SELECTING AND EVALUATING SUBSURFACE FLOW AND CONTAMINANT TRANSPORT MODELS: ISSUES AND PROBLEMS

by

Marios Sophocleous

Kansas Geological Survey
Open-File Report 88-35

Lawrence, KS, October 1988

Kansas Geological Survey
Open-file Report

Disclaimer

SELECTING AND EVALUATING SUBSURFACE FLOW AND CONTAMINANT TRANSPORT MODELS: ISSUES AND PROBLEMS

Marios Sophocleous

Kansas Geological Survey

Invited presentation at the 31st Annual Meeting of the Association of Engineering Geologists in Kansas City, Missouri, October 16-21, 1988

Mathematical models and computers are by now standard tools of most scientists, engineers and planners engaged in water resource activities.

The foundation of a model is the conceptual model that describes the processes to be modeled. A conceptual model is the technical expert's preliminary perception of the physical behavior of a groundwater flow and transport system based on available data, experience and basic hydrologic principles. Development of the conceptual model requires a clear understanding of the study objectives and compilation of available field data. A conceptual model for unsaturated zone transport, for example, is shown in this slide, which gives a summary of the hydrogeochemical processes controlling pollutant transport in a waste disposal site.

Water and hazardous waste input are shown as precipitation and waste burial respectively. Water input as precipitation can either infiltrate the soil surface or it can be lost from the disposal site as runoff. The infiltrated water is stored in the soil profile and therefore subject to redistribution. If this water is redistributed upwards and evaporated or removed by shallow plant roots, then it is likely to play no role in the transport of hazardous substances. Water redistributed into the waste zone is water that generates the trench leachate, as indicated in the slide, and any water moving through the waste zone is likely to contain dissolved hazardous substances. The amount of hazardous chemicals leaving the waste trench via the water pathway is determined by the flow rate of water into and out of the waste zone, and the concentration of hazardous substances in the leachate, which in turn is determined by the solubility and leach rate controls. The two-primary paths taken by stored water in the soil profile are drainage and plant water uptake. If the water being used by the plants has passed through the waste zone then the plants will become contaminated and bring hazardous substances, such as radionuclides, to the surface, as shown by the dotted line in the slide. Drainage water that has passed through the waste zone will contain hazardous chemicals initially at the waste leachate concentration. However, there are at least three processes occuring that can reduce hazardous chemical concentrations before drainage water reaches groundwater. These three processes are a decay process, such as radioactive decay or biodegradation, a retardation process, such as sorption/precipitation, and a dilution process, such as hydrodynamic dispersion.

This slide clearly shows that radionuclide or any other hazardous substance transported along the water pathway is the combined or integrated result of many processes. A prediction or environmental assessment of hazardous waste transport based solely on information about one process, such as a drainage rate, or a sorption coefficient, is rarely meaningful. Such conceptual model provides a common basis for most transport models being used to describe water and pollutant transport.

The next level of a model is the mathematical model. This is the set of equations used to quantify the processes identified by the conceptual model. After the equations to be solved are developed, methods of using the computer to solve them are needed. These solution techniques are called numerical algorithms, and form the next level of the model. The numerical algorithms chosen for a model determine how accurately and efficiently the computer will solve the transport equations. The final level at which the model may be examined is the computer code. This is the instruction set that is prepared to implement the algorithms on a specific computer or in a specific computer language.

Water resource and waste disposal managers and engineers are confronted with a seemingly vast variety of computer codes which are potentially useful for performing a groundwater contaminant transport study. There are many publications, such as the one by van der Heijde and others of the Holcomb Research Institute, that provide an inventory of available codes. Such code inventories present a confusing array of possible choices. Consultants and managers, nevertheless, need to know which codes might be appropriately selected for application to their particular waste disposal or other problems.

A procedure for selecting the appropriate subsurface flow and transport computer codes for modeling a waste disposal site is based on nine ideal model-development steps shown in this slide, as outlined by Battelle scientists. These nine steps are:

1) Define the study objectives
2) Collect and analyze data characterizing the study site
3) Formulate a conceptual model of the site or study problem

4) Identify the appropriate process-descriptive equations

5) Select the appropriate analytical or numerical codes

6) Couple/interface the selected codes

7) Evaluate the selected code(s) performance

8) Run site-specific simulations, and

9) Compare simulation results with study objectives.

