MODEL ABSTRACTION IN HYDROLOGIC MODELING

Yakov Pachepsky, Soil Scientist, USDA-ARS, ypachepsky@anri.barc.usda.gov; Andrey Guber, Postdoctoral Scholar, University of California Riverside, aguber@anri.barc.usda.gov; Rien van Genuchten, USDA-ARS, rvang@ussl.ars.usda.gov; Thomas Nicholson, Senior Technical Advisor, US NRC, tjn@nrc.gov; Ralph Cady, Senior Performance Assessment Analyst, US NRC, rec2@nrc.gov; Jirka Šimunek, Professor and Hydrologist, University of California Riverside, jiri.simunek@ucr.edu; Timothy Gish, Soil Scientist, USDA-ARS, tgish@hydrolab@arsusda.gov; Diederik Jacques, Scientific Staff Member, SCK·CEN, djacques@sckcen.be; Craig Daughtry, Research Agronomist, USDA-ARS, cdaughtry@hydrolab.arsusda.gov

Abstract
Model abstraction (MA) is a methodology for reducing the complexity of a simulation model while maintaining the validity of the simulation results with respect to the question that the simulation is being used to address. The MA explicitly deals with uncertainties in model structure and in model parameter sources. It has been researched in various knowledge fields that actively use modeling. We present (a) the taxonomy of model abstraction techniques being applied in subsurface hydrologic modeling, (b) the systematic and comprehensive procedure of the MA implementation including (1) defining the context of the modeling problem, (2) defining the need for the model abstraction, (3) selecting applicable MA techniques, (4) identifying MA directions that may give substantial gain, and (5) simplifying the base model in each direction. The need in MA may stem from (a) difficulties to obtain a reliable calibration of the base model, (b) the error propagation making the key outputs uncertain, (c) inexplicable results from the base model, (d) excessive resource requirements of the base model, (e) the intent to include the base model in a larger multimedia environmental model, (f) the request to make the modeling process more transparent and tractable, and (g) the need to justify the use a simple model when a complex model is available. The example illustrates the MA application in field-scale simulations of water flow in variably saturated soils and sediments. The MA (a) can result in the improved reliability of modeling results, (c) make the data use more efficient, (c) enable risk assessments to be run and analyzed with much quicker turnaround, with the potential for allowing further analyses of problem sensitivity and uncertainty, and (d) enhance communication as simplifications that result from appropriate model abstractions may make the description of the problem more easily relayed to and understandable by others, including decision-makers and the public.

INTRODUCTION

Model abstraction (MA) is defined as *methodology for reducing the complexity of a simulation model while maintaining the validity of the simulation results with respect to the question that the simulation is being used to address.* The need in MA addresses the concerns that the use of overly complex simulation models may cause an excessive burden of data collection and computations as well as difficulties in interpreting simulation results and conveying the simulation approach to both technical and lay audiences. The presumed risk of leaving some important process or feature out often leads the model users to employing fairly complex flow and transport models. However, many of the detailed features, events and processes represented in these complex models may have limited influence at a specific site.

The feasibility of MA has been demonstrated in many research and engineering fields that give ample examples of models having strikingly different complexity and yet the same accuracy, for example, in regional ground water assessments (Kelson et al., 2002), chemical engineering (Diwekar, 1994), marine ecology (Stillman et al., 2000), runoff generation (Jakeman and Hornberger, 1993), population dynamics (Stephens et al., 2002), demographic projections (Smith, 1997), battlefield simulations (Sisti and Farr, 2005), etc. In contaminant hydrology, MA is possible because the complexity of flow and transport pathways at a specific site is easy to perceive but difficult to represent in mathematical equations of the model without making strong simplifying assumptions. Different sets of plausible assumptions lead to different models that are consistent with the available observations. The multiplicity of models reflects the multiplicity of possible conceptual approaches to representation of complex subsurface processes in mathematical form tractable within limitations of existing computer and measurement technologies (Neuman and Wierenga, 2003).

The objectives of this work were (a) to develop a systematic overview of MA techniques being applied in subsurface hydrologic modeling, (b) to suggest the procedure of the MA implementation, and (c) to develop a case study of MA application.

