Brief overview of selected groundwater modelling guidelines

MARY C. HILL
US Geological Survey, 3215 Marine St., Boulder, Colorado 80303, USA
e-mail: mchill@usgs.gov

HUGH MIDDLEMIS
Aquaterra, Australia
e-mail: hugh.middlemis@aquaterra.com.au

PAUL HULME
Environmental Agency Science Group – Air, Land, Water; United Kingdom
e-mail: paul.hulme@environment-agency.gov.uk

EILEEN POETER
Colorado School of Mines, International Groundwater Modelling Center, Golden, Colorado, USA
e-mail: epoeter@mines.edu

JOHANNES RIEGGER
University of Stuttgart, Stuttgart, Germany, riegger@iws.uni-stuttgart.de

SHLOMO P. NEUMAN
University of Arizona, Tucson, Arizona, USA
e-mai: neuman@hwr.arizona.edu

HOWARD WILLIAMS
Pattle Delamore Partners Ltd, P.O.Box 389, Christchurch, New Zealand
e-mail: howard.williams@pdp.co.nz

MARY ANDERSON
University of Wisconsin, Madison, Wisconsin,USA
e-mail: andy@geology.wisc.edu

Abstract This paper briefly reviews 12 sets of guidelines for groundwater modelling. The guidelines originate from 8 countries on three continents. After reviewing terminology differences, shared, unique, and conflicting aspects of selected sets of guidelines are presented.
Key words groundwater model guidelines; model calibration; inverse modelling; conceptual models; perceptual models

1. INTRODUCTION

Hydrogeological investigations and groundwater modelling are dynamic and inexact. They are dynamic in the sense that (1) the state of any hydrological system changes with time, (2) new scientific techniques with which to evaluate these systems are continually developed and (3) new data challenge previously held concepts about the systems. They are inexact in the sense that groundwater systems are complicated and are largely inaccessible, so we cannot evaluate them comprehensively in detail, and we invariably do not have sufficient data to do so (even if we had the ability).

Over the last 70 years, many ideas and procedures have been introduced to address hydrogeological investigations, including groundwater modelling. Beginning with the four-part series of articles concluded by Freeze *et al.* (1992), attempts have been made to present these ideas and procedures more comprehensively, and these efforts have evolved such that now there exist many sets of guidelines for the development of groundwater models.

This paper seeks to increase communication and coordination between guideline developers. This goal is thought to be advantageous for two reasons. First, these developers are some of the most active groundwater modellers in the world and it is likely that increased communication will improve all sets of guidelines. Through this effort it is hoped that groundwater modelling will mature more quickly into a method that can be used reliably to investigate and manage groundwater systems. Second, differences in the sets of guidelines are likely to cause confusion for governmental agencies that depend on the guidelines. Such confusion is likely, for example, when managing groundwater systems that cross national boundaries or when groundwater models are used in litigation. Awareness of differences in the sets of guidelines by the modelling community can help address the consequences of inconsistencies.

This paper identifies and presents information about selected sets of these guidelines. There are many sets of guidelines and many of the sets of guidelines are described in manuscripts of considerable length. It is not the intent of this work to consider all sets of guidelines or to comment comprehensively on the sets of guidelines considered. Instead, this work is intended to provide enough information about selected sets of guidelines to encourage communication between guideline developers and users.

Cumulatively, the sets of guidelines cover a wide range of the modelling process, such as determining the scope and objectives, conceptualizing the system, data management, model development, sensitivity analysis, simulating predictions, evaluating prediction uncertainty, documentation and reports, and review. An individual set of guidelines may have a narrower range. This paper includes information about the entire range of the modelling process.

After this introduction, the paper consists of sections 2 through 6. Section 2 lists the selected sets of guidelines considered here and the major people and institutions involved in their development and use. Section 3 lists major differences in terminology used in the different sets of guidelines and identifies the terminology used in this work. Sections 4, 5, and 6 categorize major ideas and procedures presented in the guidelines as (1) shared by most of the sets of guidelines, (2) unique, or (3) currently in conflict. Short discussions are included to clarify the ideas and procedures involved.

2. SETS OF EXISTING GUIDELINES

The sets of guidelines included in this work were chosen mostly based on their level of development, prominence, inclusion of new or controversial ideas, or some combination of these considerations. At a minimum, the sets of guidelines included in this work were required to cover a broad range of the problems encountered when simulating groundwater systems. Most are in whole or in part applicable to many other types of systems as well. Table 1 lists the sets of guidelines, regulatory agencies that use the guidelines, and references. Table 2 lists contacts for the guidelines.

Table 1 Selected sets of groundwater modelling guidelines.

[Governmental organizations: FH-DGG, Hydrogeology Section of the German Geological Society; USGS, United States Geological Survey; USNRC, United States Nuclear Regulatory Commission.]