A conceptual model based on the site characterization data and consistent with study objectives is the hub of a model. Other model components are arranged as a wheel on that hub. Clockwise progress around the wheel, following the nine steps is required to complete the model. During the development of a simulation model, the hub may require repeated modifications and revisions in order to produce a well-rounded and balanced wheel.

I would put particular emphasis in this talk on steps 5 and 7, i.e., how to select computer codes and how to evaluate code performance. It is not easy to implement the nine model development steps I have mentioned. Successful code selection depends on other factors that are not apparent in those steps. In particular, much depends on the technical knowledge and experience of a user who is the modeling practitioner. By current groundwater transport modeling technology, codes are at best only tools in the hands of a user, and the presently available codes simply cannot transform a computer into a thinking machine capable of giving correct answers to every question about contaminant migration.

A difficulty in applying flow and transport models is that answers to specific problems will not necessarily be unique. It is quite possible for different users to arrive at dissimilar conclusions about a particular groundwater problem, while applying the same codes. This situation could be the consequence of holding different views on the hydrologic data and the system's conceptual model. So, to clarify the chosen modeling approach, a user should consider a number of technical issues, some of which are shown on this slide, and justify their resolution. The technical issues are simply questions as to what constitutes the correct way to describe the modeled system. However, there are no absolute or uniquely correct answers to these issues. I do not have time to go through these technical issues, so I will only mention one issue here.

-Boundary conditions do not automatically follow from field data and need to be established by a user. Yet, these peripheral problem constraints essentially determine a modeling problem's solution. Boundary conditions along a region's periphery are usually set as either known hydraulic head or water flux depending on a user's interpretation of the hydrology. That interpretation is part of the conceptual model that a user must be responsible for developing. Areal recharge sources are also important internal conditions that cannot be neglected. Usually the boundary flux is not actually known and can be manipulated to obtain nearly any desired result. But boundary flux must be made consistent with all other known hydrologic factors. In any case, ambiguous interpretation of boundary conditions can lead to uncertain modeling results.

To formulate and develop a proper model that can accurately simulate groundwater transport at a particular site, certain general transport modeling issues, in addition to the specific technical issues I mentioned, must also be taken into account. Again, due to time limitations, I will only enumerate just a few recognized general issues here, although in many cases these transport modeling issues do not have absolute answers. For example:

-What kinds of codes are needed?

-Should a simple or complicated code be used?

-How should selected codes be coupled?

-What is the appropriate modeling dimension?

-What modeling emphasis should be placed on each process?

-Should a steady-state or transient code be used?

-What numerical solution method is best?

-Is there a unique answer to a simulation problem?

-Is a deterministic or a stochastic approach best?

-What constitutes a tested model?

-What data requirements are imposed by the selection of codes and can these data requirements be reasonably met in the field?

Selection

Once the conceptual model is developed and the relevant equations have been identified, one can proceed with the code selection process. Selection of the appropriate analytical or numerical code suitable for analysis of a contamination problem at a specific site requires a thorough analysis of a variety of factors. Some of the factors to be considered are technical, such as numerical accuracy, efficiency, etc., and others are subjective, such as the experience a user may have with a particular code. If one views codes as "tools", then it is clear that an experienced user with a less elegant tool may perform better than a user with an unfamiliar tool-even if the unfamiliar tool is of better design.

Having identified all appropriate process-descriptive equations, or at least having identified the basic processes believed to be involved (that is the conceptual model), the kinds of codes required are nearly determined. A preliminary code selection starts by examining available code documentation and summary reports. The better the code documentation is, the more likely a user will be successful in choosing the right code. There is probably no such thing as a "best" code for all study purposes and objectives. The determination of adequate code performance rests with the evaluation process. A possible plan for code selection can be outlined as follows:

Selection Procedure

*Has the code been thoroughly tested and used by independent groups of scientists?

*Is an adequate summary description of the code available?

*Is it the most updated version of a family?

*Does it model the important processes?

*Does a user's manual exist?