MODEL ABSTRACTION TECHNIQUES

The methodology of MA has been developing for more than 30 years in various research and engineering fields (Meisel and Collins, 1973; Zeigler, 1976; Innis and Rextad, 1983; Fishwick, 1995; Frantz, 1995; Caughlin and Sisti, 1997; Davis and Bigelow, 2003; Van Ness and Scheffer, 2005). Most of the MA techniques are specific to the type of mathematical models used in a specific field. The synopsis and annotated examples of MA in modern flow and transport modeling are presented in (Pachepsky et al., 2005). These MA techniques are summarized in Fig. 1. Two main targets of abstraction are (1) the model structure, i.e. the formal description of specific processes and their interactions that affect flow and transport variables, and (2) the parameter determination, i.e. the estimation of constant and functions serving as coefficients in model equations.

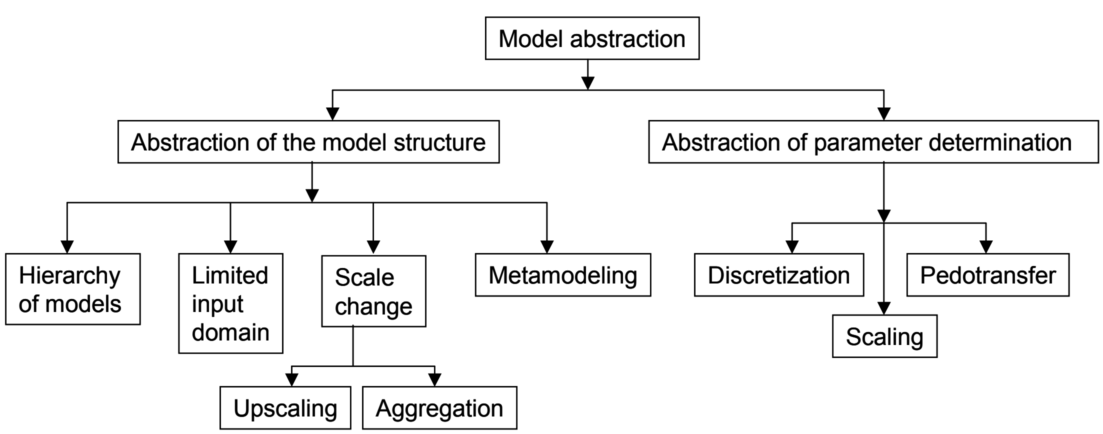

Figure 1. Categories of model abstraction techniques relevant to flow and transport modeling in subsurface hydrology.

Structure of hydrologic models is changed with MA via (a) using pre-defined hierarchies of models, (b) delimiting input domain, (c) scale change done by either upscaling or aggregation, and (d) metamodeling. A predefined hierarchy contains a series of progressively more simple conceptual and corresponding mathematical representations of porous media. The class of MA techniques based on the delimiting input domain utilises the fact that some features, events, or processes may be not relevant for a given set of scenarios or for a given set of model outputs. Scale change provides transitions between four operational scales – core, profile/pedon, field, and watershed scales that are of interest in contaminant hydrology. MA with scale change alters model equations, variables and parameters with two classes of methods: upscaling and aggregation. Upscaling MA methods use the fine-scale model and the fine scale media properties to derive the coarse-scale model equations and to relate the coarse-scale and fine-scale transport parameters. On the contrary, MA with aggregation does not assume any relationship between model parameters and equations at the fine and at the coarse scales. Parameters of the coarse-scale model are lumped, and are subject to calibration with field data. Aggregation is also being done without the change in the model equations by combining several soil horizons or geologic strata. Metamodeling seeks to simulate the input-output relationships of the complex model with a simple statistical relationship or, recently, with more complex data mining tools such as artificial neural networks.