Name for this report	**Used or officially adopted by regulators?**	**Reference and Comments**
AUS	Murray-Darling Basin Commission (MDBC); informally adopted by State water agencies	Middlemis (2001). Downloaded at: www.mdbc.gov.au/publications/si_and_e_reports.htm. Includes very helpful annotated bibliography and comparison of codes and GUIs.
UK	Environment Agency (England & Wales)	Environment Agency (2002)
GLUE	Ideas used in UK guidelines	Beven (2001, 2004)
FH-DGG	Not officially adopted, but widely used	Arbeitskreis "Hydrogeologische Modelle" FH-DGG (1999, 2002); Riegger (2004)
A&W	Not officially adopted, but widely used	Anderson & Woessner (1992)
USNRC	Hydrogeologic modelling strategy in contractor report is being evaluated and used by USNRC staff as a technical resource	Neuman & Wierenga (2003), NUREG/CR-6805 <http://www.nrc.gov/reading-rm/doc-collections/nuregs/contract/cr6805/> Includes some USGS guidelines.
USGS	Parts used by the US-NRC guidelines	Hill (1998), Hill *et al*. (2001), Tiedeman *et al*. (2003), Rielly & Harbaugh (2004).
ASTM	Used by some at USEPA. Funded by USEPA, USGS, and Dept. of the Navy. Some reference in litigation and RFPs (Requests for Proposals).	Documented in standards. For example: D5447-93e1, apply a model to a site D5490-93, compare model to data D5611-94, sensitivity analysis D5718-95, document application D5880-95, flow & transport modelling D5979-96, conceptualization D5981-96, calibration D6000-96, water level reporting D6170-97e1, select code
DK	Danish EPA	Hans Jørgen Henricksen, GEUS
HarmonIT Harmoniqua Harmoni-ca Harmoni-riv	European Union	Generic Framework papers: http://www.genericframework.org/uk/techdocu.htm http://harmoniqua.wau.nl/Summary.htm quality assurance http://www.harmoni-ca.info/HarmoniCA/Public/index.php
NZ	Ministry for the Environment	Guidelines for audit and review of models. http://www.pdp.co.nz

Table 2 Contact information for the selected sets of groundwater modelling guidelines.

[--, not applicable; ASTM, American Society for Testing and Materials; PNNL, Pacific Northwest National Laboratory; FH-DGG]

Name for this report	**Web access**	**Contact person**	**Professional affiliation, email**	**Country**
AUS	www.mdbc.gov.au/publications/si_and_e_reports.htm	Hugh Middlemis Noel Merrick	Aquaterra Simulations, hugh.middlemis@aquaterra.com.au University of Technology Sydney, nmerrick@uts.edu.au	Australia
UK	--	Paul Hulme Mark Whiteman	Environment Agency: Science Group, paul.hulme@environment-agency.gov.uk… Policy & Process (Hydrogeology), mark.whiteman@environment-agency.gov.uk	United Kingdom
GLUE	--	Keith Beven	University of Lancaster, k.beven@lancaster.ac.uk	United Kingdom
FH-DGG	www.fh-dgg/ak-hgm	Johannes Riegger	University of Stuttgart, riegger@iws.uni-stuttgart.de	Germany
A&W	--	Mary Anderson Bill Woessner	University of Wisconsin, andy@geology.wisc.edu University of Montana. gl_www@selway.umt.edu	USA
USNRC	--	Shlomo Neuman Phil Meyer	University of Arizona, neuman@hwr.arizona.edu PNNL, philip.meyer@pnl.gov	USA
USGS	http://pubs.water.usgs.gov/wri984005/	Mary Hill	U.S. Geological Survey, mchill@usgs.gov	USA
ASTM	www.astm.org	Jim Rumbaugh	Environmental Systems, Inc jrumbaugh@groundwatermodels.com	USA
DK	--	Hans Jørgen Henricksen	GEUS hjh@geus.dk	Denmark
HarmonIT	http://www.harmonit.org	14 European organizations	Project Coordinator harmonit@harmonit.org http://www.genericframework.org/uk/partners.htm	Europe
Harmon-iqua	http://harmoniqua.wau.nl/	Huub Scholten	Wageningen University, huub.scholten@wur.nl	Nether-lands
NZ	http://www.pdp.co.nz	Howard Williams	Pattle Delamore Partners Ltd, Howard.williams@pdp.co.nz	New Zealand

3. TERMINOLOGY

Table 3 lists inconsistencies in terminology encountered in the sets of guidelines considered and the terms used in this paper. Part I of the table lists terms for which the differences are just a matter of style or spelling. Part II lists terms for which the differences reflect fundamental concepts and ideas.

Table 3 Terms that are used differently in different sets of guidelines, and the term used in this paper.

[Terms used in this paper are listed in **bold**. For topics not covered in this paper, a term is not selected. Selected terms were determined in part by the opinions of the authors, which are identified in the third column using the following initials: ma, Mary Anderson; mh, Mary Hill; ph, Paul Hulme; hm, Hugh Middlemis; sn, Shlomo Neuman; ep, Eileen Poeter; jr, Johannes Riegger, hw, Howard Williams. ASTM, American Society for Testing and Materials.]

