6-15-1992

Ground-Water Modeling Issues in Ground-Water Development: Model Calibration and Verification

Paul K.M. van der Heijde

Paul K.M. van der Heijde, *Ground-Water Modeling Issues in Ground-Water Development: Model Calibration and Verification*, *in* UNCOVERING THE HIDDEN RESOURCE: GROUNDWATER LAW, HYDROLOGY, AND POLICY IN THE 1990S (Natural Res. Law Ctr., Univ. of Colo. Sch. of Law 1992).

Ground-Water Modeling Issues in Ground-Water Development: Model Calibration and Verification

by

Paul K.M. van der Heijde
International Ground Water Modeling Center
Colorado School of Mines
Golden, CO 80401

presented at

Rocky Mountain Ground Water/C.U. School of Law Conference

UNCOVERING THE HIDDEN RESOURCE:
GROUND WATER LAW, HYDROLOGY, AND POLICY
IN THE 1990S

June 15-17, 1992
Boulder, Colorado

Calibration

Introduction

The second stage of model application, calibration, starts with the design or improvement of the model grid and the preparation of an input file by assigning nodal or elemental values and other data pertinent to the execution of the selected computer code.

The actual computer simulation then takes place, followed by the interpretation of the computed results and comparison with observed data. The results of this first series of simulations are used to further improve the concepts of the system and the values of the parameters.

Sensitivity runs are performed to assist in the calibration procedure.

More data may be needed during the calibration process. In some cases the code is used initially to design a data collection program. The newly collected data are then used both to improve the conceptualization of the system and to prepare for the predictive simulations.

Calibration is the process of adjusting model inputs until the resulting predictions give a reasonable good fit to observed data (NRC 1990). Model inputs include:

- constitutive coefficients and parameters (e.g. hydraulic conductivity, dispersion coefficients and partition coefficients);
- forcing terms (e.g. sources and sinks for water or contaminants);
- boundary conditions (specified heads, concentrations, and fluxes).

Calibration (continued)

Commonly, calibration is started with the best estimates of values for model input based on measurements.

The degree of allowable adjustment of any parameter is generally directly proportional to the uncertainty of its value or specification and limited to its expected range of values or confidence interval (Konikow and Patten, 1985).

Usually, the model is considered calibrated when it reproduces historical data within some acceptable level of accuracy, determined prior to the calibration exercise.

Background

Most groundwater models are based on a detailed description in space and time of the physical and chemical processes involved in the movement of water and the transport of contaminants. Simplifications made are in either the physical structure of the system under study or the physical or chemical processes involved. In surface water modeling these simplifications often take the form of parameter lumping or spatial averaging. In such a system no account is taken of variations within the modeled area of inputs such as recharge, pumping, or hydraulic parameters.

The relationship in a physically based lumped parameter model between input and output is based on field observations, which are used to fit the model. Relative few parameters are involved.

In groundwater modeling an almost infinite number of parameters have to be determined, due to the discretization in the spatial domain. Many different input combinations may result in the same system response (or fit to the historical observations).

The primary practical solution to this "ill-posed" problem is the use of "prior information" to guide or constrain calibration.

Calibration (continued)

Notes:

- A good fit to historical data does not guarantee good predictions, particularly if the historical fit is based on a small amount of data, or if it does not test the model capabilities that are required for making predictions (NRC 1990).

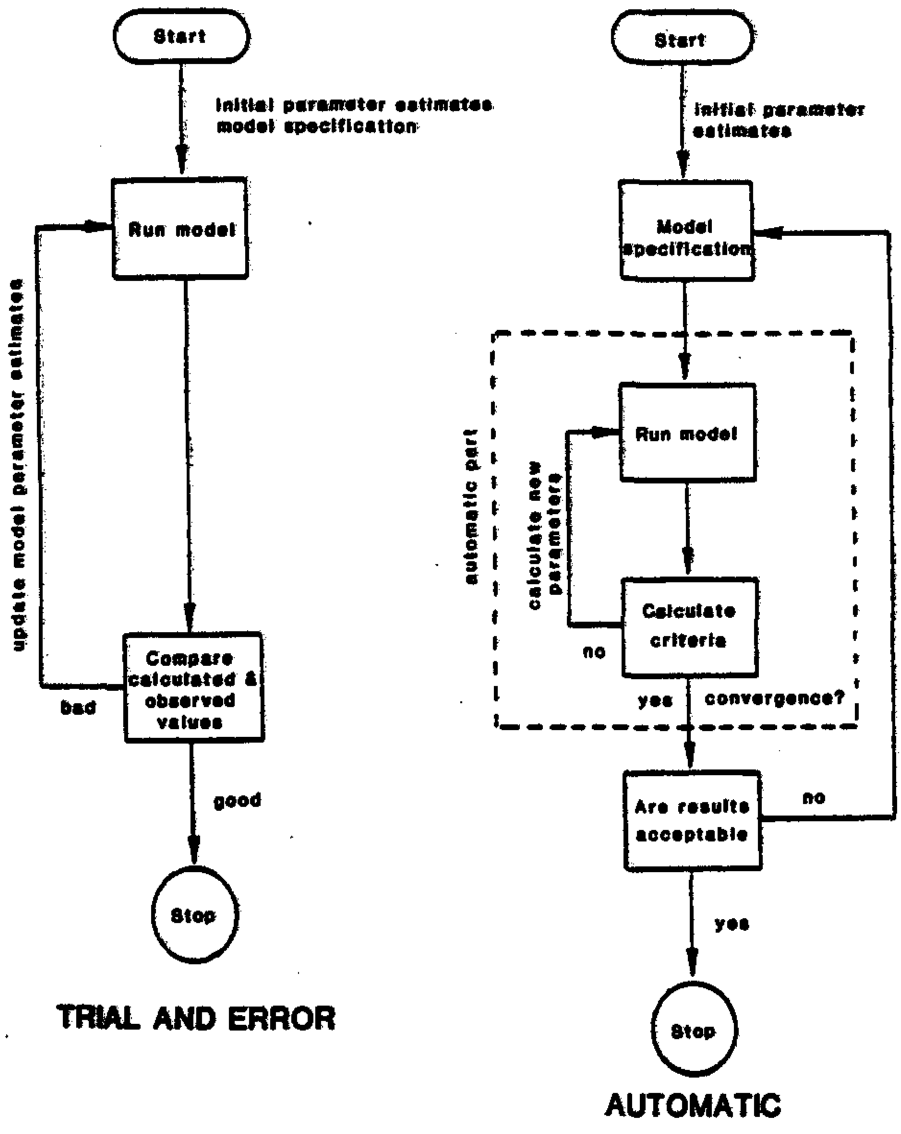
- Calibration of a deterministic model is often accomplished through a trial-and-error adjustment of the model input data. Because a large number of interrelated factors influence model output, this may become a highly subjective procedure (Konikow and Patten, 1985).

- Automatic parameter identification procedures may help to eliminate some of the subjectivity inherent in model calibration (Konikow and Patten, 1985).

- The success of model calibration is dependent on the validity of the underlying model formulation; if the model's structure ignores important sources, geological heterogeneities, physical processes, or chemical reactions, parameter estimation and model calibration will be reduced to a fitting exercise that forces available inputs to compensate (usually inadequately) for an proper formulation (NRC 1990).

- The hydrological experience and judgement of the modeler continues to be a major factor in calibrating a model both accurately and efficiently (Konikow and Patten, 1985).