MA applied to parameterization of hydrologic models does not provide a substitution for the model calibration, but rather seeks to obtain reasonable estimates of parameter values and their variability. Such estimates are useful for setting initial values of parameters for the model calibration, using non-calibrated model in pilot studies and field campaign designs, assigning values of parameters that are shown to be not sensitive, etc. Abstraction of parameter determination affects parameter estimation via (a) discretization, (b) scaling, and (c) pedotransfer functions. The discretization abstraction techniques replace the continuously varying spatial fields of parameters with piece-wise constant spatial distributions. The scaling abstraction techniques are useful when models are coarsened, i.e. the grid cell size is substantially increased but no changes in the model structure are made. The scale-dependence in model parameters is usually encountered in such cases, and scaling relationships are needed to convert the original parameter values and measurement data to the parameters values of the coarsened model. The pedotransfer techniques convert the readily available data to the hydraulic and transport parameters of the unsaturated flow and transport. Those parameters are notoriously difficult to measure, and a substantial effort has been made to estimate these parameters from the data available from soil survey or borehole logs. The empirical functions used for such estimating are often called pedotransfer functions. Large databases have been assembled to encompass variety of soil properties in developing pedotransfer functions in different parts of the world. New powerful heuristic tools, such as artificial neural networks and regression trees, appeared to be useful in the development of pedotransfer functions. Still, the accuracy of pedotransfer functions outside of their development dataset remains essentially unknown, and the use of ensembles of pedotransfer functions is strongly recommended.

MODEL ABSTRACTION IMPLEMENTATION

The MA implementation has to conform requirements of objectiveness, systematic implementation, comprehensiveness, and efficiency (Neuman and Wierenga, 2003). The MA process starts with an existing *base* model that can be calibrated and used in simulations. The *key output* of the model is defined that provides the necessary and sufficient information to decide on issues of interest. The base model may need abstraction for one or more of the following reasons (a) the base model includes a complex description of processes that cannot be observed well and yet need to be calibrated; the calibrated values of parameters of those processes are very uncertain, (b) the base model propagates uncertainty in the initial distributions, parameters, and forcing in a manner that creates an unacceptable uncertainty of the key output, (c) the base model produces inexplicable results in terms of the key output, (d) the base model requires an unacceptable amount of resources for computations, data preprocessing, or data post-processing, e.g. the base model is not suitable to be used as a part of a real-time modeling system that requires short computer runtimes, (e) the base model lacks transparency to be explicable and believable to the users of the key output.

The MA process includes the following steps (1) justify the need for the MA, (2) review the context of the modeling problem, (3) select applicable model techniques, (4) determine the MA directions that may give substantial gain, (5) simplify the base model in each direction. Statistical criteria based on guidelines by Hill (1998) can be used to justify the need in MA in case it is related to the uncertainty in calibrated parameter values or in the model key output. The context of the modeling problem has to be reviewed to realize what details and features of the problem are omitted or de-emphasized when the abstraction is performed, and thus to warrant the comprehensiveness and objectiveness of the MA process. MA can lead to simplifications via (a) the number of processes being considered explicitly, (b) process descriptions, (c) coarsening spatial and temporal support, (d) the number of measurements for the reliable parameter estimation, (e) reduced computational burden, (f) data pre-processing and post-processing. Detailed description of the suggested MA process can be found in Pachepsky et al. (2005) where classes of MA techniques are specified that lead to each type of the simplification. Each abstracted model has to be parameterized and confirmed in the uncertainty context.

MODEL ABSTRACTION CASE STUDY

The MA case study was design to illustrate the MA process and techniques. The main objective of the test case was to demonstrate how the MA can be applied to understanding and prediction of soil water fluxes at a relatively humid site where transport may be affected by the presence of soil macropores and related preferential flow phenomena. The test site was located in Bekkevoort, Belgium (Jacques, 2000). The 1.5-m deep test trench was dug in a loamy soil and was instrumented to measure soil water content and soil matric potential in 60 locations at five depths hourly, soil temperature in 6 locations at five depths hourly, and soil water fluxes in three locations at two depths once in one to four days. Vegetation was removed and a thin layer of fine gravel covered the soil. Soil water monitoring continued for 384 days. Soil was sampled at 90 locations at three depths to measure soil water retention in laboratory.