*Does the code allow exclusion of irrelevant processes?

Discard the code when any answer is no.

Weigh the remaining codes according to their ability to handle:

*appropriate geometry and geologic structure

*appropriate mix of boundary conditions

*input of characterization parameter distributions.

*the extent of the code's application record & reputation

*the quality of the code user's manual and documentation

*familiarity and experience with the code by the users

Evaluation/Verification

Perhaps the most difficult part of the model selection plan is code evaluation. The purpose of this step is to confirm that selected codes will actually work as intended.

One model verification technique is to compare the output of a new code with the output of another code that solves the same equations using similar algorithms, a process known as benchmarking. As indicated in the slide, this verification procedure only tests the accuracy of the new code relative to the code being used for comparison. This technique is used to verify that the algorithms chosen have been coded properly and that no syntax or input-output errors exist in the new code.

The validity of the algorithms chosen can be tested by analyzing the computer solution in the three ways shown in the slide: that is 1) using analytical solutions for comparison; 2) analyzing for consistency with numerical theory; and 3) checking for mass balance between water input, output, and storage predictions.

Analytical solutions are solutions to the transport equations obtained using standard mathematical techniques. They represent the true solution to the mathematical model so they provide an ideal method of testing the accuracy of the numerical solution techniques. Unfortunately, analytical or so-called "closed form" solutions do not exist for many sets of equations. In fact, lack of analytical solutions for transport equations has motivated the development of computer techniques.

A second way to test algorithms is to use some of the theoretical concepts of numerical methods that are available. For example, water flow and mass transport models predict water content, water velocity, solute concentrations, etc., at discrete points in space and time. They do not provide continuous solutions as do analytical techniques. The specific points where solutions are provided are determined by the model user in the form of a "grid" of points. The number and location of the grid points can affect the solution obtained. The sensitivity of the computer solution to the grid is sometimes predictable from numerical theory. Other numerical theory is available that predicts the stability and convergence behavior of specific algorithms. Use of this theory to analyze the results of a model can be very helpful in determining the usefulness of the algorithms that have been chosen.

The third method listed in the slide for checking numerical algorithms is checking for mass balance. The mathematical model for pollutant transport reflects the principle of conservation of mass. In terms of model predictions this means that the proper relationship must exist between the predictions of water input, output, and water stored. If the algorithms chosen to solve the transport equations are doing an accurate job, they will preserve the mass balance of the system.

It is important to remember that verifying the accuracy of a particular algorithm only assures the model user that the computer technique used is solving the equations properly. It does not guarantee that the equations that make up the mathematical model, correctly describe the situation or processes being simulated. The only way to demonstrate that the conceptual and mathematical models chosen are adequate is to compare model predictions to experimental data sets.

Validation

The definitions and meanings of the terms verification and validation often hamper successful communication among both modelers and managers. Verification means to test a code's numerical implementation and accuracy, that is assuring that the computer program actually performs as designed. Validation calls for comparing code predictions with actual measurements of a particular system's behavior, that is assessing the model's accuracy in predicting real-world events. Typically, analytical solutions are used to verify a code, and site-specific

data sets are used to validate. To validate a computer model, it must first be calibrated against a real-world system. Calibration involves developing values for the constants and coefficients in the computer program from field data, in order to accurately predict real world events. The actual measurements used in model validation should be:
1) independent of those used in the calibration process, and 2) should reflect some change in the system, i.e., different stresses, boundary conditions, or time-varying information. When a simulation model is able to reproduce actual field observations, within acceptable limits of deviation, it is then called validated, at least under the specific conditions. A "double-blind" approach, in which the experiment and the modeling prediction are done independently, is preferrable to obtain an unbiased comparison.

Given that real field data are insufficient and sometimes of questionable accuracy, one may well ask: do the highly sophisticated models really represent the field situation? It appears that they provide insights to mechanisms and understanding of the systems, but most probably they do not provide very good predictions in general of future behavior. Sensitivity analysis, which quantifies a model's response to input parameter changes, gives insight into mechanisms and dependencies, but to make predictions you not only have to have something that is physically true and represented in your model, but you need good data also. Predictions are very difficult to make at the present time, especially for reactive contaminants, for site specific, small scale or local predictions. Models, however, can be effectively used in comparing several alternatives for management, rather than give an absolute answer. Models have also been proven successful for simulation and limited prediction of conservative, non-reactive, solutes in relatively large-scale settings away from the pollution source, where averaging effects have already resolved a number of difficult-to-simulate complications.

