PNNL-13447

Pacific Northwest
National Laboratory
Operated by Battelle for the
U.S. Department of Energy

Transient Inverse Calibration of Hanford Site-Wide Groundwater Model to Hanford Operational Impact – 1943 to 1996

C. R. Cole
M. P. Bergeron
S. K. Wurstner
P. D. Thorne
S. Orr
M. I. McKinley

May 2001

Prepared for the U.S. Department of Energy
under Contract DE-AC06-76RL01830

DISCLAIMER

PACIFIC NORTHWEST NATIONAL LABORATORY
operated by
BATTELLE
for the
UNITED STATES DEPARTMENT OF ENERGY
under Contract DE-ACO6-76RLO183O

Printed in the United States of America

Available to DOE and DOE contractors from the
Office of Scientific and Technical Information,
P.O. Box 62, Oak Ridge, TN 37831-0062;
ph: (865) 576-8401
fax: (865) 576-5728
email: reports@adonis.osti.gov

Available to the public from the National Technical Information Service,
U.S. Department of Commerce, 5285 Port Royal Rd., Springfield, VA 22161
ph: (800) 553-6847
fax: (703) 605-6900
email: orders@ntis.fedworld.gov
online ordering: http://www.ntis.gov/ordering.htm


This document was printed on recycled paper.
(8/00)

PNNL-13447

Transient Inverse Calibration of Hanford Site-Wide Groundwater Model to Hanford Operational Impacts—1943 to 1996

C. R. Cole
M. P Bergeron
S. K. Wurstner
P. D. Thorne
S. Orr
M. I. McKinley

May 2001

Prepared for the U.S. Department of Energy
under Contract DE-AC06-76RL01830

Pacific Northwest National Laboratory
Richland, Washington 99352

Summary

Pacific Northwest National Laboratory has embarked upon a new initiative to strengthen the technical defensibility of and develop a more robust capability to incorporate uncertainty in the groundwater flow and transport model at the U.S. Department of Energy Hanford Site in Southeast Washington State. One aspect of the initiative is developing and using a three-dimensional transient inverse model approach to estimate the hydraulic conductivities, specific yields, and other site-wide scale parameters, including their uncertainties, by using data on the transient behavior of the unconfined aquifer system resulting from Hanford Site waste management since 1943. Over the historical period of Hanford operations, the large volumes of wastewater discharged to a variety of waste facilities resulted in large water table changes over most of the Hanford Site and created significant groundwater mounds (in excess of 20 m) under waste management facilities in the central part of the Site. Since 1988, the mission of the Hanford Site has changed from producing weapons to restoring the environment, and wastewater discharges have declined significantly, which has caused significant water table declines.

The three-dimensional transient inverse calibration, which was recommended by an external peer review panel, is being performed using UCODE, a universal inverse modeling code developed jointly by the U.S. Geological Survey and the International Groundwater Modeling Center of the Colorado School of Mines. The work uses the existing consolidated site-wide groundwater model implemented with the Coupled Fluid Energy and Solute Transport code (CFEST), which is the forward model whose parameters are estimated by UCODE. The transient inverse calibration uses over 76,000 water level measurements taken in about 1200 wells at Hanford since the mid 1940s. Because of the long run times and large number of simulations, a serial computational approach for the coupled flow and transport inverse would require a year or more of computational effort. Thus we developed an innovative parallel computational approach that uses an isolated network of 23 computers. The approach uses a recently developed parallel version of UCODE that communicates with a parallel task manager to propagate the multiple simulation tasks (i.e., the forward model runs) for simultaneous computation on the dedicated computers. In addition, a customized version of the forward model code CFEST was developed to simplify the specification of inverse model parameters and the large number of observations.

The existing consolidated site-wide groundwater model (referred to as the prior model in this report) was calibrated using 1979 data and a steady-state inverse approach in conjunction with trial and error transient model calibration runs using estimates of artificial discharges and a limited set of representative head observations taken between 1979 and 1996. The conceptual model for the prior model also provided the conceptual basis for the initial three-dimensional transient inverse modeling studies discussed in this report. The extended (1943–1996) calibration period dataset for the initial transient inverse modeling consisted of new estimates of artificial discharges and river stage variations before 1979 and a complete set of head observations from 1943 to 1996. To evaluate and test these new data, the prior model was used to simulate this time period. This simulation provided some good insights into the capability of the prior model to duplicate historical trends in water table changes and, in particular, groundwater mound building and decline during the entire period of Hanford operations. Results of these preliminary simulations with the prior model indicate that it was not a good predictor of the water table configuration before 1979. While the prior model was generally capable of replicating overall trends over most of the Site, its parameter estimates led to a significant over-prediction of the historical growth and decline of

groundwater mounds in the 200 West Area. Simulated heads in this area were 10 to 15 m higher than those observed in wells during the period of maximum discharges and mound building near discharge facilities during the 1950s, 1960s, and 1970s.

Recalibrating the prior model using the transient inverse calibration procedures in UCODE significantly improved its ability to simulate historical trends in water table changes over the entire Site, but some parameters took on unrealistic values, indicating that parameter zonation and/or conceptual model improvements are needed. The improved inversing methodology provided in UCODE and the additional information provided in the longer calibration period (1943 to 1996) improved the ability of the recalibrated prior model to simulate historical trends in water table changes. All goodness-of-fit measures were significantly improved over those for the previously calibrated prior model for the same simulation period. The most noteworthy improvement by the transient inverse-calibrated prior model is its ability to fit historical trends of water table changes and mound-building observed near major discharge facilities in the 200 West Area. This improvement in the overall fit resulted in improved statistical performance in all categories (mean residual, range of residuals, and sum of squared residuals) for the 1943–1996 period of calibration.