Meaning to be conveyed	Terms used in guidelines	Author opinions
Pat I. Superficial differences		
Subsurface water[1]	**Groundwater** Groundwater Don't care or conflicted	Sn mh, hm, ma, ph, ep, hw
Simulating, simulated, one who simulates	**modelling/modelled/modeller** modelling/modelled/modeller Don't care	 hm, mh, ep, ph, hw
Quantities measured in the laboratory or field	**Measurements** Observations	mh, ma, hw hm, ma
Measurements or values derived from measurements that are compared with simulated dependent variables	**Observations** Targets	sn, mh, ph, hw hm, ma
The calculated values that are compared to the [observations, targets]	**Simulated equivalents or simulated values** Predictions	mh, ep, hm, ph, hw
Calculated system state for future or hypothesized conditions	**Predictions** Alternative unknown	mh, ep, hm, ph, hw
Part II. Indicative of fundamental differences		
Qualitative description of the hydrology and hydrogeology to be represented in the groundwater flow model.	**Conceptual model** Hydrogeologic Model (HGM) Perceptual model[2]	mh jr ph
Quantitative description of the hydrology and hydrogeology to be represented in the groundwater flow model, including definition of hydrogeologic units, recharge distribution, surface water bodies, and so on -- everything but the values of parameters.	Hydrogeologic Framework Model (HFM)	
Quantitative description of the hydrogeology to be represented in the groundwater flow model, including definition of hydrogeologic units.	**Hydrogeologic Framework Model (HFM)** Hydrogeologic Model Concept Conceptual model[2]	mh, hm, ep jr ph, hw
Quantitative description of hydrogeology and the related processes including respective model calibration and evaluation; ready to use	Hydrogeologic Model	jr
The degree to which a model application resembles or is designed to reproduce the details of the hydrogeological system.	Complexity Fidelity	hm, mh, ep, ph, hw ASTM, jr

Meaning to be conveyed	Terms used in guidelines	Author opinions
The correspondence between the prediction of interest and the level of model complexity	Fidelity	ASTM, jr
A test of the model application by checking if simulated values reasonably match a reserved data set that was excluded during model calibration.	Verification Validation Test	hm, hw FH-DGG mh, ep
Analysis of model results with respect to uniqueness, accuracy, sensitivity and model application range; quantification of uncertainty	Evaluation	jr, hw
Process of comparing simulated and measured values and changing the model to address inconsistencies.	Calibration Model refinement Parameter identification	mh, sn, ep ph, hm, hw
Process of adjusting parameter values to reduce inconsistencies in model fit.	Parameter estimation Calibration	mh, hm ph, hw
Spatial assignment of parameters for a unique and accurate relation between measurements and not directly observable values (volume / mass /energy flows) under hydrogeologic constraints	Calibration	jr
Use of optimization methods to adjust parameter values.	Inverse modelling Parameter estimation using optimization Parameter optimization	 mh, hm ph

1. In Great Britain and New Zealand, only 'groundwater' is used. Both options are used in Europe and North America.
2. Perceptual and conceptual models as presented in the UK guidelines are discussed in section 5.2.

4. SHARED IDEAS AND PROCEDURES

4.1 Overall perspective

The following statement, which is modified slightly from the AUS guidelines, states some basic ideas that are shared by all the sets of guidelines considered:

The aim of most guidelines is to reduce and reveal model uncertainty for the users of modelling studies, including resource management decision makers and the community. This is achieved by promoting transparency in modelling methodologies and encouraging innovation, consistency, and best practice. Guidance is provided to non-specialist modellers and auditors or reviewers of models by outlining the steps involved in scoping, managing, and evaluating the results of groundwater modelling studies. The guidelines serve modelling specialists by providing a baseline set of ideas and procedures from which they can innovate.

The guidelines are intended for use in raising the minimum standard of modelling practice and allowing appropriate flexibility, without limiting necessary creativity or rigidly specifying standard methods. The guidelines also should not limit the ability of modellers to use simple or advanced techniques, appropriate for the study purpose. Techniques recommended in the guidelines may be omitted, altered, or enhanced, subject to the modeller providing a satisfactory explanation for the change and negotiation with the client and/or regulator as required. Not all aspects of the guidelines would necessarily be applicable to every study. It also is acknowledged that standardization of modelling methods will not preclude the need for subjective judgment during the model development process.

The guidelines are to be applied to new groundwater flow modelling studies and reviews of existing models. The guidelines should be seen as a best practice reference point for framing modelling projects, assessing model performance, and providing clients with the ability to manage contracts and understand the strengths and limitations of models across a wide range of studies (scopes, objectives, budgets) at various scales in various hydrogeological settings. The intention is not to provide a prescriptive step-by-step guidance, as the site-specific nature of each modelling study renders this impossible, but to provide overall guidance and to help make the reader aware of the complexities of models, and how they may be managed.