History matching/calibration using trial and error and automatic procedures



Mercer and Faust 1981

History Matching Calibration

Example of transmissivity adjustment

$$h_{c_{i,j}} - h_{o_{i,j}} = \Delta h_{i,j} \rightarrow \Delta T_{i,j} = f(\Delta h_{i,j})$$

$$T_{i,j_{new}} = T_{i,j_{old}} + \Delta T_{i,j} \quad (\text{ or } T_{i,j_{old}} * \text{Corr}_{i,j})$$

or

$$T_{i,j_{new}} = T_{i,j_{old}} + g\{\Delta T_{i,j}, \Delta T_{i+1,j}, \Delta T_{i-1,j}, \Delta T_{i,j+1}, \Delta T_{i,j-1}\}$$

h_c = calculated head

h_o = observed head

T = transmissivity

ΔT = calculated or given change in T

Examples of such a correction

- $Corr_{i,j} = (1 + \Delta T_{i,j})$ as in previous example

- $Corr_{i,j} = \dfrac{TGC_{i,j}}{TGO_{i,j}}$

where TGC is calculated hydraulic gradient and TGO is observed hydraulic gradient or in F.D. approximation:

$$TGC_{i,j} = \left(\frac{\delta h}{\delta s}\right)_{i,j} = \frac{1}{2.\Delta x}\sqrt{\{(h_{i+1,j} - h_{i-1,j})^2 + (h_{i,j+1} - h_{i,j-1})^2\}}$$

and same for $TGO_{i,j}$ assuming these values can be derived from water-level map.

Calibration

Least Squares

Minimize $F = \sum_{i=1}^{N} [h_{o_i}(x_1\, y_1\, z_1\, t) - h_{c_i}(x_1\, y_1\, z_1\, t)]^2$

Using sensitivity coefficients or sensitivities:

$$\frac{\delta F(h)}{\delta T_j} = \sum_{i=1}^{N} \left[-2\,(h_{o_i} - h_{c_i})\,\frac{\delta h_{c_i}}{\delta T_j}\right] = 0$$

↑ sensitivity coefficient

1. initial guess for parameters

2. iterate with improved estimate

Calibration

Trial and Error Methods

Disadvantages

- No methodology exists to guarantee that simulations will proceed in direction leading to best set of parameters

- It is difficult to determine when best set has been reached, expecially in case of large number of calibration parameters (type and number of zones)

- No practical way of determining how many other sets of parameter values could yield similar correspondence between $h_{observed}$ and $h_{calculated}$

- Difficult to decide whether or not additional parameters or a more refined model would improve model fit

- No way of quantitavely assessing the predictive reliability of the model

History Matching/Calibration

Trial and error approach

- **Completion depends on:**
  - objectives for analysis
  - complexity of flow system
  - length of observed history
  - budget
  - expertise of modeler
  - patience of modeler (or manager)

- **Automatic**
  - completion achieved when preset matching criteria are met

Advantages and Disadvantages of Trial and Error and Automatic History Matching Procedures

Automatic History Matching

Advantages

- less subjective
- fewer computer runs
- statistical estimates of confidence

Disadvantages

- programs less well documented
- statistical training necessary
- still a research tool

Trial and Error

Advantages

- well documented programs
- conceptually straightforward

Disadvantages

- time consuming
- subjective

Calibration

Example: simple indirect solution of inverse problem involving repetitive solution of forward problem while adjusting T in a trial and error fashion

Step 1: Use prior information to formulate hypotheses for spatial distribution of constant zones of T

Step 2: Perform sensitivity analysis for zones of T

Step 3: Adjust T to improve agreement of computed values of h with measured values of h

Step 4: Check reasonability of final parameter distribution

Step 5: Perform sensitivity analysis for other variables and parameters. Often, all parameters and variables have a level of uncertainty. This means that the level of uniqueness of the solution diminishes even further.

Calibration (continued)

Model calibration can be performed to steady-state or transient data sets.

Most calibrations are performed under steady-state conditions but may also involve a second calibration to a transient data set (Anderson and Woessner 1992).

Steady-state:

- to determine time-independent parameters, e.g. transmissivity, hydraulic conductivity, leakance
- to determine long-term (average) values for parameters which might change over time, e.g. recharge rates, leakage rates

Transient:

- to determine time-dependent parameters, e.g. storativity
- to determine time-dependent values for parameters which change over time, e.g. recharge rates
- to check initial calibration of time-<u>in</u>dependent parameters

Calibration continued)

Calibration Targets for Flow

Field-measured values of heads and fluxes form the *sample information* or *calibration values* (Anderson and Woessner 1992).

The calibration value with its associated error forms the *calibration target*, which should be determined before calibrating the model.

The associated error might consist of:

- measurement error
- scaling error (representativeness of measurement for model variable)
- interpolation error (from transferring measured information to nodal values)

Field-measured fluxes might include:

- base flow
- spring flow
- infiltration from a losing stream
- evapotranspiration from the water table

Additional calibration information for a flow model can be obtained from velocities and solute distributions.

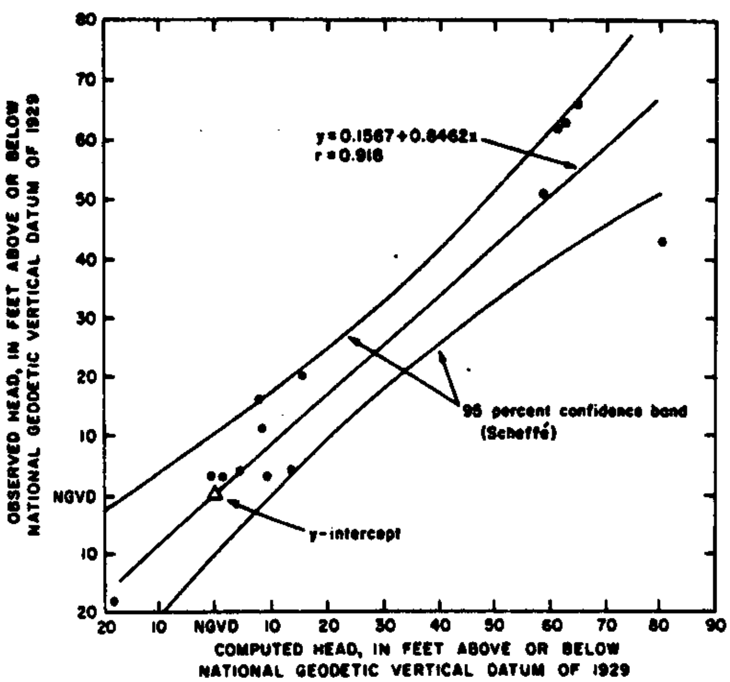
Such additional calibration information might increase the likelihood to obtain a "unique" solution.

Calibration (continued)

Measures of Calibration

1. Qualitative

- comparison between contour maps of measured and simulated heads, providing information on the spatial distribution of the error.
- contouring the calibration error (residuals)
- representation of calibration error per cell or element
- scatterplot of measured versus simulated heads; deviation of points from the straight line should be randomly distributed (might include confidence intervals for the linear regression):



- tabulation of measured and simulated heads for each node.