The majority of changes in the parameter estimates derived from the transient inverse calibration of the prior conceptual model using the UCODE methodology produced new estimates and linear confidence intervals consistent with prior knowledge. However, estimates for the specific yield of the Hanford Formation (from 0.06 to 0.07) and the Ringold Formation (between 0.20 and 0.21) from the three-dimensional transient inverse calibration were not consistent with current understanding of these sediments. The unrealistic estimates for these two parameters, combined with the increased estimates from the inversing for natural recharge as well as boundary fluxes for Cold Creek Valley and Rattlesnake Hills Springs, indicate that the conceptual basis of the prior model is incomplete. This initial three-dimensional transient inverse modeling study indicates that other conceptual model components not considered in the prior model are needed to approximate historical aquifer system behavior. The increased estimates of natural recharge resulting from this initial study, for example, is believed to indicate that recharge to the unconfined aquifer system of the prior model resulting from intercommunication of the unconfined aquifer system with the uppermost confined aquifer associated with the Columbia River Basalt is on the same order of magnitude as natural recharge. Thus this is the first major conceptual model to be evaluated as studies of alternative conceptual models are undertaken.

In summary, a three-dimensional transient inverse model approach for estimating site-wide scale flow and transport parameters, including their uncertainties and using data on the transient behavior of the unconfined aquifer system over the entire historical period of Hanford operations, has been developed and applied. The initial application of this new methodology considered only the conceptual basis embodied in the prior site-wide model. Only the flow model and corresponding data set were used in this three-dimensional transient inverse calibration effort; subsequent efforts will examine both flow and transport. Comparisons of the goodness-of-fit measures for the newly calibrated model with those for the prior model illustrate that the new model will strengthen the technical defensibility of the final calibrated model(s) and provide the capability to incorporate uncertainty in model predictions. These initial results, however, indicate that improvements in the conceptual model framework of the prior model are required. Studies are under way to implement and test these improvements.

Glossary

CFEST	Coupled Fluid, Energy, and Solute Transport
CHARIMA	Charriages de Rivierres Mailles (a computer code for simulating transport of river sediments)
DOE-RL	U.S. Department of Energy, Richland Operations Office
GIS	Geographic Information System
GEOFEST	Geological Finite Element Synthesis Tool
HEIS	Hanford Environmental Information System
HGWP	Hanford Groundwater Project
IGWMC	International Groundwater Modeling Center
CFEST- INV	Coupled Fluid, Energy, and Solute Transport Inverse computer code
MASS1	Modular Aquatic Simulation System 1D
MRGT	Multiple Realization Generator Task
MTT	Multicomponent Mass Transport
ONWI	Office of Nuclear Waste Isolation
PNNL	Pacific Northwest National Laboratory
PERL	Practical Extraction and Report Language–(a freeware language designed for text Manipulation)
PSPL	Puget Sound Power and Light
SGM	Site-Wide Groundwater Model
TRANSS	A Simplified Model for Radioactive Transport
UCODE	Universal Inverse Modeling Code
USGS	U. S. Geological Survey
VTT	Variable Thickness Transient

Acknowledgments

Many people contributed to the development of this report. The authors wish to express their thanks to Dr. Eileen Poeter (IGWMC), Dr. Evan Anderman (Calibra Consulting), and Dr. Mary Hill (U.S. Geological Survey) for their technical guidance in the implementation of UCODE to Hanford Site specific data and information. Special thanks go to Dr. Poeter for the development and testing of the parallel version of UCODE that was invaluable to our entire effort. The authors also wish to thank Dr. Sumant Gupta (CFEST Co.) for his technical assistance and implementation of modifications to CFEST to simplify its use with UCODE. The authors also express their thanks to Doug Hildebrand (DOE-RL) for his ongoing encouragement and support for our efforts and the useful review comments and suggestions he provided on early drafts of the report.

Thanks and appreciation also go to other PNNL staff who provided invaluable assistance in the important areas of data processing and model data input preparation and final report preparation. Will Nichols assisted in the development of historical artificial discharge database used in the calibration. Greg Guensch and Marshall Richmond generated the data used to approximate the historical river stages used to represent the Columbia River-boundary condition during the period of calibration. Mark Williams provided technical support in translating historical river stage to model boundary input files. Mary French and Jennie Harland assisted in the tedious analysis and processing the large volumes of hydraulic head and artificial discharge information used in the calibration effort. Tim Scheibe provided technical peer review and Wayne Cosby and Sheila Bennett provided technical editing support.

Contents

Appendixes

Figures

Tables

1.0 Introduction

In 1996, the U.S. Department of Energy Richland Operations Office (DOE-RL) initiated a project to consolidate multiple groundwater models at the Hanford Site into a single site-wide groundwater model (SGM). From that process, RL selected a three-dimensional groundwater flow and transport model developed by the Hanford Groundwater Project (DOE-RL 2000) as the preferred alternative for the initial phase of the SGM consolidation process. The existing consolidated site-wide groundwater model was calibrated using 1979 data and a steady state inverse approach along with trial and error transient model calibration runs using estimates of artificial discharges and a limited set of representative head observations from 1979 to 1996. It will generally be referred to in this report as the "prior model."

In the autumn of 1998, an external peer review panel was convened to conduct a technical review of the selected Hanford SGM, or prior model. The three-member review panel was asked to comment on three specific questions:

1. Are the conceptual model and technical capabilities embodied in the numerical implementation of the proposed site-wide groundwater model adequate to meet the anticipated needs, requirements, and uses for the Hanford Site?
2. If not, what refinements, modifications, or alternative conceptual models should be investigated to further improve the conceptual model and its numerical implementation to meet the anticipated Hanford Site needs, requirements, and uses?
3. Are there major conceptual model, parameter, and data uncertainties that can and should be resolved by collecting additional data for the proposed model to be adequate for Hanford Site needs, requirements, and uses?