4.2 Model Transparency

The goal of model transparency mentioned in section 4.1 is stressed in many guidelines and, indeed, is a major reason for the guidelines to be developed. Transparency means that the ideas and assumptions used to build a model application are clearly stated and can be tested. The complexity of the systems and the model applications and the tools used to develop groundwater-model applications (including guidelines, visualization software, database software, and so on) rarely result in applications that are completely transparent. However, the goal of transparency is important.

4.3 Valid conceptual models that start simple and build complexity as needed are crucial and fundamental

All sets of guidelines stress the importance of valid conceptual models. Most guidelines suggest some form of parsimony. For example, the UK guideline outlines how conceptual models and the model applications should be continually updated/refined from an initial “appropriately simple” approach. Model updates and refinements arise as ongoing modelling studies and more data on system responses to natural and imposed stresses produce improved understanding of the system processes and interactions. The UK guidelines note that “The first (conceptual) model is not the best and it is not the last”. In the FH-DGG guidelines, this is expressed in their “Hydrogeological Model” HGM as “Model Maintenance.”

A related concept is the step-wise method of model development: Refine the conceptualization (more simple or more complex) and(or) add more parameters as needed to obtain model fit.

Some guidelines also stress consideration of the information provided by the observations. The USGS guidelines present sensitivity measures of the information provided by observations, as discussed in section 5.

The UK guideline outlines a comprehensive modelling approach based on the need to develop understanding of the studied systems. Conceptual models are developed as quantitative descriptions of the real system using observed field values. These are then tested using a variety of methods including lumped water balances, purely investigative numerical models, and during the development of historical numerical models.

Investigative modelling refers to building numerical models of alternative conceptual models to test hypotheses and to define key processes. Trial simulations are run to explore initial understanding without necessarily “calibrating” any model. In the FH-DGG guidelines, this is called the “scenario technique.”

Historical modelling refers to building numerical models that adequately represent the historical behaviour and specifically the key flow mechanisms of the real system. Field data is used both as model input and to compare against the model outputs.

4.4 Consider predictions of interest in model development

All guidelines suggest the importance of considering the model and the calibration in light of predictions, but they differ on how this is accomplished. Some new procedures have been introduced in some sets of guidelines and are described in Section 5.

4.5 Use hydrogeologic data to constrain the model

The hydrogeologic data of concern generally includes the time-invariant data such as stratigraphy, layer elevations and extents, hydraulic conductivity data, and so on. It also can include some time-variable data such as recharge, stream-aquifer interaction, abstraction configuration and stresses, and so on. The emphasis placed on hydrogeologic data differs between the sets of guidelines. Especially, the proper role of hydraulic-conductivity measurements is in contention, as discussed in Section 6.

4.6 Use least-squares objective functions as one measure of model performance

All of the methods encourage the use of some type of least-squares or maximum-likelihood objective function to quantify how well the model fits the observations. These two are the same for a given regression if, as is common, the statistical parameters are known. Alternatives, such as the sum of absolute values, have been used, but rarely and mostly in research papers.

The types of least-squares objective functions commonly used can be classified based on how the quantities are included in the objective function (observations, prior information, or regularization) are weighted. Possibilities include: simple least-squares objective functions that have no weighting or equal weighting, weighted least-squares objective functions that have a diagonal weight matrix, and generalized least-squares objective functions that have a full weight matrix.

An alternative to a single least-squares objective function including all the observations is to divide the observations to create multiple objective functions (a recent reference of this approach is Vrugt *et al*. (2003). The multi-objective functions are each least-squares objective functions, so the agreement cited in this section applies.

4.7 Use other measures of model performance

All guidelines make numerous suggestions about how to use evaluate models. Most performance measures are based on comparing observations to simulated values. Here are a few of the suggestions shared by most sets of guidelines considered.

<u>Use more than hydraulic heads as observations</u>

Hydraulic heads are the most commonly available hydraulic data in most systems. Many of the guidelines directly address the advantage of having other types of hydraulic data, such as flows, advective transport derived from concentrations, concentrations used directly, temperature, and so on. The procedures suggested for including these different types of observations vary; some of the ideas are presented in

Section 6.

Use more than fit to observations to judge a model

To assist the end-user to assess whether model performance is acceptable and meets the level of complexity required, qualitative and quantitative model performance measures are proposed in many guidelines. For example, in the UK 30-year time-variant simulations are common and modelled results are routinely compared to long-term trends and seasonal behavior. The FH-DGG guidelines stress that consideration of the uniqueness and accuracy of model fit is essential. A way to determine these quantitatively is to display the objective function based on least squares versus parameter combinations. The USGS guidelines stress the normality, randomness, and magnitudes of the weighted residuals, and note that very good fits can result from undesirable fitting of observations errors.