Calibration (continued)

Measures of Calibration (continued)

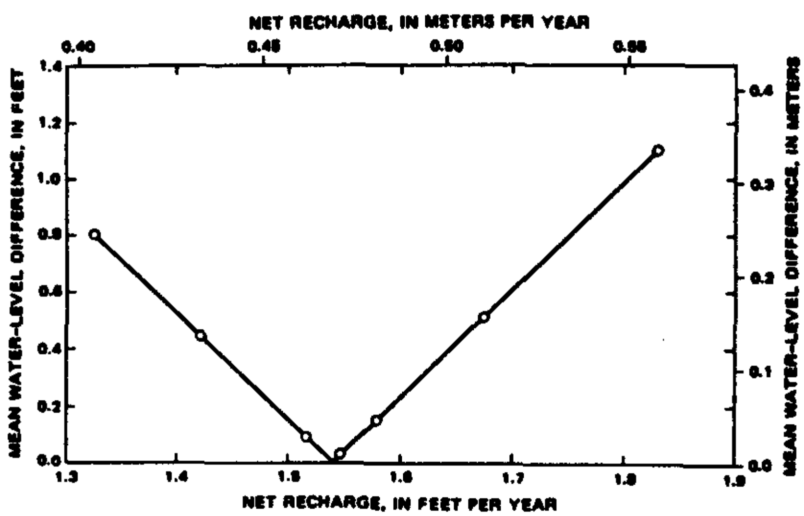
2. Quantitative (calibration criterion or performance criterion)

- *mean error (ME)* is the mean difference between measured heads and simulated heads:

$$ME = \frac{1}{n}\sum_{i=1}^{n} (h_m - h_s)_i$$

with n the number of calibration values.

The ME can be represented in a graph against various values of the calibrated parameter (Anderson and Woessner 1992):



Both negative and positive differences are incorporated in the ME and may cancel out each other.

Calibration (continued)

Measures of Calibration (quantitative - continued)

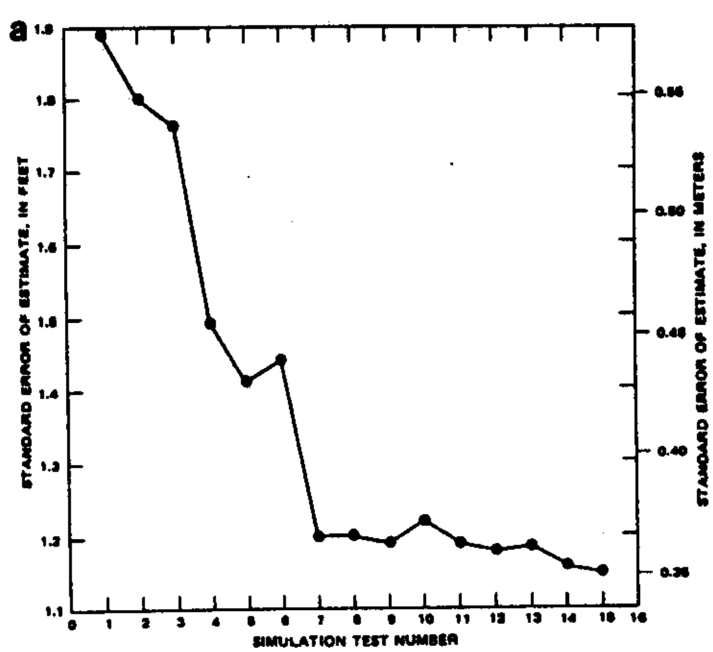
- The *mean absolute error (MAE)* is the mean absolute value of the differences in measured and simulated heads:

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |(h_m - h_s)_i|$$

- The *root mean squared (RMS) error* or the *standard deviation* is the average of the squared differences in measured and simulated heads:

$$RMS = [\frac{1}{n}\sum_{i=1}^{n} (h_m - h_s)_i^2]^{0.5}$$

The systematic reduction of the standard error of estimate (SE) and RMS error may be shown graphically:



Calibration (continued)

Levels of Calibration (Anderson and Woessner 1992)

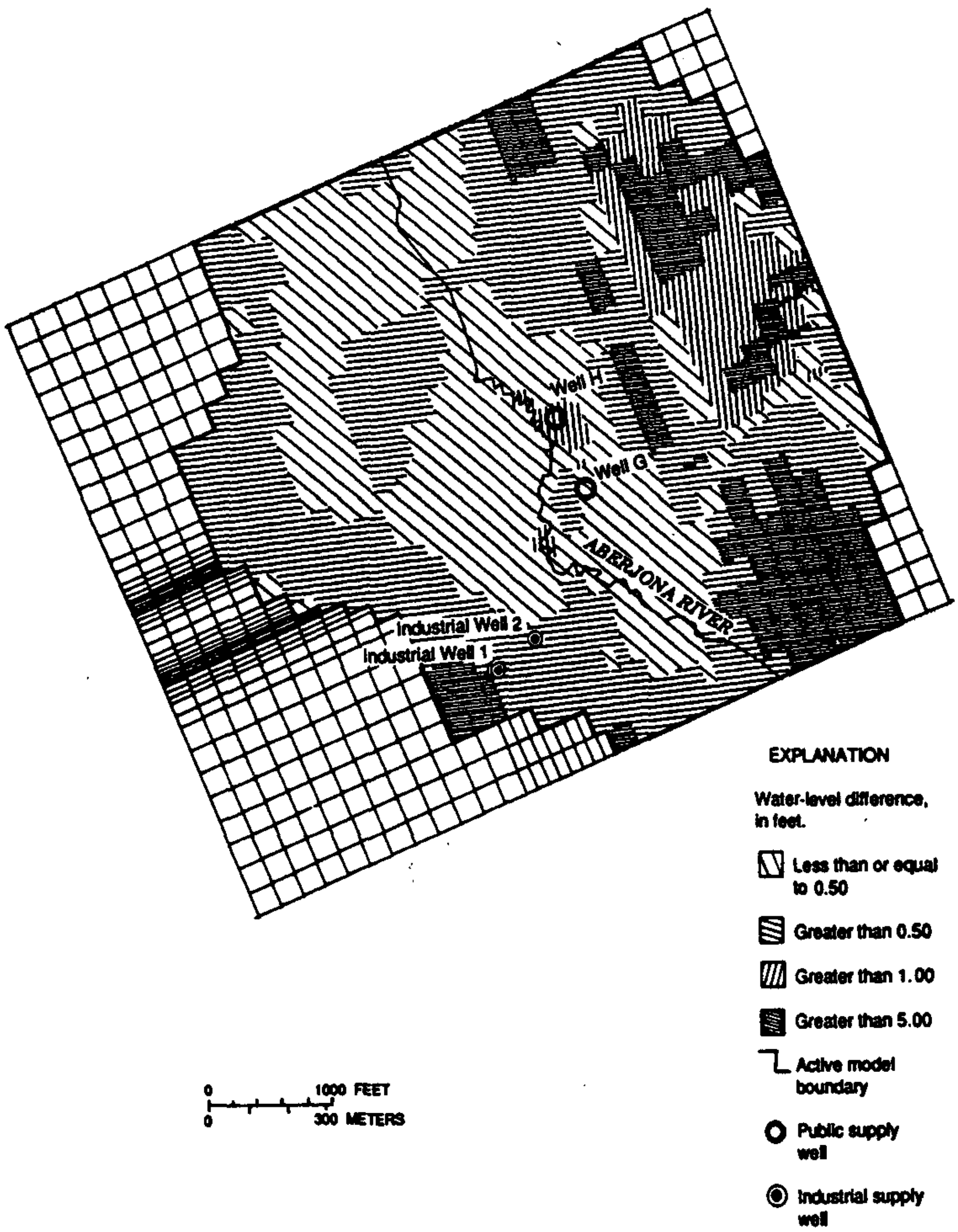
- level 1 — simulated value falls with target (highest degree of calibration)
- level 2 — simulated value falls within two times the associated error of the calibration target
- level 3 — simulated value falls within three times the associated error of the calibration target
- level n — simulated value falls within *n* times the associated error of the calibration target (lowest degree of calibration)