A formal report transmitted to RL on January 14, 1999 documents the results of their review.[(a)] The panel agreed that the concept of developing a broadly applicable site-wide groundwater model was excellent. A summary of other key points of the review as they relate to this three-dimensional inverse model calibration effort follows.

1.1 Key Expert Panel Recommendations

With regard to question 2 on improvements in the modeling framework, which is relevant to the inverse modeling discussed in this report, the panel made several general comments and related recommendations that centered on a broad theme of uncertainty, which they summarized as follows:

- *The existing deterministic modeling effort has not acknowledged that the prescribed processes, physical features, initial and boundary conditions, system stresses, field data, and model parameter values are not known and cannot be known with certainty. Consequently, predictions of heads and concentrations in three dimensions over time will be uncertain as well.*

(a) S. Gorelick, C. Andrews, and J. Mercer. January 14, 1999. *Report of the Peer Review Panel on the Proposed Hanford Site-Wide Groundwater Model*. Letter report to U.S. Department of Energy Richland Operations Office, Richland, Washington.

- *A new modeling framework must be established that accepts the inherent uncertainty in model conceptual representations, inputs, and outputs. Given such a framework, the expected values of heads and concentrations, as well as the range (distribution) of predictions, would be products of the site-wide groundwater model.*
- *A priority task is to construct a comprehensive list of alternate conceptual model components and to assess each of their potential impacts on predictive uncertainty.*
- *Assessment can be initiated with hypothesis testing and sensitivity analysis within the general framework already established with the existing site-wide model. If uncertainties due to alternate conceptual models are significant, then a Monte Carlo analysis is required to estimate both the expected value of the prediction and its uncertainty.*

The transient inverse calibration effort described in this report was undertaken as part of the effort to address the panel's specific recommendation that the concept of uncertainty be acknowledged and embraced from the outset and that a new modeling framework be established that is stochastic rather than purely deterministic, with both the expected values of heads and concentrations and the range of (distribution) predictions being products of any modeling effort. A companion report (*Uncertainty Analysis Framework—Hanford Site-Wide Groundwater Flow and Transport Model*) discusses the uncertainty analysis framework being developed for the Hanford SGM that will enable the key uncertainties in model predictions of groundwater flow and transport to be quantified.

Development of the three-dimensional transient inverse modeling approach described in this report addresses the panel's specific comments and recommendations on model calibration (note the underlined portions of the panel's comments that follow). The panel's comments and recommendations on model calibration indicate that the previous calibration process and consequent estimates of hydraulic conductivity made with the prior model were not defensible for the following reasons:

1. *Parameter estimation was based on the selection of a single snapshot of hydraulic heads in 1979 that was assumed to represent steady-state conditions. Given the transient nature of areal recharge and source fluxes from disposal of wastewater, this approach is questionable. Further work should aim to justify this assumption and/or to perform a transient calibration.*
2. *The zonal parameterization of transmissivities resulted in 262 parameter values that were estimated. The data used in the inverse procedure considered 217 hydraulic heads and 52 local estimates of transmissivity. This is a clear example of over-parameterization. Resulting transmissivity estimates lead to simulated heads that match observed heads, but the predictive value of the model is low.*
3. *Hydraulic conductivities for each of the model layers were calculated based on transmissivities estimated from a 2D model of the entire unconfined aquifer. The panel believes that, in general, hydraulic conductivities in a 3D model should be estimated using a 3D inverse model. Short of 3D estimation, an assessment must be undertaken regarding the use of detailed stratigraphy and "text-book value" hydraulic conductivities as the basis for disaggregating transmissivities for a 2D unconfined aquifer into hydraulic conductivities in 3D.*
4. *The head data used in the inverse model were, in fact, not head data. Rather, they were interpolated values at model node locations. These interpolated values carry a bias. The*

parameter estimation procedure provides two pieces of information: the parameter estimates and the covariance of these estimates. When the "data" used in the inversion process are values interpolated at all nodal locations, the covariance of the parameter values is artificially reduced and the estimates are unreliable. That is, the creation of data through interpolation leads to biased estimates of model parameter values and artificial estimates of model parameter uncertainty.

5. *The panel is also concerned about the effect of using transmissivities from wells that are partially screened in the aquifer to serve as observed transmissivities for the entire thickness of the alluvial aquifer. An additional concern is the selection of weights used in the matching procedure for heads and transmissivities.*
6. *Within the framework suggested earlier, parameter uncertainty estimates are an essential part of the model and its ability to provide an expected range of predicted values. Proper parameter estimates and parameter uncertainty estimates (covariances) should be developed and used to assess the uncertainty in predicted heads and concentrations.*

The three-dimensional inverse methodology being adapted for use in future Hanford SGM model calibration and subsequent uncertainty analyses, when fully implemented, will address each of the panel's six comments and recommendations on model calibration.

1.2 Purpose of Report

In response to the external peer review recommendations,[(a)] Pacific Northwest National Laboratory (PNNL) has embarked upon a new initiative to strengthen the technical defensibility of the predictions made with a site-wide groundwater flow and transport model at Hanford (Figure 1.1). In FY 2000, the focus of the initiative was on the characterization of major uncertainties in the current conceptual model that affect model predictions. The long-term goals of the initiative are to develop and implement an uncertainty estimation methodology in future assessments and analyses using the site-wide model.

One aspect of this uncertainty analysis framework is developing a calibration approach and methodology that can be implemented to assist in the testing and evaluation of alternative conceptual models of the Hanford aquifer system. This report focuses on the development and initial implementation of the calibration approach and methods that are being used.