Prescriptive performance measures should not be applied blindly, as model performance can only be gauged against observations that are usually imperfect and incomplete, and the model must replicate processes that might be poorly understood or inadequately measured.

The utility of some qualitative comparisons are in dispute. For example the issue of using contoured hydraulic heads is discussed in section 6.5.

4.8 Consider alternative models

All sets of guidelines stress the importance of considering alternative models because system dynamics are rarely clearly defined. The methods used to generate alternative models and evaluate the results vary. Both deterministic and stochastic methods are considered. Generation of alternative models is a very interesting problem that has not been addressed thoroughly.

All methods evaluate alternative models through comparisons with field data and eliminate or reduce emphasis on models that reproduce field data poorly. This idea was originally stressed by the GLUE developers as part of considering only models that adequately fit the observations, and is now widely accepted.

5. UNIQUE IDEAS AND PROCEDURES

Below are short statements from the authors of six of the sets of guidelines describing briefly what they see as unique about their guidelines.

5.1 AUS

Scoping a Modelling Study

The scoping process is a key initial step in a model study, with the outcomes being specific study objectives, model complexity, and the required/available resources of time, budget, data, and technical expertise.

Detailed information is provided in the AUS guide including, for a range of complexity, the broad data requirements, timeframes for model development, broad budget requirements, and examples of specific objectives, for use by project managers in scoping their project.

Model Complexity

The ASTM guides proposed the term "fidelity", which was adapted to "complexity" for the AUS guide. In simple terms, model complexity can be described by the "quick-cheap-good" paradox. The end-user can readily obtain a model with one or two of these three attributes, but not all three. If a model is required to be done quickly, it also can be done cheaply, but the results may not be good enough on which to base important resource development or management decisions. Alternatively, if a good, reliable model is required, then it is not likely to be able to be developed quickly or cheaply. Thus, it is crucial to establish at the scoping stage the specific details of the study objectives, the water resources issues/scenarios, the model purpose, the development stages, and resources.

In less simple terms, the "quick-cheap-good" attributes are better defined in terms of a hierarchical scale of model complexity. The level of model complexity needs to be discussed and agreed upon by the end-user and the modeller to ensure that it suits the study purpose, objectives, and resources available for each study, including long term staged development and technology transfer.

Water managers also should be included in the scoping and model design process (if they are not already part of the project team), as they will use the model results to allocate water resources and/or to assess the impacts of proposed developments and/or to implement resource management policies. It is important for the overall project objectives that potential fatal flaws in the modelling approach are identified and rectified at an early stage, rather than presenting government agencies with the results of a study that may not be regarded as scientifically sound.

Model Reviews

The AUS guide proposes a unique model review framework and detailed checklists, with reviews recommended at all stages throughout the study, consistent with the objectives, scope, scale, and budget of the project. A model review provides a process by which the end-user can check that a model meets the project objectives. It also provides the model developer with a specification against which the modelling study will be evaluated. The level of review undertaken will depend on the nature of the project. Less complex models require less detailed reviews. Reviews necessarily add expense to the modelling process. The client and contractor must be clear at the outset as to which party is to bear the cost of each review. The reviews included in the AUS guidelines are listed in Table 4.

Table 4 Reviews prescribed by the AUS guidelines.

Type of review	Parts of model reviewed	Procedure provided in AUS[1]	Suggested reviewers
Appraisal	Report.	36 questions; App. E	Representative of stakeholders
Peer review	Report.	200 questions; App. E, plus 10 pass/fail criteria; App.G	Other professional modeller.[2]
Audit	Report, model data files, simulations, and output.	200 questions; App. E, plus 10 pass/fail criteria; App.G	Other professional modeller.[2]
Post-audit	Report, model data files, simulations, and re-runs with actual stresses.	200 questions; App. E, plus 10 pass/fail criteria; App.G	Professional modeller.[2]

1. The listed appendices are in Middlemis (2001).
2. Attributes of suitable experienced model reviewers are summarized in Item 11 of Appendix C of the AUS guide

5.2 UK

The Environment Agency's Guidance Notes distill the practical experience of more than 20 modellers who have worked on groundwater resources projects over the last 30 years. Each of the 31 two or three page topics was written by an invited specialist and included modellers from consultancies, academics, and the regulator. Each topic is directly relevant to the Agency's operational use of regional groundwater modelling for water-resources management and the guidelines do not deal in any detail with source protection or contaminant transport modelling. Two appendices describe case studies that illustrate the integrated use of the various topics.

The UK guidelines use the terms *perceptual model* and *conceptual model*. Most guidelines refer to a conceptual model as some simplified understanding of the real system. However, in the various guidelines the term is used to cover both a qualitative and a quantitative understanding. In the UK these are distinguished. Beven (2002) points out that much more complexity is recognized than can be represented in a mathematical model. He refers to what we know about a system as the *perceptual model* and the mathematical representation as the *conceptual model.* The Environment Agency in the UK makes the same distinction between the qualitative understanding of a system – the model in your head – and the quantitative description of that understanding in the conceptual model (Hulme *et al.*, 2003). This leads to a conceptual model that can be properly tested because it is described using numbers.