Problems in Calibration:

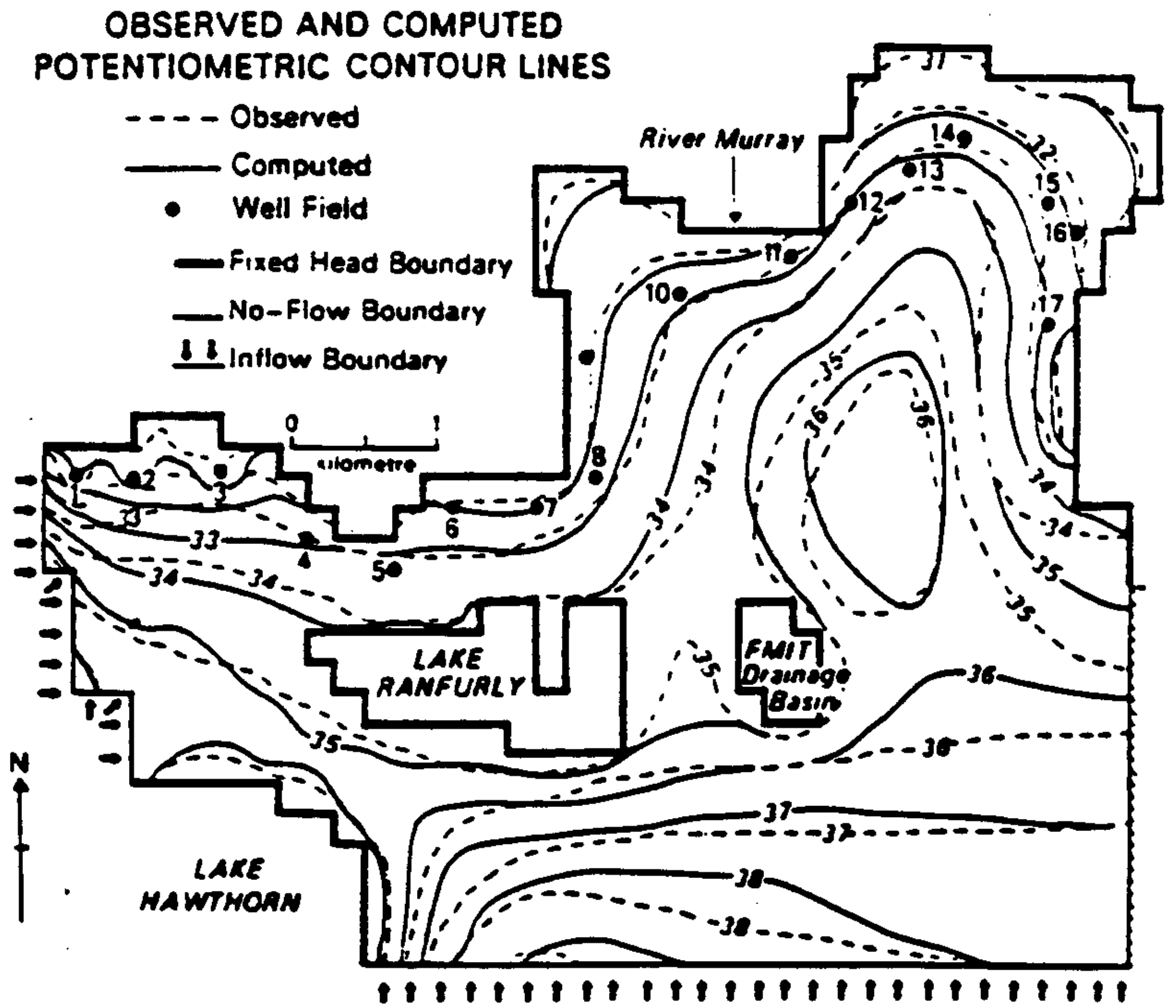
- inverse problem is mathematically ill-posed
- measurements are not available for all locations
- where measurements exist they are not accurate
- most important parameters might not have been measured

Calibration (continued)

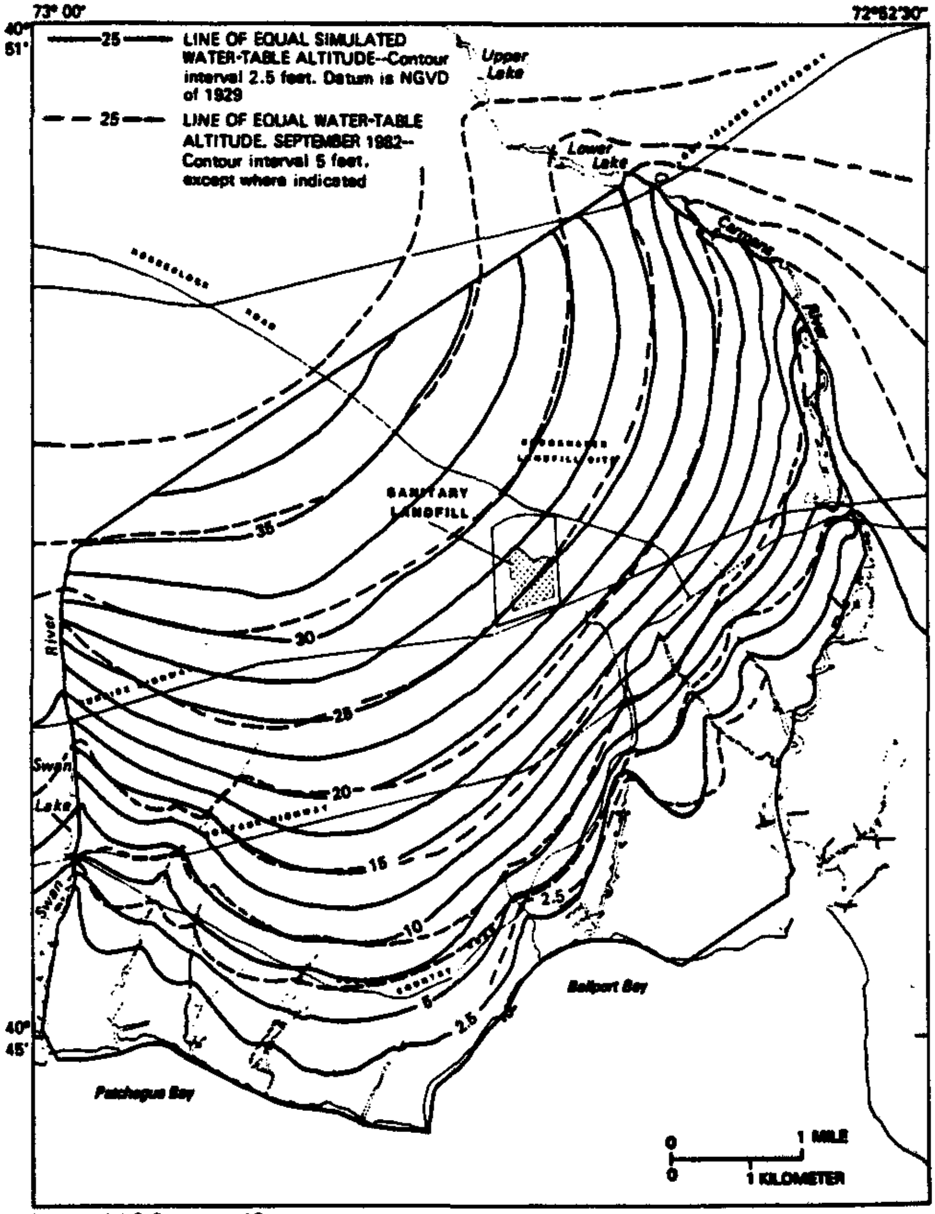
Representation of the spatial distribution of error of residual calculated as the difference between measured (or interpolated) heads and simulated heads for an unconfined aquifer in Woburn, Massachusetts (de Lima and Olimpio, USGS-WRI 89-4059, 1989)