The overall technical approach that will be used for assessing uncertainty in future predictions of groundwater and contaminant transport at Hanford will follow closely the recommendations of the external peer-review panel.[(a)] The initial phase of activities will focus on model identification and parameter estimation that entails a broad evaluation of multiple numerical constructs of the model based on alternative interpretations of components of the conceptual model. A central part of this estimation process will involve calibrating individual numerical constructs of alternative interpretations of key components of the conceptual model to historical impacts of Hanford operations on the aquifer system. A

(a) Gorelick S, C Andrews, and J Mercer. January 14, 1999. "Report of the Peer Review Panel on the Proposed Hanford Site-Wide Groundwater Model." Letter Report to U. S. Department of Energy Richland Operations Office, Richland, Washington.

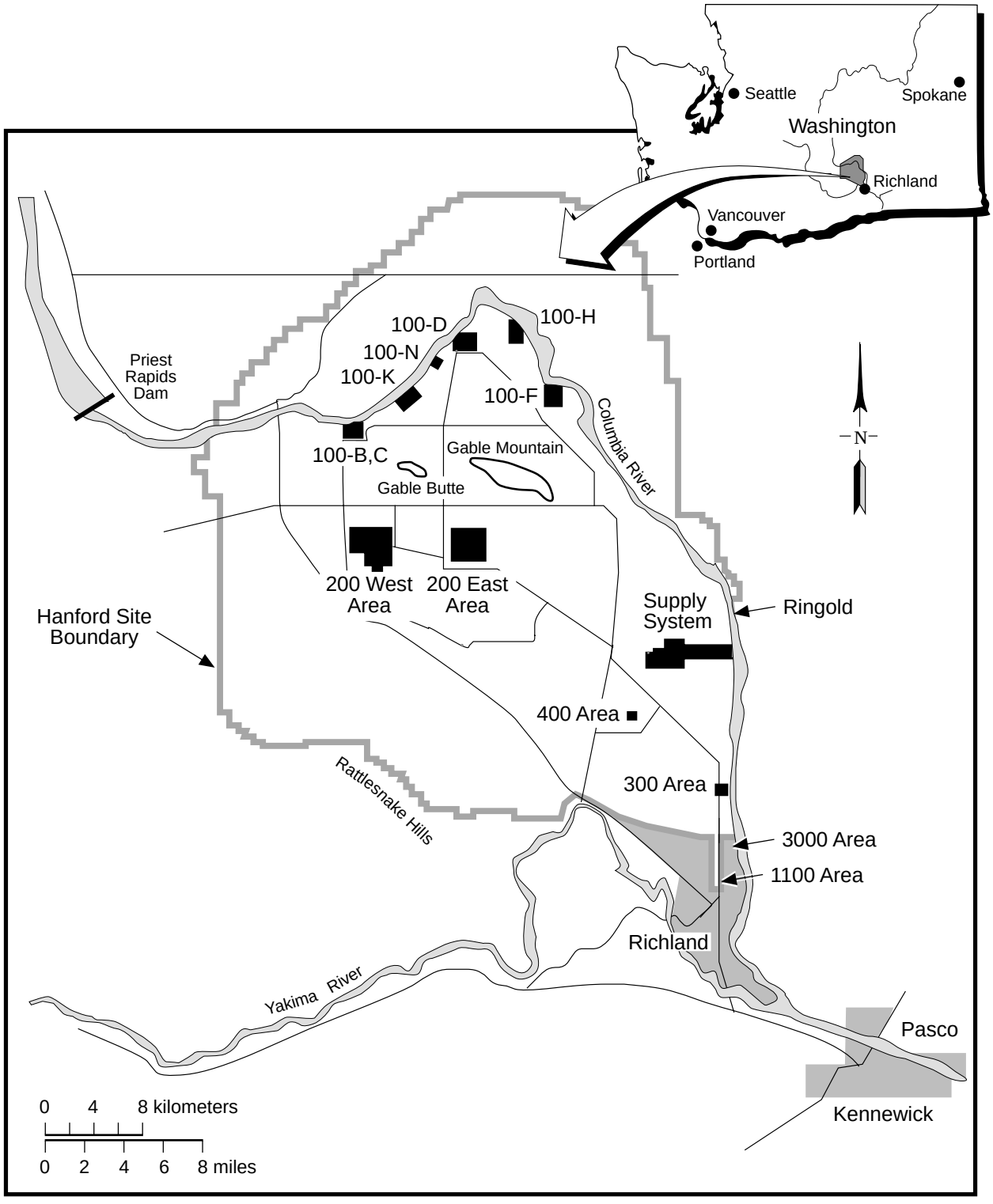

Figure 1.1. Location of the Hanford Site

more complete description of the Uncertainty Analysis Framework is provided in the companion report, *Uncertainty Analysis Framework—Hanford Site-Wide Groundwater Flow and Transport Model*.

The calibration process will involve a three-dimensional transient inverse calibration of each numerical model to historical observations of hydraulic and water-quality impacts to the unconfined aquifer system from Hanford operations since the mid-1940s. From the beginning of operations in 1943, Hanford activities discharged large volumes of wastewater to a variety of waste facilities. These operational discharges raised the water table and created groundwater mounds. They were the source of contaminant plumes along the Columbia River and in the central part of the site. Since 1988, Hanford's mission has changed from weapons production to environmental restoration. Wastewater discharges have declined significantly, which has caused the water table to decline significantly over the past decade.

Results of this initial model identification and parameter estimation phase have aided in the development of plausible alternative conceptual models. Results also provide statistical information needed to incorporate the effects of alternative interpretations of aspects of the conceptual model and associated parameter uncertainty into predictions of future groundwater flow and contaminant transport.