The most important attitude that a manager should keep in mind is that "models are to be used as decision-making tools". Answers derived from modeling cannot be used to replace good management judgements. Predictions and projections obtained from models should be used to support study conclusions - not to determine them in an absolute way based on whether or not results fit within criteria. A careful and extensive examination of model reliability is needed before a model's predictive accuracy can be trusted. There are many hydrologic transport modeling aspects that contribute to predictive uncertainties, such as incompletely described processes (like field-scale dispersion, chemical-media interactions) and the limited measureability of hydraulic properties. A manager needs to be familiar with those difficulties and take them into account in making judgements based on modeling results. The subjects of stochastic hydrology, parameter sensitivity analysis, and uncertainty analysis all contribute to a determination of confidence limits in model predictions. A manager who expects to have to defend decisions based on modeling results should strongly consider including a sound reliability analysis which will describe the confidence assignable to an assessment based on modeling results. Such an analysis will be a final determination of whether or not code selection for meeting specific study objectives was successful.

During the last decade a tremendous effort has been directed toward unravelling the interactive physical, chemical and microbiological mysteries of flow and transport processes in the subsurface, with contributions being made by hydrologists, geologists, soil scientists, geochemists, microbiologists and engineers. The effort has increased our conceptual understanding of the major mechanisms affecting flow and transport, perhaps to the point of frustration when realizing the complexity of the transport problem. This complexity and variety of contaminant transport processes has led to increasing specialization within the area of transport modeling. Unfortunately, segmented, disciplinary research has contributed to a lack of experimental and theoretical understanding of the contaminant transport through variably saturated media. Thus, a more unified and interdisciplinary approach is needed that considers the most pertinent physical, chemical and biological operating processes.

Development of models is not enough. Models by themselves have no credibility with the scientific community or the public at large. Models need to be tested with field data. We are now at a point when our ability to compute far exceeds our ability to collect field data which constitutes the input for the computer models. There is a great <u>need for at least a few well-controlled field studies</u>. In very few instances can data from existing sites be used because the boundary and initial conditions are seldom known. Well designed field studies are costly, time consuming, and often non-rewarding in terms of number of publications per scientific man-hour. Computer models can be very useful in the design of such experiments. For example, some preliminary computer simulations can give useful information on the optimum placement of sensors and on the frequency of sampling.

Once comprehensive models for predicting contaminant transport have been developed and tested, it may be possible to make simplifications in the models without substantially changing the outcome. How and where these simplifications can be made is one of the challenges for future research. For example, it has been shown that, if one is interested in the concentration distribution of a single tracer leaching out of a soil profile, models which are steady state with respect to water flow are often sufficient.

In conclusion, I would like to stress that models in whatever shape or form are no substitute for rigorous hydrogeologic analysis and reasoning, indeed they demand more of it. Applying a model is an exercise in thinking about the way a system works. Models require a significant lead in time for development, spanning several years, before they can be expected to produce results of the required accuracy and calibre. Used wisely they can be effective tools of the trade in reconnaissance studies which preceed field investigations, interpretive and predictive work, hypothesis verification and sensitivity analysis. To avoid model misuse, it is important to know and understand their limitations and sources of error. Neither the modeling nor the experimental approach alone is self-correcting and likely to lead us toward the progressive acquisition of comprehensive knowledge and understanding. It is only by going back-and-forth between experimental data and theoretical models that we can advance, albeit in a tortuous and laborious way.

Modeling is surely not a substitute for experimentation, but a possibly more rational basis for experimentation. We need at least a few detailed, sound and comprehensive experiments as a basis for devising models, for supplying their parameters, and for validating the model results. Reciprocally, such results can help economize experimentation by guiding it to where it is needed most. At its best, modeling offers a vehicle for theorists and experimentalists to begin journeying together.