Measured and comuputed hydraulic heads of the Parilla Sand aquifer for January 1980

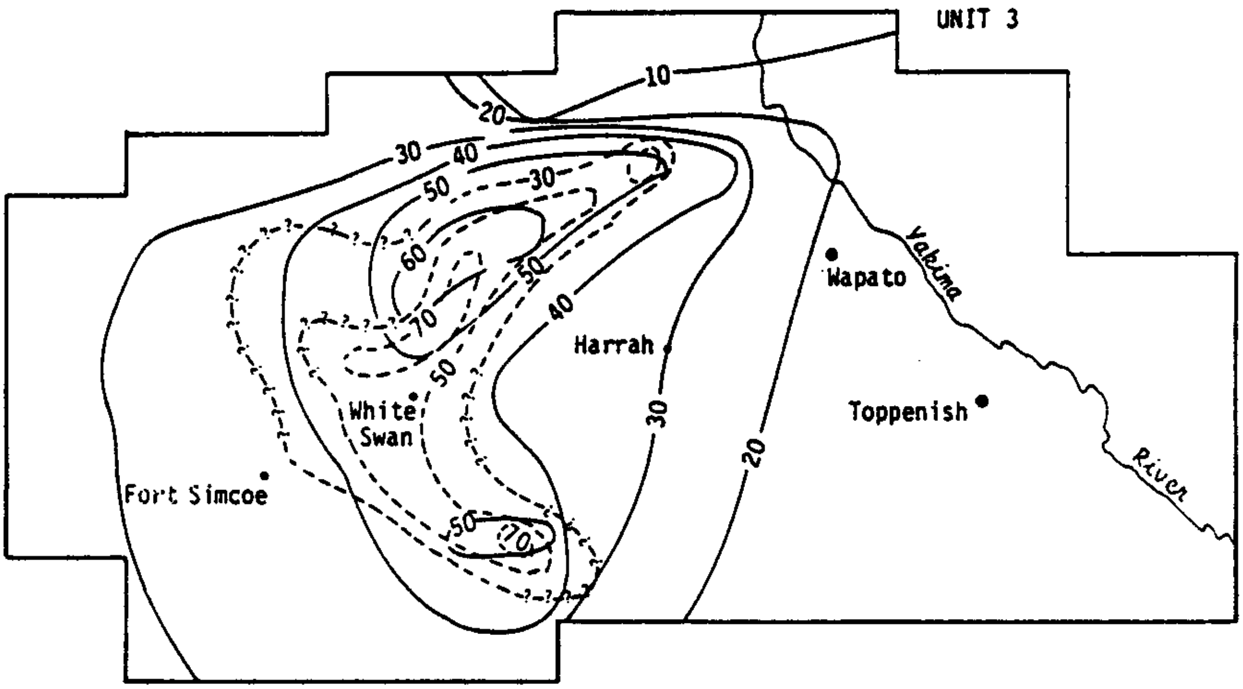


Simulated and observed (September 1982) water-table



Base from N.Y.S. Department of Transportation, Bellport, 1981; Howells Point, 1981, NY, 1:24,000

Comparison of measured and calculated 1960-70 head decline in unit 3



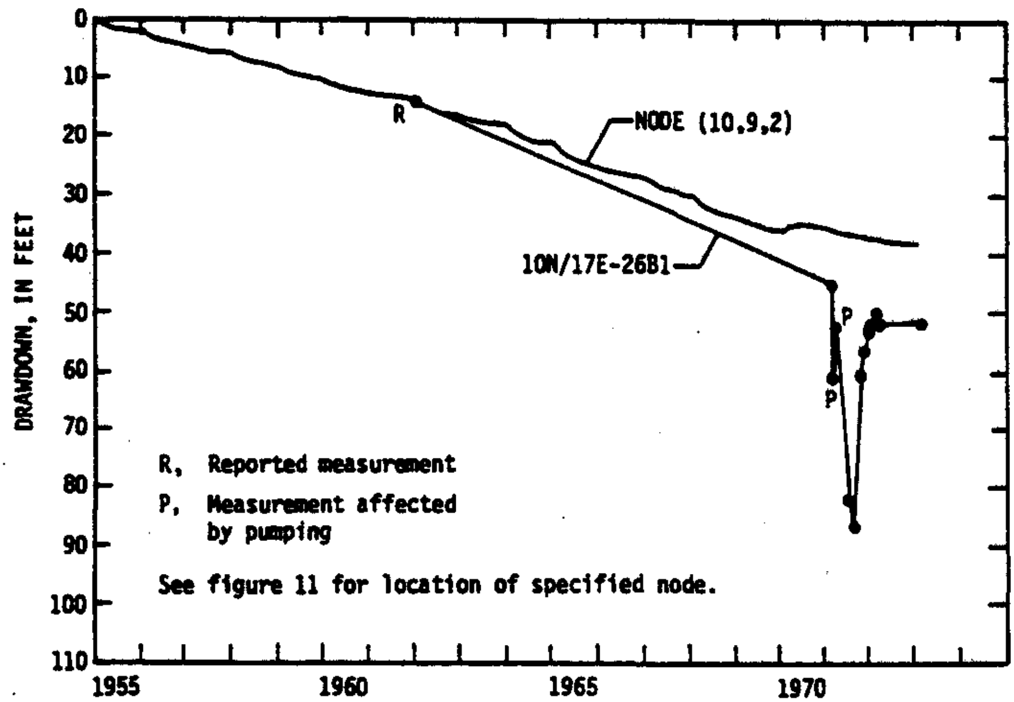
EXPLANATION

--- 30 --- Line of measured equal head decline, queried where approximately located

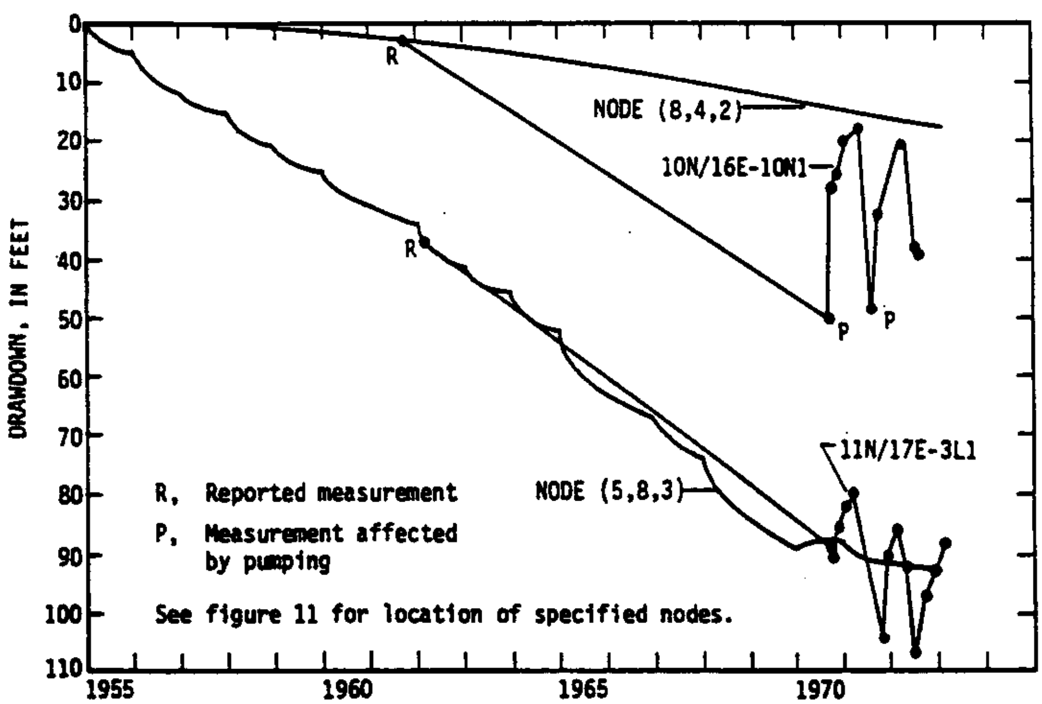
— 40 — Line of calulated equal head decline. Interval variable, in feet