1.3 Scope of Report

The scope of this report is to 1) describe the development of the technical approach and methods being used to perform a three-dimensional transient inverse calibration of the Hanford SGM and 2) summarize results of the initial application of this newly developed technical approach in the recalibration of the prior model using information on the historical impacts to the aquifer between 1943 and 1996. Section 2 of this report provides general background information on the prior model, including the chronology of its development and the selection of the computer code to implement it. Section 3 summarizes the understanding of the Hanford Site aquifer system (i.e., the conceptual model) and its current numerical implementation. Section 4 presents the technical approach and methods adapted or developed to perform the transient inverse calibration. Section 5 summarizes the approach used in setting up the transient inverse recalibration of the prior site-wide conceptual model for the historical period of Hanford Operations (1943–1996). Section 6 summarizes the results of preliminary simulations of the 1943–1996 calibration period using the prior model for comparisons with results from the new approach. Section 7 discusses the results and evaluation of the transient inverse simulation of the prior conceptual model for the 1943–1996 period. Section 8 summarizes the main conclusions resulting from this initial transient inverse calibration effort. Section 9 provides the cited references.

Appendix A presents an overview of the modifications of the Coupled Fluid, Energy, and Solute Transport (CFEST)-96 code to facilitate its interaction with the UCODE universal inverse code. Appendix B contains an overview of the MasterTasker, Perl-based code[(a)] developed to manage the multiple simulations (150 to 250 individual simulations) required to perform the calibration. Appendix C describes the modeling to estimate the transient behavior of the Columbia River for use as boundary conditions in the transient inverse calibration period from 1943–1996. Appendix D provides selected plots of water table elevation and associated distribution of head residuals resulting from the preliminary simulations of the 1943–1996 period using the prior model. Appendix E shows selected plots of water table elevation and the associated distribution of head residuals resulting from preliminary simulations of the 1943–1996 period using the best-fit model from the transient inverse process.

(a) A freeware language designed for text manipulation.

2.0 Background

The model being considered in this initial calibration effort is based on an existing three-dimensional numerical model for groundwater flow in the Hanford unconfined aquifer that was developed and enhanced as part of the Hanford Groundwater Project (HGWP) (Thorne and Chamness 1992; Thorne et al. 1993; 1994; Wurstner et al. 1995; Cole et al. 1997). The three-dimensional model was developed to better understand future changes in water levels and to enhance predictions of contaminant-plume movement being monitored by the HGWP (Cole et al. 1997). Applications and developments made on the HGWP's three-dimensional site-wide model of the Hanford unconfined aquifer are routinely reported in the Hanford annual groundwater monitoring report (e.g., Hartman and Dresel 1997). The existing model, which is referred to as the prior model in this report, was calibrated using 1979 data and a two-dimensional steady-state inverse approach (Wurstner and Devary 1993; Jacobson and Freshley 1990) in conjunction with additional trial and error three-dimensional transient model calibration runs using estimates of artificial discharges and a limited set of representative head observations between 1979 and 1996 (Cole et al. 1997).

Hydraulic property data used in the development of the prior model were obtained from the results of hydraulic tests documented in Bierschenk (1959), Kipp and Mud (1973), Deju (1974), Lindberg and Bond (1979), Graham et al. (1981), DOE (1988a), Liikala and Aaberg (1988), Thorne and Newcomer (1992), Peterson (1992), Connelly et al. (1992a, 1992b), Swanson (1992), Thorne et al. (1993), Connelly (1994), and Swanson (1994). Information was also obtained from new tests and tests that were previously undocumented. Information on the subsurface geologic framework came primarily from interpreting geologic descriptions of samples acquired during well drilling. These interpretations were based on work by Lindsey et al. (1991, 1992), Lindsey (1992), Lindsey and Jaeger (1993), Lindberg (1993a, 1993b), Hartman and Lindsey (1993), and Swanson (1992) in the 100, 200, and 300 Areas of the Hanford Site, which use the lithofacies units outlined in Lindsey (1991).

Many of the wells used to define the geologic framework of the prior model were drilled to basalt as part of a study for a proposed nuclear power plant (PSPL 1982). Other information used in defining the top of basalt came from wells drilled for the Basalt Waste Isolation Project (DOE 1988a), which studied the basalts underlying the Hanford Site for disposal of high-level nuclear waste. Approximately 550 wells were used to define the three-dimensional hydrogeologic structure of the unconfined aquifer system. Many of these wells were used to determine the elevation of the top of basalt, and not all have been interpreted over their entire depth. Information on the southern part of the Hanford Site and the Richland area came from studies conducted by the U.S. Geological Survey (Ebbert et al. 1993), from Liikala (1994), and from private well logs filed with the Washington State Department of Ecology (Ecology). Information on the construction of Hanford Site wells was obtained from Chamness and Merz (1993) and the Hanford Environmental Information System (HEIS) database.

2.1 Groundwater Flow Model Selection and Chronology

The three-dimensional groundwater flow and transport model developed for the Hanford Groundwater Project is implemented numerically using the CFEST code (Gupta et al. 1987; Cole et al. 1988; Gupta 1997). The CFEST code was originally designed to support the radioactive waste repository

investigations under DOE's Civilian Radioactive Waste Management Program (Gupta et al. 1987). The chemical-waste-management community has also effectively used it for conducting exposure assessments, evaluating remediation alternatives, and designing extraction and control systems for aquifers.

