Advances in Water Resources **15** (1992) 1–3

Editorial

In King Henry IV of Shakespeare (Act III, Scene 1), Glendower boasts to Hotspur: 'I can call spirits from the vasty deep'. 'Why, so can I, or so can any man', responds Hotspur unimpressed. 'But, will they come, when you do call for them'.

The call for validation is probably as old as the first theory propounded by man. The history of science is the continuing story of human endeavour to describe and predict nature. Most early theories had only a descriptive purpose. Although one would presume that models describing the present may be also suitable for predicting the future (a presumption not always true), those early theories were rarely used to make predictions. This may be one of the reasons why early theories could remain 'acceptable' for a long time. As humans intensified their quest of harnessing the forces of nature, the purpose of theories shifted from description to prediction. This trend can be recognized at the dawn of the Renaissance when many old theories were discarded, either because they had little utility as predictive tools or they failed to withstand the test of 'validity'.

For the purpose of illustrating some of these ideas, let us consider the phenomenon of free fall of objects as approached by Galileo and Aristotle (the description given here is due to R.H. March, *Physics for Poets*, 2nd edn. McGraw-Hill, New York, 1978). Aristotle proposed that a falling object acquires instantaneously a constant speed of fall which is dependent on medium type and the object weight. Galileo, on the other hand, postulated that, in the absence of the medium resistance, the speed of falling objects is zero at the onset of fall and increases linearly with time; the speed of fall being independent of the object shape and weight. Fortunately, both of these theories are well-posed in the sense that it is, in principle, possible to prove them wrong by doing measurements. Typical results of such measurements will prove surprising. As illustrated in Fig. 1, should we actually measure the speed of fall of an object and plot it against time, we would see that for all practical purposes, there is a very good agreement between Aristotle's theory and reality. Whereas, Galileo's theory is far from satisfactory based on any goodness-of-fit criterion. It is well known, of course, that Galileo's theory does not fit the measurements because these are affected by the medium resistance. When accounting for such effects, it would be possible, using Galileo's theory, to obtain a fit even better than that of Aristotle. Alternatively, one should provide data from experiments carried out in vacuum; it will then become evident that Aristotle's theory is indeed very wrong. The point is that describing nature and providing a fit to measurements is necessary but not sufficient. A model is useful only if it can be used to make predictions for new conditions and future situations. Aristotle's theory does not have a predictive ability. The horizontal line in Fig. 1 can be drawn only after a set of measurements, for a given object and in a given medium, have been performed. For each new situation and each new object the experiment has to be carried out anew. It is important to realize that sometimes a simple non-physical theory may appear to fit the data better than a less simple but physically based theory. That does not, however, prove that the simple theory is a better one. One may contend that a good and useful model is one with a predictive ability. A second characteristic of a useful model is that it can in principle be proven wrong. The concept of validation is relevant, and essential, to models which have these two characteristics.

One may wonder why there is a need to deal with validation as a separate concept and why there has been so much debate about it in recent years. We believe that this is because traditional modelling approaches have not been successful in hydrology. A standard modelling procedure consists of stating basic laws and building them into a computer code, which can then be used for simulating the performance of the real system under various conditions. This methodology has worked well when applied to man-made prototypes (e.g. bridges, dams, machinery); the most common case in technological developments. It has failed, however, when applied to models in hydrology (and other natural sciences) because of complications associated with spatial and temporal scale effects. Geo-hydrological media are mostly heterogeneous and virtually impossible to characterize sufficiently; processes that are not relevant at small scales become dominant at larger scales; new processes emerge as a result of changes in scale. This has become even more apparent in recent decades in the wake of the increased demand for long-term accurate predictions, especially for hazardous waste disposal. The problem is that, following the traditional point of view, many people view validation as a foolproof check of a model's accuracy in predicting reality. Experience has taught that uncertainties and mechanisms associated with large spatial and temporal scales, commonly encountered in geo-hydrology, defy such a level of model validation as traditionally perceived. In view of the above discussion, one is not

Advances in Water Resources (**15**) (1992)–

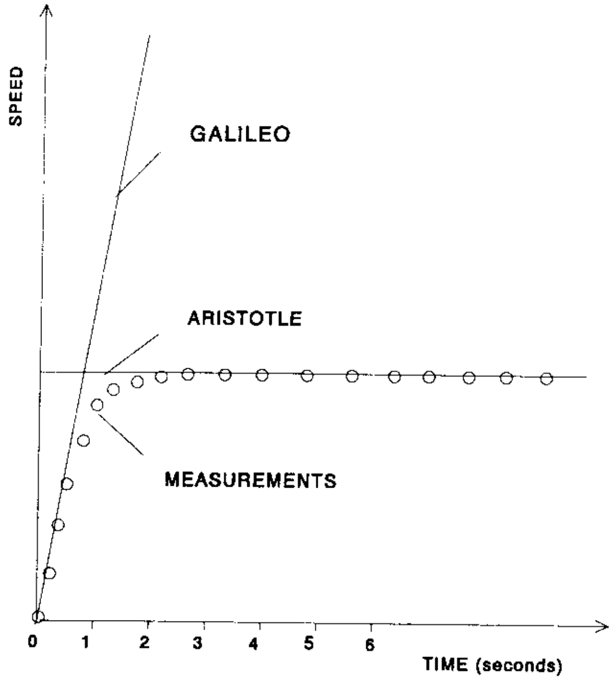

Fig. 1. Free fall of objects (after March, 1978).