Skrivan USGS WRI 1987 82-4010

Comparison of measured and computed drawdown at node (10,9,2)



Comparison of measured and computed drawdown at nodes (8,4,2) and (5,8,3)

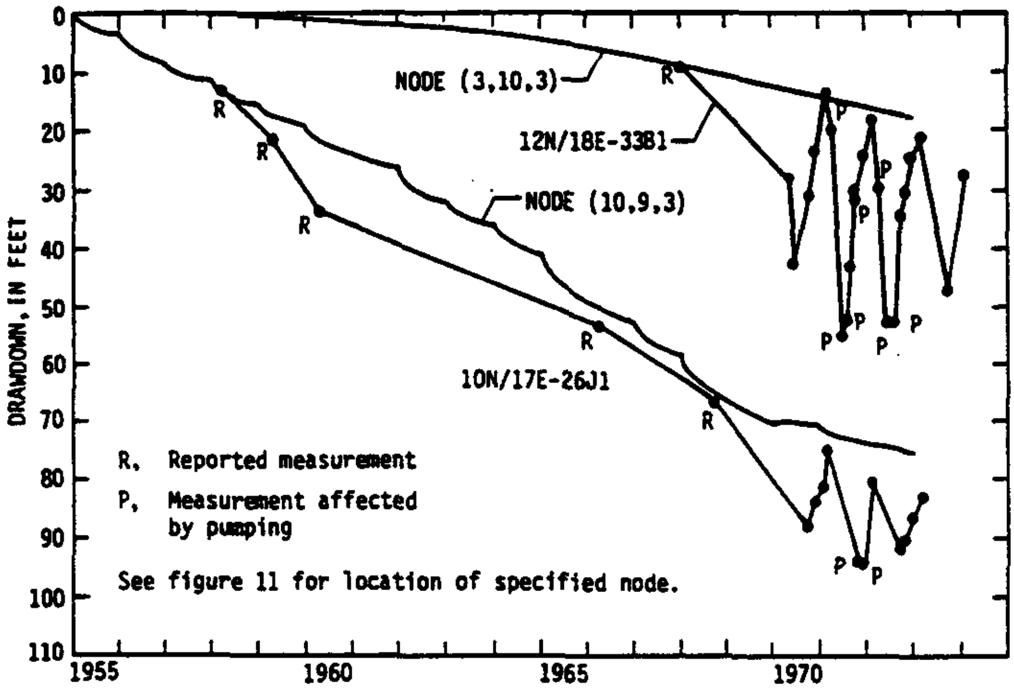


Sources and discharges simulated in the calibrated model

Data in cubic feet per second

Sources of water		
Areal recharge		118.1
Flow across boundaries		204.7
Recharge ponds at major industry		2.5
Recharge ponds at North Main Street well field		12.4
Total sources		337.7
Discharges of water		
Flow across boundaries		43.4
Pumping (layers 1 and 2)		17.6
River leakage:		
Christiana Creek	17.3	
Baugo Creek	5.7	
Elkhart River	37.6	
Pine Creek	4.6	
St. Joseph River	209.0	
Little Elkhart River	2.8	
Total river leakage		277.0
Total discharges		338.0
Percent difference between sources and discharges		.09

Comparison of measured and computed drawdown at nodes (3,10,3) and (10,9,3)



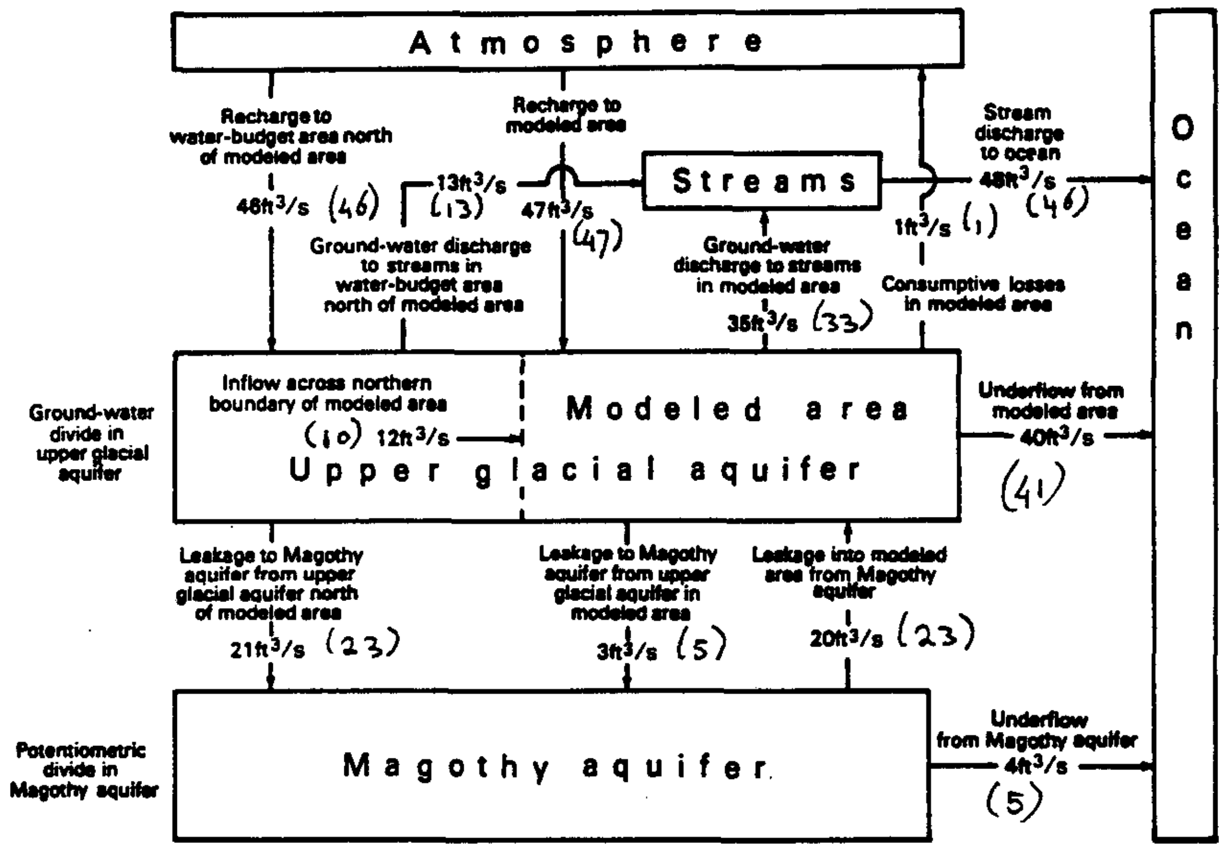
Model-simulated and measured seepage

Streams and reaches	Model-simulated seepages: Positive value, river reach gaining; negative value, river reach losing (ft^3/s)	Measured seepages: Range based on measurement error. Positive value, river reach gaining; negative value, river reach losing (ft^3/s)
Christiana Creek		
Reach 1	12.3	4.4 ------- 18.3
Reach 2	2.6	-3.7 ------- 11.7
Reach 3	-12.4	-18.5 ------- -3.7
Baugo Creek		
Reach 4	4.5	4.4 ------- 6.8
Elkhart River		
Reach 5	13.4	3.6 ------- 50.5
Reach 6	8.9	-34.9 ------- 9.9
Reach 7	11.5	2.5 ------- 45.5
Pine Creek[1]		
Reach 8	2.8	1.8 ------- 3.3
Reach 9	1.5	1.4 ------- 2.3