2.1.1 Code Selection

Evans et al. (1988) and Wurstner et al. (1995) described the capabilities and approach used in the CFEST code and its selection for the Hanford Groundwater Project. The chronology in the continuing development of the PNNL site-wide model of the unconfined aquifer is outlined below. CFEST is an approved code for working on Hanford Federal Facility Agreement and Consent Order (Tri-Party Agreement) (Ecology 1989) milestones related to risk assessment (DOE 1991). The CFEST software library was tested extensively and brought under strict software quality assurance/quality control procedures by the Office of Nuclear Waste Isolation (ONWI) when it was developed for the Civilian Radioactive Waste Management Program. The supercomputer version (CFEST-SC), developed to run on all major UNIX workstations (Cole et al. 1988), was used for all flow and transport modeling before FY 1997. In FY 1997, the refinement of the site-wide three-dimensional model continued with its application to contaminant transport of selected contaminant plumes (Cole et al. 1997). An updated version of the CFEST code called CFEST96 (Gupta 1997) was used in this effort. This version of the code runs in a PC or UNIX environment, the direct solver has been replaced with an iterative solver, and the disk storage requirements have been reduced from the previous version of CFEST.

Results from CFEST are graphically displayed using the ARC/INFO®(a) geographic information system (GIS) and Tecplot®.(b) The ARC/INFO® GIS package is also used to store fundamental hydrogeologic data and information used to represent the three-dimensional conceptual model and to construct the three-dimensional numerical model. The three-dimensional visualization software package known as EarthVision®(c) is used to manipulate hydrogeologic data for the conceptual model.

2.1.2 Chronology of Site-Wide Groundwater Model Development

Summarizing from the chronology discussed in Wurstner et al. (1995) and Kincaid et al. (1998), a site-wide flow and transport model has been under continuous development by PNNL staff since the early 1960s as part of PNNL's continuing involvement in Hanford's groundwater monitoring efforts. The capabilities of the site-wide flow and transport model are refined and updated as additional information is gathered and as conditions and application needs change at Hanford. PNNL's Hanford Site unconfined aquifer model consists of a conceptual model and database that defines current system understanding.

Early flow models were two dimensional (e.g., the Variable Thickness Transient [VTT] code) (Kipp et al. 1972), and transport modeling, depending on application, was of the advective type (e.g., Hanford Pathline Calculation code) (Friedrichs et al. 1977), quasi-three-dimensional particle tracking type (e.g., the Multicomponent Mass Transport [MMT] code) (Alhstrom et al. 1977), or multiple streamtube type (e.g., the TRANSS code) (Simmons et al. 1986). Early flow-models were calibrated with a stream-tube

(a) ARC/INFO is a registered trademark of Environmental Systems Research Institute, Inc., Redlands, California.
(b) Tecplot is a registered trademark of Amtec Engineering, Inc., Bellevue, Washington.
(c) EarthVision is a registered trademark of Dynamic Graphics, Inc., Alameda, California.

approach that used available field measurements of transmissivity, river stage, disposal rates to ground, and head in an iterative approach to determine the Hanford unconfined aquifer transmissivity distribution (Transmissivity Iterative Calculation Routine) (Cearlock et al. 1975). Freshley and Graham (1988) describe applications of the VTT, MMT, and TRANSS codes at the Hanford Site.

In the mid 1980s, the CFEST code was selected for upgrading PNNL's two-dimensional modeling capability. CFEST has been used to model Hanford and a number of other sites in two and three dimensions (Dove et al. 1982; Cole et al. 1984; Gale et al. 1987; Foley et al. 1995). Evans et al. (1988), in a Hanford Site groundwater monitoring report for 1987, discuss selection of the CFEST code for application to modeling flow and transport in the Hanford Site's unconfined aquifer.

Initial flow modeling with the CFEST code was two-dimensional as it had been with the previous VTT code. New data were used to recalibrate the CFEST two-dimensional groundwater flow model of the Hanford Site unconfined aquifer. A steady-state finite-element inverse calibration method developed by Neuman and Yakowitz (1979) and modified by Jacobson (1985) was used in this effort. All available information on aquifer hydraulic properties (e.g., transmissivities), hydraulic heads, boundary conditions, and discharges to and withdrawals from the aquifer were included in this inverse calibration. Evans et al. (1988) described inverse calibration efforts, Jacobson and Freshley (1990) described final calibration results, and Wurstner and Devary (1993) described the calibrated two-dimensional model of the unconfined aquifer.

Two-dimensional flow models used extensively at the Hanford Site before disposal operations ceased were generally adequate for predicting aquifer head changes and directions of groundwater flow. This is because groundwater levels were somewhat stable through time across the Hanford Site. However, in the early 1990s, it was recognized that a three-dimensional model was needed to accurately calculate future aquifer head changes, directions of groundwater flow, mass transport, and predictions of contaminant concentrations. The three-dimensional model was needed because there is significant vertical heterogeneity in the unconfined aquifer, and the water table was dropping over most of the Hanford Site in response to cessation of large liquid disposals to the ground. Development of a three-dimensional model began in 1992 (Thorne and Chamness 1992) and was completed in 1995 (Wurstner et al. 1995). Thorne et al. (1994) interpreted the hydrogeology of the Hanford Site unconfined aquifer as an alternating series of transmissive units that are separated from each other in most places by less transmissive or mud units. Accounting for this vertical heterogeneity is particularly important for unconfined-aquifer predictions at the Hanford Site as future water table changes result in the dewatering of hydrogeologic layers. The water table is near the contact between the Hanford Formation and the underlying, and much less permeable, Ringold Formation over a large part of the Hanford Site. Water-level declines caused by decreased discharge at disposal facilities is causing and will continue to cause dewatering of the highly permeable Hanford Formation sediments in some areas (Wurstner and Freshley 1994). This may result in aquifer transmissivity changes of an order of magnitude or more that would not be properly accounted for by two-dimensional flow and transport models that average vertical properties at each spatial location. As a result, a two-dimensional model cannot accurately simulate changes in groundwater levels, groundwater flow direction, and contaminant transport because the three-dimensional routing of groundwater flow and contaminant mass resulting from the vertical heterogeneity cannot be properly accounted for.