The base model was the Richards equation of water flow in variably saturated porous media. The model was calibrated using almost 200,000 measurements of soil water content. The key output of the model was the total soil water flux at two measurement depths and at the bottom of the soil profile over three wetting-drying periods.

The need in MA was justified by the fact that the calibrated base model predicted substantial runoff whereas no runoff was observed at the site during the monitoring period. The base model produced inexplicable results in terms of the key output (Jacques et al, 2002).

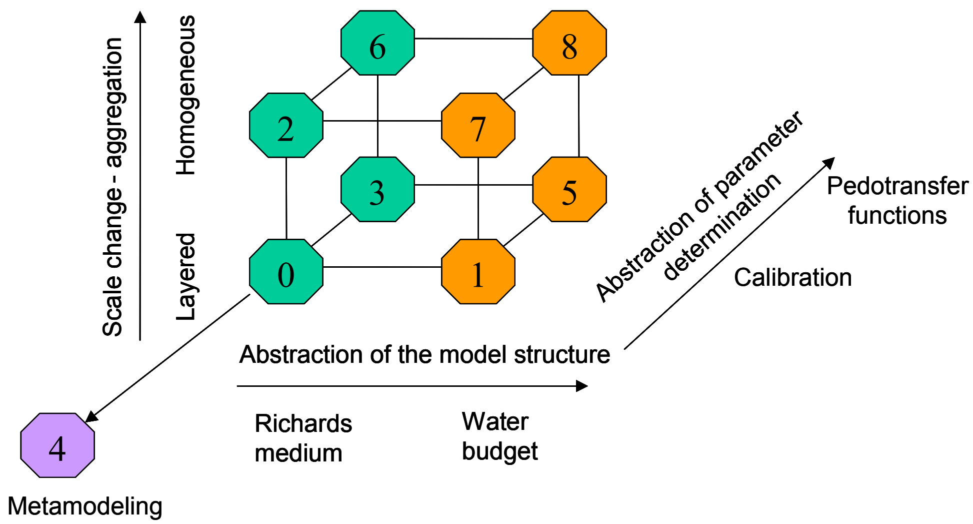

Figure 2. Design of the model abstraction application. The Richards media was simulated with the HYDRUS-1D software (Simunek et al., 1999); the water budget model was MWBUS (Pachepsky et al., 2005). The model numbering is explained in the text.

Four classes of applicable model abstraction techniques were selected as shown in Fig. 2. The base model 0 was abstracted. Model structure abstraction was achieved using the hierarchy of model (models 1,7,5, 8 in Fig. 2), aggregation (models 2, 6, 7, and 8), and metamodeling (model 4). The hierarchy of models included the Richards equation and a simple soil water budget model. Aggregation was done by replacing the layered soil profile with the homogeneous profile. Metamodeling was done with the backpropagation artificial neural network that used daily and weekly precipitation as the input to estimate monthly cumulative soil water fluxes. An ensemble of 22 pedotransfer functions was used in model abstraction of parameter determination. HYDRUS1-D and MWBUS software packages were used to calibrate the abstracted models and to run Monte Carlo simulations to evaluate the uncertainty in calibrated model outputs.

The abstraction with the hierarchy of models was useful. The simple soil water budget model was less accurate in predictions of soil water content compared with the Richards model. However, unlike the more complex Richards model, this simple soil water budget model correctly predicted the absence of runoff and measured cumulative soil water fluxes (Fig. 3). The prediction of runoff was an artifact of the Richards model calibration in absence of measured boundary fluxes. This abstracted model appeared to be instrumental in both explaining behavior of the complex model and in predicting the key output – soil water fluxes.

The abstraction with aggregation was not useful in this case. The Richards model was less accurate with respect to soil water contents and continued to generate large simulated runoff when a homogeneous soil layer was introduced.