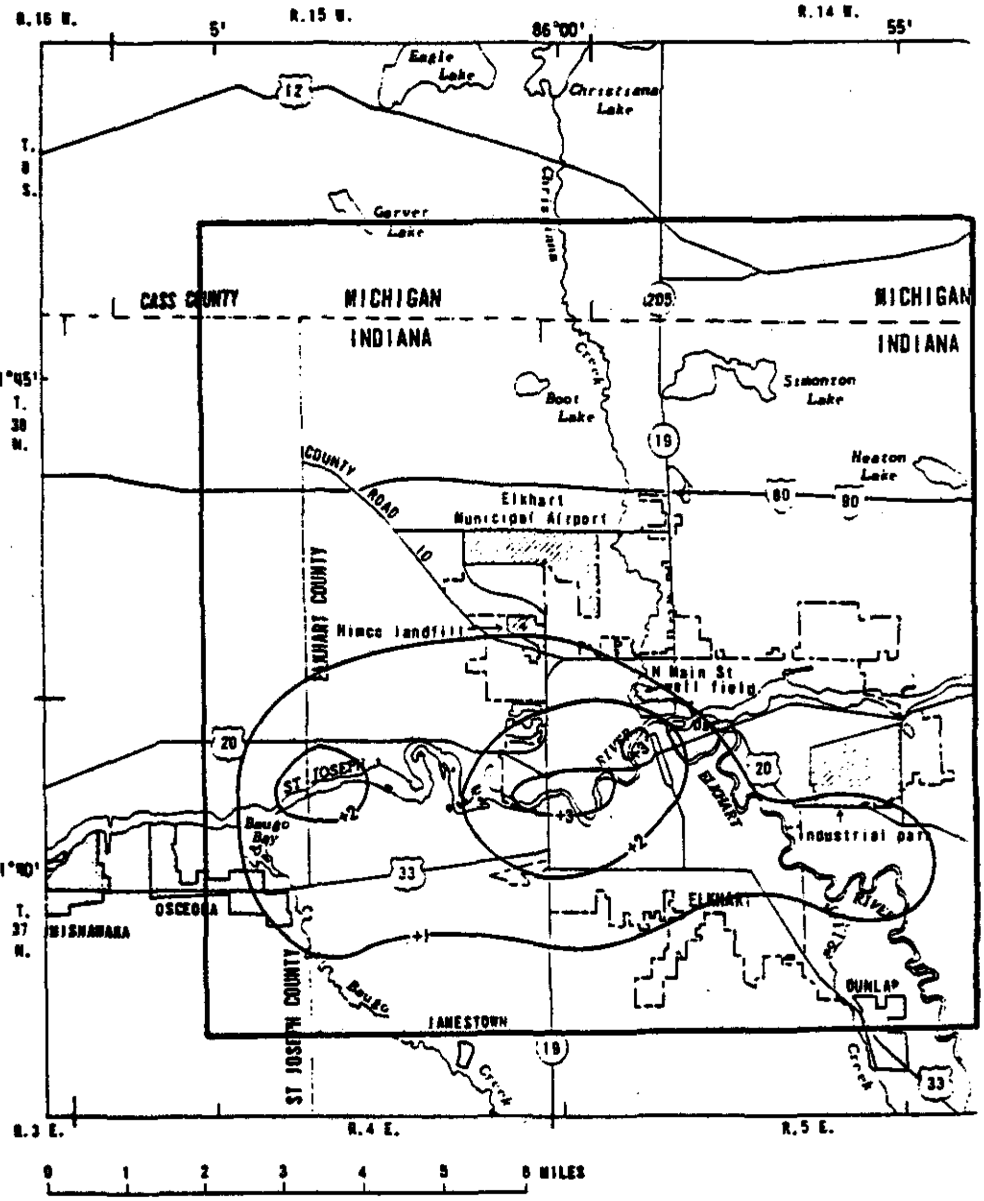
[1]Seepages for Pine Creek were measured in September 1979. Seepages for all other streams were measured in May-June 1979.

Components and rates of flow in water-budget area

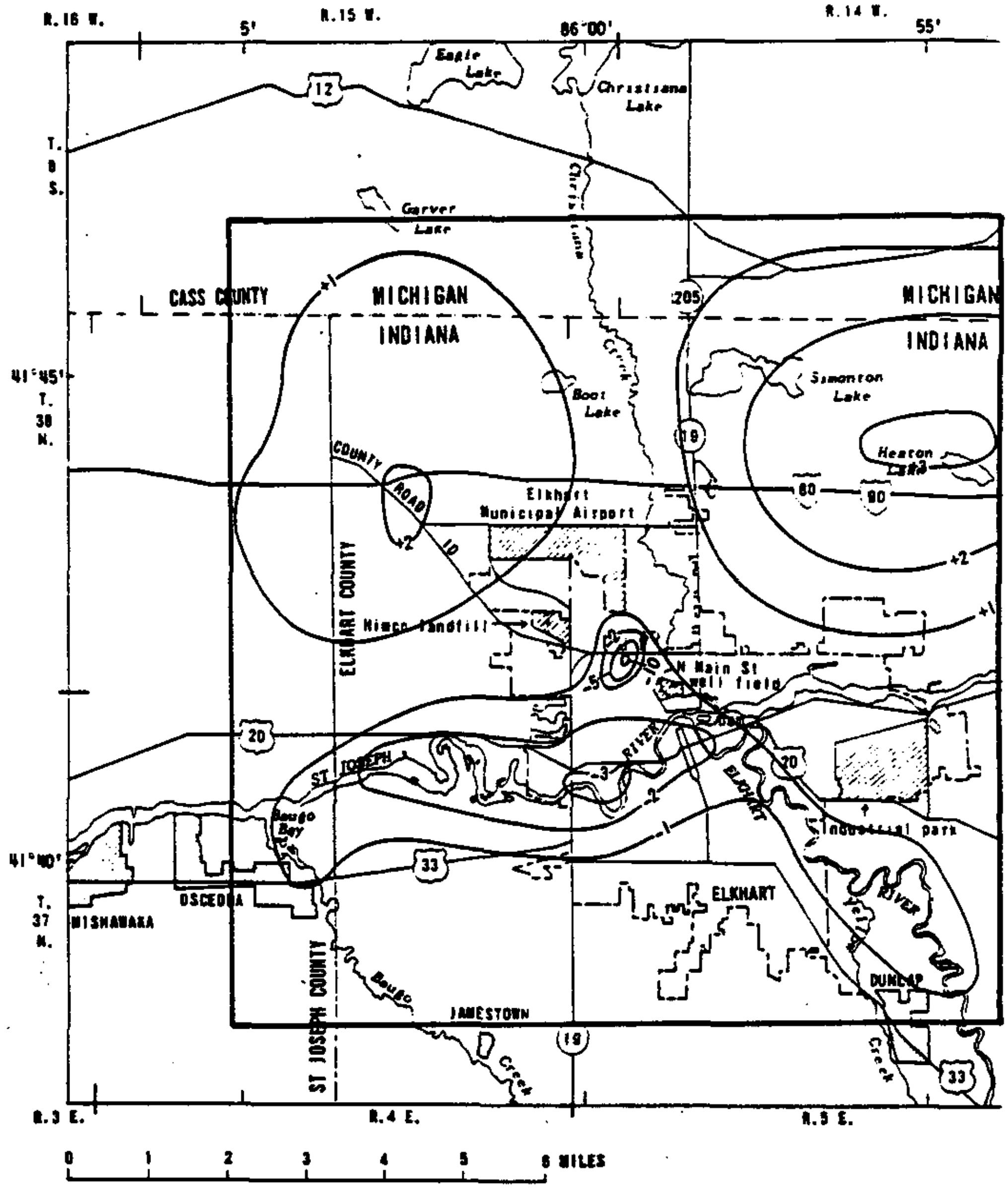
46 ft³/s = determined from water-budget analysis
(46 ft³/s) = determined by calibrated model