The initial three-dimensional model of the Hanford Site unconfined aquifer (Wurstner et al. 1995) was calibrated in a two-step process. In the first step, the two-dimensional model was recalibrated with a

steady-state, statistical inverse method implemented with the CFEST-INV computer code (Devary 1987). The two-dimensional transmissivity distribution from this inverse modeling was preserved during the calibration of the three-dimensional model as is described in Wurstner et al. (1995).

The site-wide model was improved further during FY 1996 and FY 1997 as part of the HGWP. The purpose of this effort was to assist the HGWP in interpreting monitoring data, to investigate contaminant mass transport issues and evaluate the future movement of existing contaminant plumes, and to identify and quantify potential groundwater quality problems for onsite and offsite use. The report on this effort (Cole et al. 1997) describes the improvements to the three-dimensional model, the model recalibration, and the application of the model to predict the future transport of existing contaminant plumes in the unconfined aquifer. The Cole et al. (1997) report presents predicted changes in transient-flow conditions in the unconfined aquifer to the year 4000.

3.0 Description of the Groundwater Flow Model

This section briefly describes the conceptual and numerical basis for the consolidated site-wide groundwater model (SGM) of the Hanford unconfined aquifer system that is used as the forward model in the initial three-dimensional transient inverse calibration effort described in this report. In this section, assumptions and conceptual model components identified by the expert review panel and other reviewers (DOE-RL 2000) as needing further investigation or improvement are described. The SGM conceptual model was developed from information on the hydrogeologic structure of the aquifer, spatial distributions of hydraulic and transport properties, aquifer boundary conditions, and distribution and movement of contaminants. Development of the basic aspects of this three-dimensional conceptual model of the Hanford Site unconfined aquifer system is documented in Thorne and Chamness (1992), Thorne et al. (1993, 1994), and Wurstner et al. (1995).

3.1 Hydrogeologic Framework

The Hanford Site lies within the Pasco Basin, a structural depression that has accumulated a relatively thick sequence of fluvial, lacustrine, and glaciofluvial sediments. The geology and hydrology of the Hanford Site have been studied extensively for more than 50 years and are summarized in a number of documents (Wurstner et al. 1995 and references therein). The Pasco Basin and nearby anticlines and synclines initially developed in the underlying Columbia River Basalt Group, a sequence of continental flood basalts covering more than 160,000 km^2. Overlying the basalt within the Pasco Basin are fluvial and lacustrine sediments of the Ringold Formation and the glaciofluvial Hanford formation and pre-Missoula gravels. Together, these sedimentary deposits compose the Hanford Site unconfined aquifer system. The saturated thickness of this unconfined aquifer system is greater than 61 m in some areas but pinches out along the flanks of the basalt ridges. Depth to the groundwater ranges from less than 0.3 m near the Columbia River to more than 100 m near the 200 Areas. Groundwater in this unconfined aquifer system generally flows from recharge areas in the west to the Columbia River in the east.

The Hanford and Ringold formations can be defined as several distinct hydrogeologic units. Data from wells across the site were used to define these hydrogeologic units based on textural composition. A brief summary of each of these units, based on descriptions in Wurstner et al. (1995), is provided in Table 3.1.

3.1.1 Implementation

The lateral extent and relationships between the nine hydrogeologic units, including the subunits of the Ringold Formation, the Hanford Formation, and the pre-Missoula gravels, were defined by determining geologic contacts between these layers at as many wells as possible. These interpreted distributions and thicknesses were input to EarthVision®, which was used to construct a database for formulating the three-dimensional Hanford Site conceptual model. The resulting numerical model contains nine hydrogeologic units above the top of the underlying basalt. The Geological Finite Element Synthesis Tool (GEOFEST) described in Foley et al. (1995) was used to transfer the interpreted water table and the extent and thickness of major hydrogeologic layers to the correct format for input to CFEST to develop the regional numerical model of the Hanford Site. Bottom elevations of layers, elevation of the water table, and grids representing hydraulic conductivity zones were put into GEOFEST.

Table 3.1. Major Hydrogeologic Units Used in the Site-Wide Three-Dimensional Model

Unit Number	Hydrogeologic Unit	Lithologic Description
1	Hanford Formation and Pre-Missoula Gravels	Fluvial gravels and coarse sands
2	Palouse Soils	Fine-grained sediments and eolian silts
3	Plio-Pleistocene Unit	Buried soil horizon containing caliche and basaltic gravels
4	Upper Ringold Formation	Fine-grained fluvial/lacustrine sediments
5	Middle Ringold (Unit E)	Semi-indurated coarse-grained fluvial sediments
6	Middle Ringold (Unit C)	Fine-grained sediments with some interbedded coarse-grained sediments
7	Middle Ringold (Unit B and D)	Coarse-grained sediments
8	Lower Mud Sequence (Lower Ringold and part of Basal Ringold)	Lower blue or green clay or mud sequence
9	Basal Ringold (Unit A)	Fluvial sand and gravel
10	Columbia River Basalt	Basalt

A depiction of the surface finite-element grid and boundary conditions used in the three-dimensional flow model is illustrated in Figure 3.1. Most of the interior surface elements are regular elements that are 750 m on a side. The total number of surface elements in the three-dimensional model is 1606, and the total number of nodes is 1784. The three-dimensional model, based on this surface grid, comprises a total of 7200 elements and 8765 nodes.