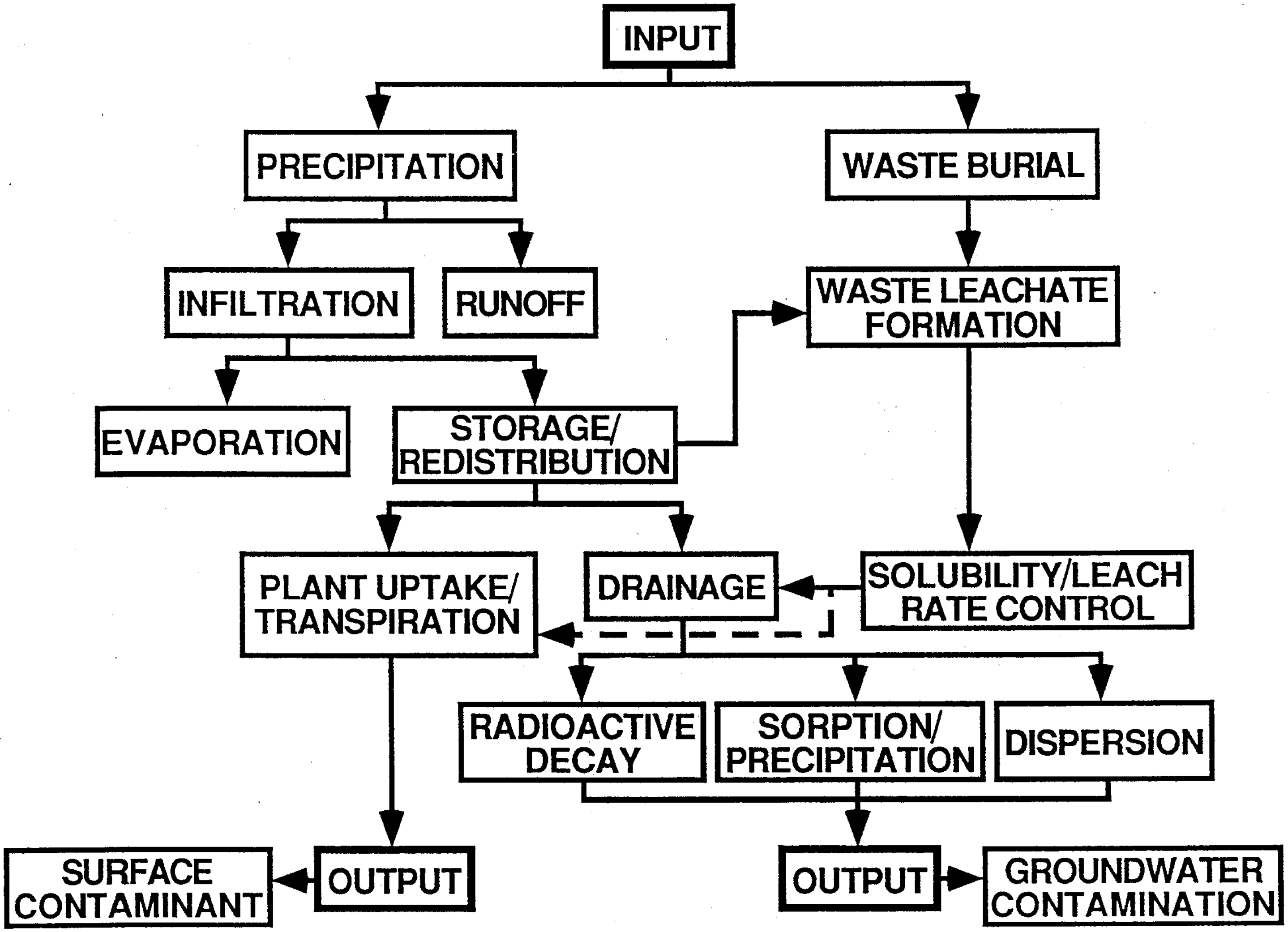
SELECTING AND EVALUATING SUBSURFACE FLOW AND CONTAMINANT TRANSPORT MODELS: ISSUES AND PROBLEMS