The UK guidelines are unique in their emphasis on post-project appraisal as an essential separate stage. The guidelines were written after a comprehensive review of all the Agency's time-variant regional models. The Agency has reviewed three recent models in adjacent chalk catchments and this has promoted debate on the scientific issues raised, for example, the estimation of time series of recharge (Environment Agency, 2004).

The guidelines include a Template Project Brief which provides an example specification of the purpose, approach, and outputs for each major task in a groundwater modelling study. These are presented not as a strict procedure but as a resource to be adapted accordingly.

5.3 FH-DGG

The FH-DGG guideline is based on the idea of a "Hydrogeological Model" (HGM), which aims to provide a consistent framework for the transfer of complex hydrogeological nature into a model. Its main focus is on the creation of a hydrogeological model concept that simplifies nature adequately with respect to the problem to be solved and the respective dominant hydrogeologic features. As the HGM is intended to enhance understanding of hydrogeological systems and the predictions of their behavior as well as to serve as a basis for analytical or numerical calculations, the calibration, evaluation, and possible re-iteration of the hydrogeological model concept prior to application is included. Thus, in the FH-DGG guideline, the terminus “Hydrogeological Model” is only chosen if the hydrogeological model concept is proofed by a sound evaluation consisting of an analysis of model results with respect to uniqueness, accuracy, sensitivity, and model application range as well as a quantification of uncertainty. That means the HGM is ready to use.

The FH-DGG guideline is not considered to be a strict recipe on modelling, but rather to provide a systematic framework for the generation of hydrogeological model concepts as well as for practical issues like problem specification, commissioning of work, efficient work flow, and structural quality assurance. The intention of the Guideline is to assist clients, consultants, and regulatory officers in groundwater resources in the assessment of the database, the choice of an adequate model approach based on the spatial scale and the data situation, as well as possible necessary revisions of the model approach.

The guideline also proposes a working and communication scheme throughout all fundamental steps in the construction and application of the HGM. By following this procedure, an efficient approach to model development should be guaranteed, irrespective of the individual hydrogeological situation and the posed problem. Particular emphasis is put on communication at specific milestones, where for quality assurance purposes a common work base must be formed jointly by clients, consultants, and regulatory officers. Thus, unnecessary iterations are avoided and necessary iterations emanating from an inadequate model approach are clearly identified.

5.4 USNRC

A comprehensive strategy for hydrogeologic modelling and uncertainty analysis has been developed by a U.S. Nuclear Regulatory Commission (USNRC) research contractor (Neuman & Wierenga, 2003). The strategy recommends that alternative models be evaluated using maximum-likelihood Bayesian averaging, and suggests rating different models using Kashyap’s criterion. This method includes the variance-covariance matrix on the parameters in model discrimination – models with parameters that are more precisely estimated are preferred, given a similar match to observations. The strategy uses Kashyap’s criterion to weight predictions from the different models to obtain a probability distribution for predictions that accounts for model structure uncertainty. The same weighting procedure can be used with other criteria, such as the AIC and BIC statistics (defined in most statistics textbooks and in Hill (1998). USNRC staff are evaluating this strategy for use in reviewing performance assessments of nuclear facilities and sites.

5.5 USGS

The USGS guidelines are unique in the statistics presented for sensitivity analysis. These statistics can be used to measure the information provided by observations for parameters and predictions, and the importance of parameters to predictions. These statistics provide an effective way to improve the utility of the model for resource managers. For example, the model can be used to clearly indicate the value of additional field data and whether new data justify recalibration of a model.

The new statistics developed as part of the USGS guidelines are independent of model fit, which makes them useful even if the model is not yet calibrated. They also are well suited to evaluation of potential new data, for which the observation is not yet known.

The USGS guidelines provide a unique perspective on weighting of observations, prior information, and regularization. Using statistical theory, the guidelines discuss the importance of using weights to account for observation error in the model development effort. The weighting approach suggested provides a systematic way to include different kinds of observations in a single objective function, or in multiple objective functions.

5.6 New Zealand

The New Zealand groundwater model audit guidelines reflect a need to provide non-experts with tools to audit and review groundwater models. The guidelines include explanations of how models work and which models suit specific problems. Procedures with accompanying checklists of pertinent questions are listed such that auditors of models may determine whether there are any modelling errors, whether the results are meaningful in the context of the particular question being asked of the model, and whether model uncertainty is a result of parameter variability and measurement errors, or model assumptions.

6. CONFLICTING IDEAS AND PROCEDURES

The following are active known disagreements reflected directly in the guidelines considered here, or simmering just beneath the surface within the broader community of groundwater modellers. A brief description of the positions is stated, and proponents are noted for positions expressed in sets of guidelines presented in this paper. Positions without a proponent listed are not expressed in the sets of guidelines considered here but are included so that opinions from elsewhere in the groundwater modelling community are expressed.