surprised to find validation hard to define. Many definitions exist but there is no consensus except perhaps for what one may call 'ideal validation'. Ideal validation may be defined as the process of substantiating that a model possesses a 'satisfactory' degree of accuracy and certainty within its entire domain of applicability and over the entire spatial and temporal scales for which the model is intended to be used; this would lead to an unqualified validity of the model. For reasons described earlier, this is impossible to achieve for geo-hydrological models. Therefore, many scientists resent the use of the label 'validated model' or even the term 'model validation'. Instead, they prefer to talk about 'invalidation', 'achievable validation', 'partial validation', 'provisional validation', etc. Various validation procedures consist primarily of the following three ingredients: (1) establishing the ability and adequacy of a model for making predictions; (2) comparing model predictions to actual measurements; (3) quantifying the uncertainties and inaccuracies of predictions. The proposed aim of various validation procedures is to obtain insight and/or increase confidence in the working of a specific model. The difference between different definitions is not a matter of semantics; it is a question of the perception behind validation. The general concern about a proper definition of aim and scope of validation is a legitimate one, given the many misconceptions about 'validated models'. Many individuals do not realize that a 'validated model' does not necessarily yield accurate predictions of reality even if it does so once. A theory which has overcome many tests is not ensured of not failing in the next one; theories can be proven wrong, but they cannot be proven right. Thus, validation must not be considered as obtaining a label; one should not seek a 'yes or no' answer to model validation. Validation is an iterative process going through cycles of quantifying inaccuracies of predictions, modification of model to reduce uncertainties of predictions, and back to quantifying inaccuracies. This process should go on as long as the model is in use. Clearly, this approach implies an extremely wide scope for validation which must address issues related to process identification, calibration, error analysis and transmission, hypothesis testing, network and experiment design, and so forth.

The scope of validation issues is reflected in the papers comprising this special issue. In total 11 papers have been selected which will be presented in two parts. Part I (this issue) is comprised of six papers. In the first paper, P. Flavelle discusses the importance of quantifying model accuracy in the framework of model validation. He proposes employing a linear regression analysis of computed values and measured data as a means of quantifying the validity of a model. Although this approach may not provide insight into 'reality', it does provide a convenient way of evaluating model errors. The evaluation method is demonstrated by applying it to a groundwater flow and contaminant transport model of a uranium tailings site. S. Luis and D. McLaughlin cast the process of groundwater modelling in a stochastic framework, which leads to formulating validation as a hypothesis testing problem. The acceptability of model predictions is measured in a stochastic sense. The method is employed to identify model deficiencies and provide standards for model performance. The approach is illustrated by applying it to results of field experiments in an unsaturated porous medium. In the next three papers, a number of validation methodologies employed in the study of transport of solutes in granular and/or fractured porous media are discussed. C.P. Jackson, D.A. Lever, and P.J. Sumner propose a validation framework which consists of calibration, testing predictions, comparison with alternative models, analysis of discrepancies, presentation, and evaluation of implications of the study. They illustrate their approach by applying it to a laboratory experiment on transport in a rock sample. P. Maloszewski and A. Zuber present an extensive review of various calibration and validation methodologies commonly employed in the interpretation of tracer tests. With the aid of many examples from laboratory and field experiments in both granular and fractured media, they illustrate major difficulties and pitfalls encountered in calibration and validation of solute transport models. H. Kimura and M. Munakata employ a 'standard' validation procedure, that is, comparing calculations with experimental results. They apply this procedure to a field experiment on transport of tracers in a fractured rock, with the aim of identifying relevant processes. This volume ends with a paper by L.F. Konikow and J.D. Bredehoeft, who adopt a critical attitude towards the concept of validation. They present a controversial view, shared by many geo-hydrologists, that 'groundwater models cannot be validated'. They see 'model validation' as a useful tool for identifying sources of errors and a better understanding of groundwater problems, leading to model improvements. However, they regard the term 'validation' as misleading and favour abandoning its use. Part I is concluded with a brief introduction by Alf Larson to the international validation projects INTRACOIN, HYDROCOIN, and INTRAVAL. These projects have served as a very useful forum for bringing modellers and experimentalists together to discuss practical as well as philosophical aspects of model validation. Various validation procedures have been put into practice and new concepts have been explored. The importance and usefulness of this kind of international cooperation cannot be overemphasized.

We hope that this Special Issue provides a basis for dealing with various validation issues and a platform for reaching consensus as to what model validation is. Part II of this Special Issue will include five papers dealing with some other aspects of model validation not discussed here, alternative frameworks for validation, and additional case studies.

S. Majid Hassanizadeh
Soil and Groundwater Research Laboratory,
National Institute of Public Health
and Environmental Protection,
PO Box 1, 3720 BA Bilthoven,
The Netherlands

Jesus Carrera
Geotechnical Engineering Department,
Civil Engineering School,
Polytechnic University of Catalonia,
c./Gran Capitan, s/n
08034 Barcelona, Spain