Changes in water levels due to a 50 percent decrease in the calibrated streambed leakance (top aquifer)



Changes in water level in top aquifer due to a 50 percent decrease in calibrated transmissivity



River leakage and boundary flows for the calibrated model and the sensitivity analyses

	Flow leaving ground-water system			Flow entering ground-water system	
	Across model boundaries		Into rivers	Across model boundaries	
	Layer 1 (ft^3/s)	Layer 2 (ft^3/s)	(ft^3/s)	Layer 1 (ft^3/s)	Layer 2 (ft^3/s)
Calibrated model	28.5	15.0	266.1	140.1	64.6
Transmissivity[1]	15.7	9.8	169.2	64.4	28.7
Streambed leakance[1]	31.1	17.0	249.5	133.5	61.9
Vertical hydraulic conductivity of confining bed[1]	32.1	14.4	254.3	133.8	63.8
Areal recharge[1]	24.9	12.5	234.0	154.2	71.7

[1]Parameters decreased by 50 percent from calibrated values.

Calibration (continued)

Sensitivity Analysis

Sensitivity analysis provides a useful way to identify the model inputs that have the most influence on model predictions, at least over a specified range. Although a detailed sensitivity analysis can be laborious and time-consuming, it is usually feasible to carry out a small scale exploratory analysis that focuses on a few critical inputs identified, most likely by intuition.

Sensitivity analysis should be performed initially at the beginning of calibration to design a calibration strategy. After the calibration is completed a more elaborate sensitivity analysis is performed to quantify the uncertainty in the calibrated model caused by uncertainty in the estimates of aquifer parameters, stresses, and boundary conditions.

During the final sensitivity analysis, calibrated values for hydraulic conductivity, storage parameters, recharge and boundary conditions are systematically changed within the previously established plausible range.

Sensitivity analysis is typically performed by changing one parameter at a time and evaluating the effects on the distribution of heads and other computed variables.

Results of a sensitivity analysis should be qualitatively discussed (Anderson and Woessner 1992):

Condition varied	Range tested	Results
Hydraulic conductivity	1000–2870 ft/d	Major effect, more intruded for high-K case. Width of transition zone doubles at aquifer base over range tested.
Anisotropy	10:1–500:1	Slightly more intruded for high-anisotropy case. Tilt and curvature of front *increase* with *higher anisotropy*.
Recharge	6–18 in./yr	Major effect, more intruded for low-recharge case. Width of transition zone doubles at aquifer base over range tested
Lateral influx	5–15 ft²/d/ft	*Minor* effect in *movement* and shape of front for range tested.
Pumping well	0–5 ft³/d/ft	Similar response as to lateral influx case.
Dispersivity	100–800 ft (α_L) 5–50 ft (α_T)	Insensitive to changes of this magnitude of α_L. *Increasing* values of α_T makes front more vertical, less intruded at aquifer base.
Layering (K_{top} : K_{bot})	1:1–1:0.01	Pronounced layering compresses saltwater front, increases curvature. Intrusion is especially limited to upper (permeable) part of aquifer.
Saltwater boundary	Atlantic Ocean–inland saltwater bodies	Transition zone translates laterally depending on concentration at aquifer top.

From Anderson et al., 1988.

Model Verification

As the set of parameters used in the calibrated model may not accurately represent field values, the calibrated parameters may not represent the system under a different set of boundary conditions or hydrologic stresses.

Model verification will help establish greater confidence in the calibration and the predictive capabilities of the calibrated model.

A model is "verified" if its accuracy and predictive capability have been proven to lie within acceptable limits of error by tests independent of the calibration data.

In general, verification is performed using a transient data set, e.g. the response of heads to drought or long-term pumping.

If only a single time-series is available, the series may be split in two sub-series, one for calibration (e.g. representing the response of the system to natural cyclic variations in stress), and another for verification (e.g.representing the response of the system to a significant changed stress condition such as caused by irrigation development).

If such data are not available, verification may be performed using an second "independent" steady-state data set (i.e. not used previously for calibration).

Note: If the parameters are changed during the verification, this exercise becomes a second calibration and the first calibration needs to be repeated to account for the changes. The second exercise cannot be considered "verification" anymore.

Post-Audits

Whenever an opportunity exists to obtain further field information regarding the system being modeled, refinements and improvements in the model should be made and previous analysis modified. Sometimes, such an opportunity is offered in the form of post-audits. Post-audits are reviews performed some time after the model-based predictions were made and often provide an opportunity for in-depth analysis regarding the inaccuracies in those predictions. However, not many of such post-audits actually take place, depriving modelers and managers from important feedback and educational experience.

Conclusion

Often, a major impediment to the efficient use of models in groundwater management is the lack of data. Data insufficiencies might result from inadequate resolution in spatial data collection (e.g., spatial heterogeneities relevant on smaller scale than sampled), or in temporal sampling of time-dependent variables (e.g., measured too infrequently), and from measurement errors.

Many types of problems can occur in the application of models. Some of these are technical, method-dependent problems such as numerical dispersion and oscillations in transport models.

Conceptual problems, often significant, can be related to the mechanisms (e.g., dispersion, adsorption, multiphase or multifluid flow), the heterogeneity of the medium, or the simplifying assumptions adopted (e.g., vertical averaging).

Finally, problems external to the model execution can occur, such as those caused by the absence of good data, model availability, available computer facilities, skilled professionals, and competent technicians.

Selected References

Anderson, M.P. and W.W. Woessner. 1992. Applied Groundwater Modeling: Simulation of Flow and Advective Transport. Academic Press, New York, New York.

Konikow, L.F. and E.P. Patten, Jr. 1985. Groundwater Forecasting. In: M.G. Anderson and T.P. Burt (eds.), Hydrologic Forecasting, pp. 221-270. John Wiley and Sons, Ltd., Chichester, U.K.

Mercer, J.W., and C.R. Faust. 1981. Groundwater Modeling. National Water Well Association, Dublin, Ohio.

National Research Council. 1990. Ground Water Models—Scientific and Regulatory Applications. National Academy Press, Washington, D.C.