *Environ Model Softw* 37:68–77
12. Crutzen PJ, Stoermer EF (2000) Global change newsletter. *The Anthropocene* 41:17–18
13. Editorial (2011) Welcome to the anthropocene. *The Economist* May 28
14. Shmueli G (2010) To explain or to predict? *Stat Sci* 25(3):289–310
15. Berthouex P, Brown L (1994) *Statistics for environmental engineers*. CRC Press, Boca Raton
16. Di Baldassarre G, Montanari A (2009) Uncertainty in river discharge observations: a quantitative analysis. *Hydrol Earth Syst Sci* 13:913–921
17. Mittelbach H, Lehner I, Seneviratne S (2012) Comparison of four soil moisture sensor types under field conditions in Switzerland. *J Hydrol* 430–431:39–49
18. Young P, Parkinson S, Lees M (1996) Simplicity out of complexity in environmental modelling: Occam’s razor revisited. *J Appl Stat* 23(2–3):165–210
19. Beven K (2007) Towards integrated environmental models of everywhere: uncertainty, data and modelling as a learning process. *Hydrol Earth Syst Sci* 11(1):460–467