6.1 Proper role of hydraulic-conductivity measurements

Position 1: Hydraulic-conductivity measurements based on laboratory tests of field samples, slug tests, aquifer tests, and so on are not relevant enough to large-scale models to be used in large-scale ground-water models of a significantly larger scale to support many defined hydraulic conductivity parameters. The difference in scale also makes the hydraulic-conductivity measurements inconsistent with being used in this way. (USGS)

Position 2: It depends on the scale. Small-scale hydraulic conductivity measurements can be useful for macroscale structured models. Position 2 is true for regional models, where the effective parameters are found between the harmonic and the arithmetic mean of the local measurements depending on the anisotropy and the direction of flow with respect to the anisotropy axes. (FH-DGG)

Position 3: Hydraulic-conductivity measurements based on laboratory tests of field samples, slug tests, aquifer tests, and so on can be used in large-scale ground-water models to support many defined hydraulic-conductivity parameters. The differences in scale between the measurements and the model do not cause problems that make the relevance of the measurements questionable.

6.2 Accounting for observation errors

Position 1: Use weighting to account for random errors in the observations. This also provides a way to normalize observations that may have different units, and therefore cannot be accumulated directly into a single objective function. (USGS)

Position 2: Weights cannot be determined well enough with available data to be useful.

6.3 Proper role for pumping test head-change data in calibrating regional models

Position 1: All models, regional or local scale, should be calibrated to pumping test head-change data (for example, drawdown data).

Greenfields sites (where there is only limited and/or short term data available), commonly have data on the short term pumping of test boreholes, and the measurement of aquifer responses in observation boreholes. This position holds that even short term pumping test head-change data is suitable for model calibration, and the implication is that calibration to such data renders a model valid for regional scale simulations and/or impact assessment purposes. The corollary (ie. if Position 2 of section 6.3 prevails over Position 1), raises the question of to what extent is the common approach of undertaking these field investigations justified in the short or long term in relation to the substantial expense involved and the argument of Position 2 that the short term data is of limited value.

Position 2: Regional models should not be expected to be accurately calibrated to pumping test data, but should be (eventually) well-calibrated to large scale stresses and long term monitoring. (AUS, FH-DGG) A regional model is designed for regional scale investigations, and its setup typically involves non-homogeneous (and sometimes non-isotropic) conditions, as well as boundary inflows/outflows, recharge, and stream-aquifer interaction features that impose regional- and local-scale gradients and curvature on the water table. However, pumping tests are usually analyzed with a range of assumptions that include homogeneity, isotropy, infinite and/or fully penetrating boundaries, and other assumptions that are not consistent with the regional model setup. A short term and/or local scale pumping test does not stress the aquifer adequately to invoke regional-scale aquifer responses, which is what the regional-scale model is designed to investigate. Therefore, a regional model should not be expected to accurately reproduce local-scale changes in head in response to pumpge, and also be expected to be suitable for its prime purpose of regional-scale investigations. Monitoring data from large- scale and long-term pumping schemes (ie. flows, water levels, water quality, etc), however, is highly valued for calibrating models. The calibrated model should, however, be developed with parameter values that are consistent with the values obtained from any pumping tests, to help address model non-uniqueness issues.

6.4 Use of contoured head data

Position 1: While heads form a quantitative calibration target, subjective assessment of the goodness of fit between model and measured groundwater level contours also is important. (AUS)

Position 2: Trying to match head contours that do not reflect conservation of mass considerations generally is not helpful. (USGS)

Position 3: Contouring head data helps to interpret local hydrogeologic conditions, including definition of parameter distributions and boundaries. Yet as interpolation has no physical background for calibration, only measured point values should be used. (FH-DGG)

6.5 Establish specific goals for model performance measures

An example of such a goal is that the largest discrepancy between observed and simulated heads needs to be less than a specified amount.

Position 1: Set goals for model performance measures at the scoping stage of model development (AUS, A&W). Propose staged development with coarse initial targets to be met before invoking more accurate targets with each successive stage of refinement as understanding improves.

Position 2: It is unclear how to establish such goals and how relevant they are. The resource manager would probably be able to best suggest goals based on tolerable prediction uncertainty, but translating that into goals applicable to model calibration is not straightforward, and perhaps not even possible. (USGS)

6.6 Use optimization methods to estimate parameter values.

Position 1: Using optimization methods to estimate parameter values helps to enhance understanding of the system. Either gradient or global search methods can be used. This does not indicate that only one model is to be produced, only that intense investigation of one of several models can be informative. (USGS, US-NRC, ASTM)

Position 2: The use of optimization methods is misguided because it emphasizes a single model. It is better to generate random samples of the possible sets of parameter values, use the related simulations to calculate weighted least-squares objective functions, and use dotty plots to investigate the results. Models that match the data too poorly are eliminated from the analysis. (GLUE, UK)

Position 3: Optimization methods are useful in calibration to estimate selected parameter values. However, as the model results depend on other influences (aquifer geometry, boundaries, initial states, etc.), a proper evaluation is needed to quantify the uncertainties of the assumptions and their influence on the results. (FH-DGG)

7. CONCLUSIONS

The many sets of guidelines developed for groundwater models reflects the importance of groundwater to people, communities, nations, and the world. Working together will help groundwater modelling reliability improve. This paper introduces the conflicting ideas and procedures so that the community can begin to work together to understand and possibly resolve differences. Development of computer programs that make it easy to experiment with different approaches will facilitate joint experiments that will help to mature groundwater modelling.