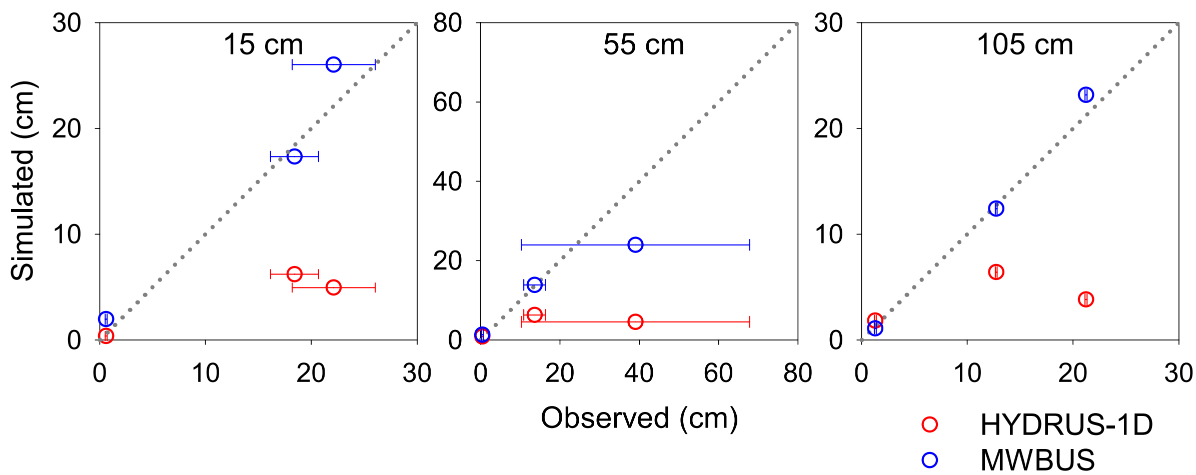

Figure 3. Simulated and observed cumulative fluxes over three wetting-drying periods at depths of 15 cm, 55 cm, and 105 cm. Error bars show standard deviations. Both HYDRUS-1D and MWBUS have been calibrated.

The abstraction of parameter determination with pedotransfer functions was useful. It showed that the Richards model with parameters in correct ranges is able to correctly simulate soil water fluxes. The ensemble of pedotransfer functions represented field water retention better than data from laboratory soil water retention measurements. Soil water content predictions with the Richards model also were more accurate with the ensemble of pedotransfer functions than with laboratory water retention data.

The neural network metamodel was extremely accurate in estimating cumulative soil water fluxes. It was also five orders of magnitude faster than the numerical algorithm coded in HYDRUS-1D.

In general, the MA process was successful in this case study. MA explained the strange behavior of the complex model, and provided the correct description of the system behavior and plausible parameter ranges. The case has also demonstrated the need in model calibration and uncertainty analysis toolboxes and appropriate user interfaces to perform the MA in the uncertainty context.

Overall, the intensive use of models in subsurface contaminant hydrology has resulted in development of many MA techniques that to-date have been used mostly in research. Potential benefits of MA include improvement of understanding and communication of modeling results, more robust predictions, and better understanding of essential factors and their representation in models. This makes MA an attractive methodology for engineering modeling applications. The

MA process can be set as a transparent step-by-step formalized procedure of justification of the use of a simplified model. An important feature of models abstraction is the explicit treatment of model structure uncertainty. The model structure, along with the data uncertainty, and scenario uncertainty, is known to introduce the uncertainty in modeling results. Unlike the uncertainty in input data, in model parameters, and in scenarios, the effect of the model structure uncertainty on the uncertainty in simulation results is usually impossible to quantify in statistical terms. Using MA, a series of models with feasible structures can be built and evaluated in a systematic manner. Each of the models is evaluated from results of an ensemble of simulations by its accuracy to measurement data and by its predictions with respect to scenarios that have not been observed.

REFERENCES

Caughlin, D. and Sisti, A.F. (1997). "A Summary of Model Abstraction Techniques", Proceedings of Enabling Technology for Simulation Science I Conference, (Orlando FL., 22-24 April 97). SPIE Conf. 3083, pp 2-13.

Davis, P. K., and Bigelow, J. H. (2003). Motivated metamodels: synthesis of cause-effect reasoning and statistical metamodeling. MR-1570. RAND, 2138, Santa Monica, CA.