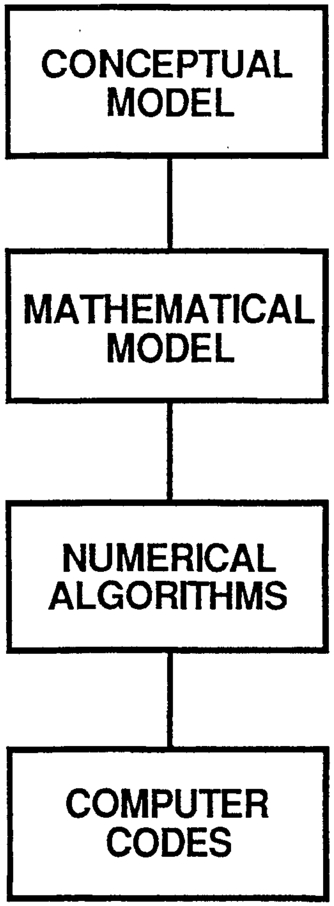
by

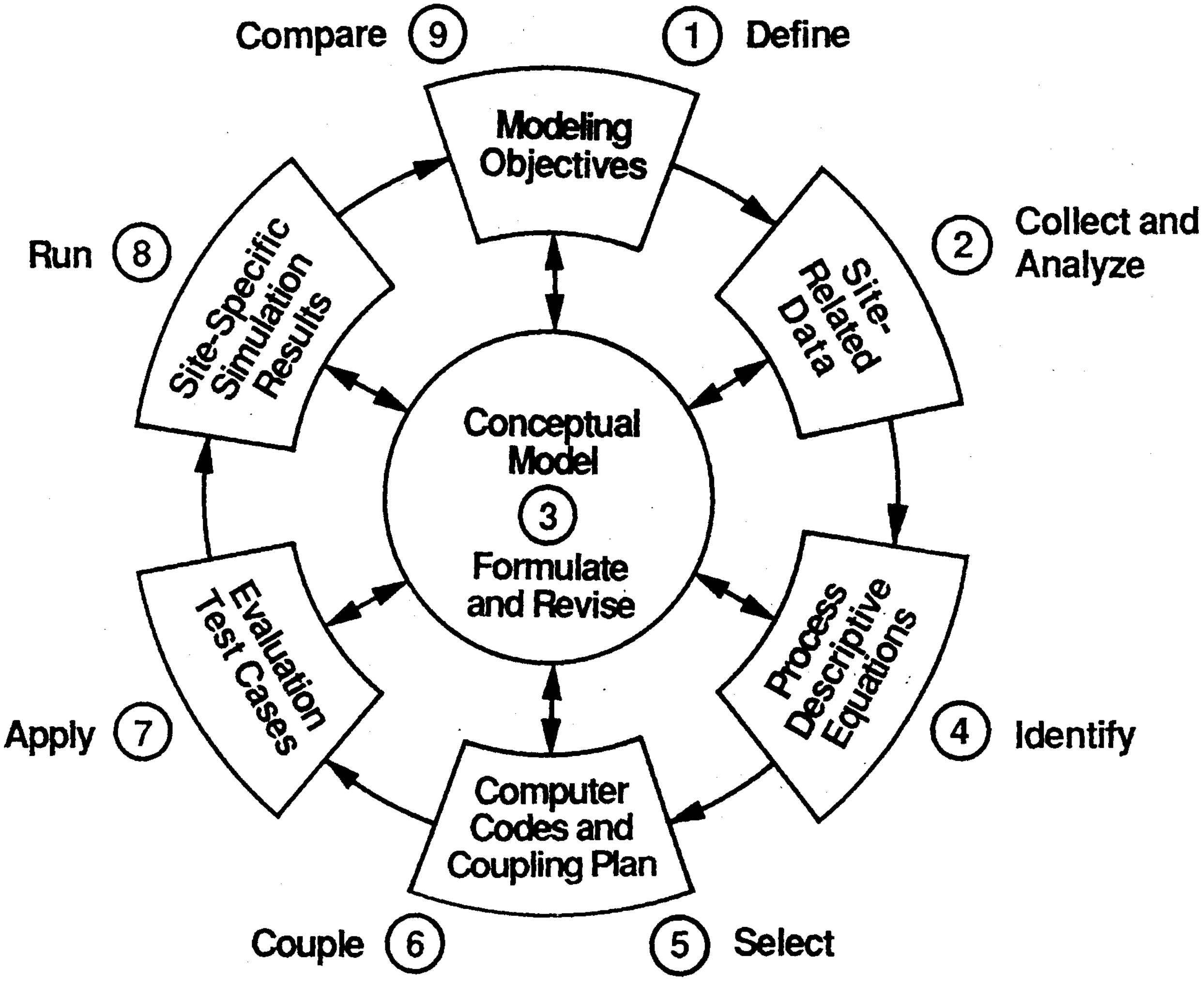
Marios Sophocleous
KANSAS GEOLOGICAL SURVEY