REFERENCES

Anderson, M. P. & Woessner, W. W. (1992) Applied Groundwater Modelling. Academic Press, San Diego.

Arbeitskreis "Hydrogeologische Modelle" der Fachsektion Hydrogeologie (1999) Hydrogeologische Modelle – Ein Leitfaden für Auftraggeber, Ingenieurbüros und Fachbehörden. *Schriftenreihe der Deutschen Geologischen Gesellschaft; Heft 10;* English translation by Johannes Riegger, „Hydrogeological models -- Guidelines for clients, consultants, and regulatory officers", http://www.fh-dgg/ak-hgm, 2004.

Arbeitskreis "Hydrogeologische Modelle und Grundwassermanagement" der Fachsektion Hydrogeologie (2002) Hydrogeologische Modell – Ein Leitfaden mit Fallbeispielen. *Schriftenreihe der Deutschen Geologischen Gesellschaft; Heft 24.*

Beven, K. J. (2001) Rainfall-Runoff modelling, the primer. West Sussex, England, John Wiley.

Beven, K. (2002) Towards a coherent philosophy for modelling the environment. *Proceedings Royal Society of London, A*. **458**, 2465-2484.

Beven, K. J. (2004) A manifesto for the equifinality thesis. Unpublished manuscript.

Environment Agency (2002) Eds. Hulme, P. *et al*. Groundwater Resources Modelling: Guidance Notes & Template Project Brief. R&D Technical Report W213. *Environment Agency (England & Wales*).

Environment Agency (2004) A comparison of Chalk Groundwater Models in and around the River Test catchment. *Environment Agency Science Project NC/03/05, Environment Agency (England & Wales)*

Freeze, R. A., Massman, J., Smith, L., Sperling, T. & James, B. (1992) Hydrogeological decision analysis, 4, The concept of data worth and its use in the development of site investigative strategies. *Groundwater* **30**(4), 574-588.

Hill, M. C. (1998) Methods and guidelines for effective model calibration. *US Geological Survey Water-Resources Investigations Report* 98-4005. http://pubs.water.usgs.gov/wri984005/

Hill, M. C., Ely, M. D., Tiedeman, C. R., D'Agnese, F. A. Faunt, C. C. & O'Brien, B. A. (2001) Preliminary evaluation of the importance of existing hydraulic-head observation locations to advective-transport predictions, Death Valley regional flow system, California and Nevada. *US Geological Survey Water-Resources Investigations Report* 00-4282, http://water.usgs.gov/pubs/wri/wri004282/.

Hulme, P., Johnson, D., Fletcher, S. (2003) Challenging our misunderstandings through modelling - lessons learned from the Environment Agency national groundwater modelling programme. *In MODFLOW 2003 - Proceedings, International Groundwater Modelling Center (IGWMC), September 16-19, 2003*.

Middlemis, H. (2001) Groundwater flow modelling guideline. *Murray-Darling Basin Commission, Perth, Australia, Project No. 125.*

Pattle Delamore Partners (2002) Groundwater Model Audit Guidelines. *Report submitted to New Zealand Ministry for the Environment, 73p plus 4 Appendices.*

Neuman, S. P. & Wierenga, P. J. (2003) A comprehensive strategy of hydrogeologic modelling and uncertainty analysis for nuclear facilities and sites. *NU-REG/CR-6805, US Nuclear Regulatory Commission, Washington, D.C.*

Reilly, T. E. & Harbaugh, A. W. (2004) Guidelines for evaluating ground-water flow models. *USGS Scientific Investigations Report 2004-5038.*

Riegger, J. (2004) Hydrogeologic models - A guideline for clients, consultants and regulatory officers. Unpublished manuscript.

Tiedeman, C. R., Hill, M. C.**,** D'Agnese, F. A. & Faunt, C. C. (2003) Methods for using groundwater model predictions to guide hydrogeologic data collection, with application to the Death Valley regional ground-water flow system. *Water Resources Research* **39**(1), 5-1 to 5-17, 10.1029/2001WR001255.

Vrugt, J. A., Gupta, H. V., Bastidas, L. A., Bouten, W. & Sorooshian, S. (2003) Effective and efficient algorithm for multiobjective optimization of hydrologic models. *Water Resoures Research* **39**(8), 1214, doi:10.1029/2002WR001746.