Diwekar, U. M. (1994). "How Simple Can It Be? - A Look at The Models for Batch Distillation," Computers & Chemical Engineering 18: S451—S457.

Fishwick, P. A. (1995). "*Simulation Model Design and Execution*". Englewood Cliffs: Prentice-Hall, 1995

Frantz, K. F.(1995). "A taxonomy of model abstraction techniques," In: Alexopoulos, C., Kang, K., Lilegdon, W. R., and Goldsman, D. (Eds.) Proceedings of the 1995 Winter Simulation Conference, pp.1413-142.

Hill, C. M. (1998). Methods and Guidelines for Effective Model Calibration. U. S. Geological Survey Water Resources Investigations Report 98-4005, Denver, CO, 90 pp.,

Innis, G., and E. Rextad (1983). "Simulation model simplification techniques," Simulation, 41: 7-15.

Jacques D. (2000). Analysis of Water Flow and Solute Transport at The Field Scale. PhD Thesis no 454. Faculteit Land bouwkundige en Toegepaste Biologische Wetenschappen, K.U. Leuven, Belgium, 255 pp.

Jacques, D., Simunek, J., Timmerman, A., and J. Feyen (2002) "Calibration of Richards and Convection–Dispersion Equations To Field-Scale Water Flow And Solute Transport Under Rainfall Conditions," Journal of Hydrology, 259: 15-31.

Jakeman, A.J. and G.M. Hornberger (1993). "How Much Complexity is Warranted in a Rainfall-Runoff Model?" Water Resources Research, 29: 2637-2649,

Kelson, V.A., Hunt, R.J., and H.M. Haitjema (2002). "Improving a Regional Model Using Reduced Complexity and Parameter Estimation," Ground Water, 40: 132-143.

Meisel, W. S. and D. C. Collins (1973). " Repromodeling: An approach to efficient model utilization and interpretation," IEEE Transactions on Systems, Man, and Cybernetics SMC-3: 349-358,

Neuman, S. P., Wierenga, P. J., and Nicholson, T. J. 2003. A comprehensive strategy of hydrogeologic modeling and uncertainty analysis for nuclear facilities and sites. NUREG/CR 6805. U. S. Nuclear Regulatory Commission. Washington, D.C. 20555-0001.

Pachepsky, Y. A., Guber, A. K., Van Genuchten, M. T., Nicholson, T. J., Cady, R. E., Simunek, J., Schaap, M. C. (2005). Model Abstraction Techniques for Soil Water Flow and Transport. NUREG/CR-6884. U.S. Nuclear Regulatory Commission, Washington, D.C.

Simunek, J., Sejna, M., and van Genuchten, M. Th. (1998). The HYDRUS-1D software package for simulating the one-dimensional movement of water, heat, and multiple solutes in variably-saturated media. Version 2.0. IGWMC - TPS - 70, International Ground Water Modeling Center, Colorado School of Mines, Golden, Colorado, 186 pp.

Sisty, A.F. and S.D. Farr. (2005). Model Abstraction Techniques: An Intuitive Overview. Air Force Research Laboratory/IFSB, Rome, New York, http://www.rl.af.mil/tech/papers/ModSim/ModAb_Intuitive.html.

Smith, S.K. (1997). "Further Thoughts on Simplicity and Complexity in Population Projection Models," International Journal of Forecasting, 13: 557 – 565.

Stephens, P. A., Frey-Roos, F. Arnold, W., and Sutherland, W. J. (2002.). "Model Complexity and Population Predictions. The Alpine Marmot As A Case Study," Journal of Animal Ecology, 71: 343–361.

Stillman R. A. , McGrorty, S., Goss-Custard, J. D. and A. D. West. (2000). "Predicting Mussel Population Density and Age Structure: The Relationship Between Model Complexity and Predictive Power," Marine Ecology Progress Series, 208: 131–145.

Van Ness, E. H., Scheffer, M. (2005). "A strategy to improve the contribution of complex simulation models to ecological theory", Ecological Modelling, 185: 153-164..

Zeigler, B. (1976). Theory of modeling and simulation. New York, New York, Wiley and Sons.