CONCEPTUAL MODEL:

A scientist's general understanding and approximation of the system being studied.


INPUT
PRECIPITATION
WASTE BURIAL
INFILTRATION
RUNOFF
WASTE LEACHATE FORMATION
EVAPORATION
STORAGE/ REDISTRIBUTION
PLANT UPTAKE/ TRANSPIRATION
DRAINAGE
SOLUBILITY/LEACH RATE CONTROL
RADIOACTIVE DECAY
SORPTION/ PRECIPITATION
DISPERSION
SURFACE CONTAMINANT
OUTPUT
OUTPUT
GROUNDWATER CONTAMINATION


CONCEPTUAL MODEL
MATHEMATICAL MODEL
NUMERICAL ALGORITHMS
COMPUTER CODES


Compare
9
1
Define
Modeling Objectives
Run
8
Site-Specific Simulation Results
2
Collect and Analyze
Site-Related Data
Conceptual Model
3
Formulate and Revise
Evaluation Test Cases
Process Descriptive Equations
Apply
7
4
Identify
Computer Codes and Coupling Plan
Couple
6
5
Select

TECHNICAL ISSUES OF SATURATED ZONE TRANSPORT MODELING

- Boundary conditions
- Aquifer transmissivity
- Hydraulic potential and velocity field
- Geologic structure and hydrology
- Fracture flow
- Contaminant sources
- Field-scale dispersion
- Chemical-media interaction
- Coupling groundwater and geochemical models

GENERAL TRANSPORT MODELING ISSUES

- Code requirements
- Code complexity
- Code coupling
- Dimensionality
- Geometrical perspective
- Balanced model
- Steady - state vs transient

GENERAL TRANSPORT MODELING ISSUES
cont'd

- Numerical methods
- Model sensitivity
- Spatial and temporal scales
- Model uncertainty
- Model verification and validation
- Data requirements

MODEL SELECTION

MODEL SELECTION PROCEDURE - 1

1. Has the code been tested and used by independent groups of scientists ?
2. Is an adequate summary description of the code available ?
3. Is it the most updated version of a family?
4. Does it model the important process ?
5. Does a user's manual exist ?
6. Does the code allow exclusion of irrelevant processes ?

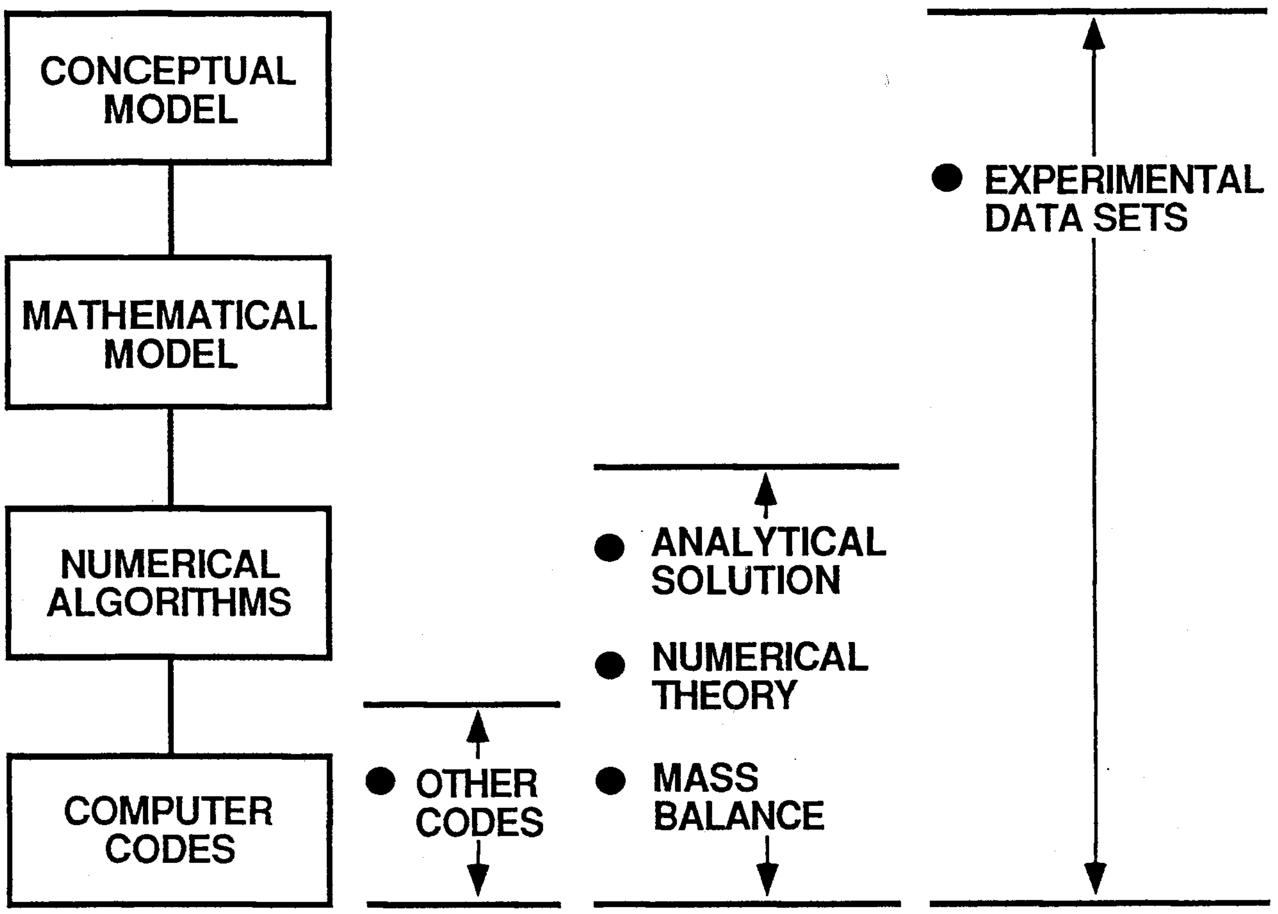
DISCARD CODE WHEN ANY ANSWER IS <u>NO.</u>

MODEL SELECTION PROCEDURE - 2

Weigh remaining codes based on

1. Appropriate geometry and structure
2. Appropriate mix of boundary conditions
3. Input requirements
4. Code application record and reputation
5. Quality of user's manual and documentation
6. Familiarity and experience with the code

MODEL EVALUATION / VERIFICATION


CONCEPTUAL MODEL
MATHEMATICAL MODEL
NUMERICAL ALGORITHMS
COMPUTER CODES
● OTHER CODES
● ANALYTICAL SOLUTION
● NUMERICAL THEORY
● MASS BALANCE
● EXPERIMENTAL DATA SETS

MODEL VALIDATION

VERIFICATION:

A confirmation that a code accurately solves the intended mathematically expressed problem.

VALIDATION:

A demonstration in which a calibrated model adequately describes physical reality.

CALIBRATION:

Adjustment of model input using alternative combinations of parameter values so as to obtain reasonable agreement with measured data.

MODEL VALIDATION REQUIREMENTS

- Independent data set not used in the calibration process.
- Data set reflecting some change in the system.

Do the highly sophisticated models really represent the field situation?

SENSITIVITY ANALYSIS:

An analysis that quantifies the change in model output results produced by a change in input parameter values.

Models are to be used as decision-making tools

MODEL RELIABILITY:

Uncertainty bounds

INTERDISCIPLINARY APPROACH

Need for well-controlled field studies

MODELING:

A tool for rigorous analysis and reasoning; a common ground for theorists and experimentalists.