3.1.2 Flow System Boundaries

The conceptual model for the existing SGM, which was used as the forward model in this initial transient inverse calibration, contains several important flow-system boundaries. The unconfined aquifer system, which extends beyond the current boundaries of the existing SGM, is recharged by groundwater by those portions of the unconfined aquifer that extend westward into the Cold Creek and Dry Creek Valleys (Figure 3.1). Additionally, the unconfined aquifer is recharged from springs and runoff that infiltrate the aquifer along the northern side of the Rattlesnake Hills and the western edge of the model (Figure 3.1). To approximate the groundwater flux entering the modeled area from these valleys and the Rattlesnake Hills, both prescribed head and prescribed flux boundary conditions were defined. For previous steady-state model-calibration runs, a prescribed head-boundary condition was used to estimate boundary fluxes entering into the aquifer system from the Cold Creek and Dry Creek Valleys as well as along the Rattlesnake Hills. These calculated fluxes, summarized in Table 3.2, provided the initial estimates of boundary fluxes used for all transient inverse calibration simulations. The expert review panel (DOE-RL 2000) in their discussion of boundary fluxes presented the following discussion:

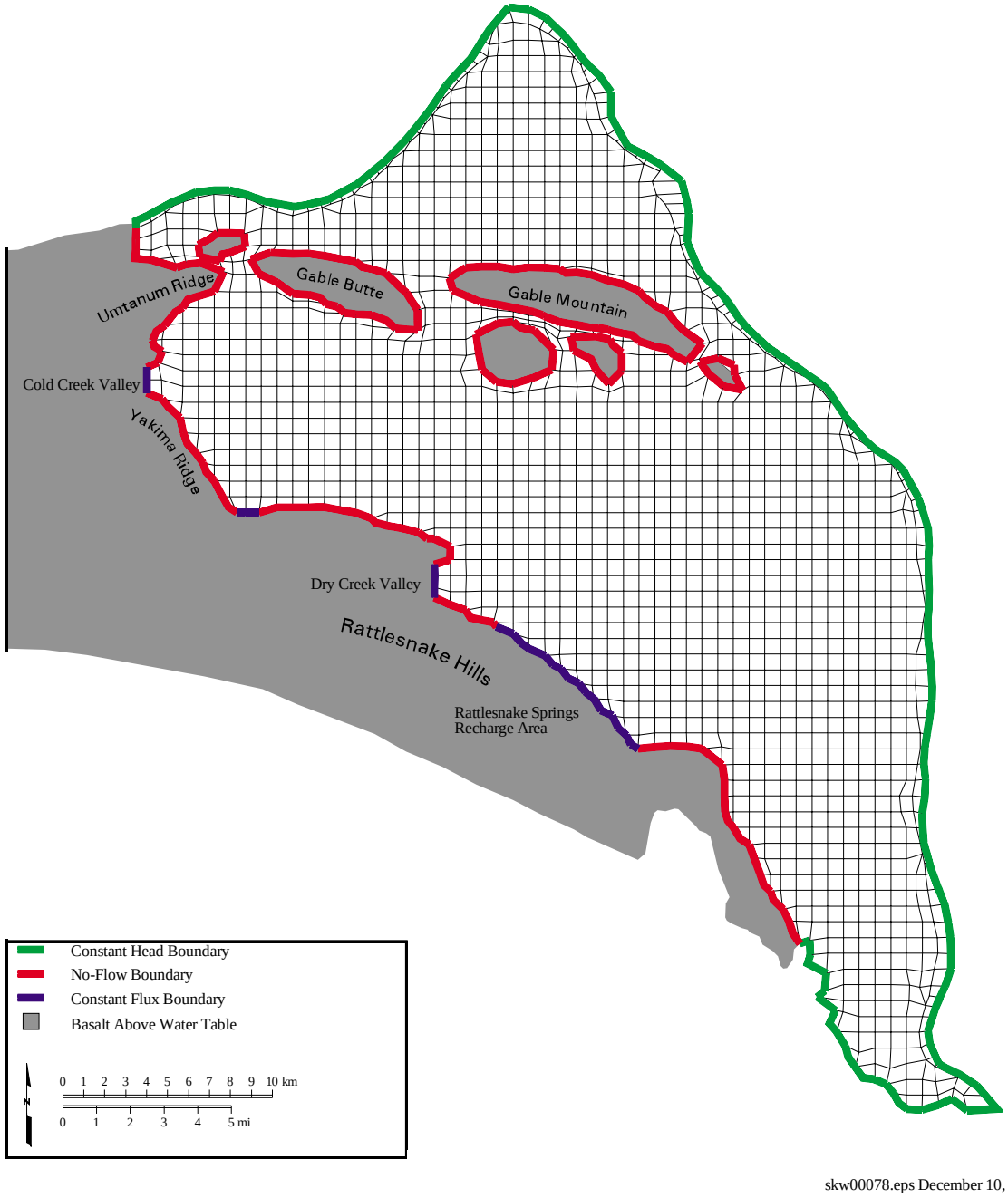

Figure 3.1. Finite-Element Grid and Boundary Conditions Used in the Three-Dimensional Flow Model

Assuming that the locations of lateral boundary fluxes are reasonable, there is an inadequate conceptual model of the existing boundary fluxes. Based on the map of recharge values used during calibration and the locations of Gable Butte and Gable Mountain, significant internal boundary fluxes apparently exist and are not considered in the active model domain. Similarly, fluxes along the western boundary are non-zero only along a small portion. Given the large drainage area in the Rattlesnake Hills and associated mountain area, some rationale must be supplied for assuming no-flow conditions, and/or those boundary fluxes must be reconsidered. Stream flow in upstream reaches of Dry Creek and Cold Creek are a likely lower boundary on underflow from these areas. A comparison of upstream stream-flow values and boundary fluxes is needed; for example, the 1997 USGS estimates of recharge from the creeks to the alluvial system are lower than values used in the calibrated model. A uniform 3D distribution of values along each flux-boundary was assumed. Some rationale for this distribution is needed, or these values must be redistributed in a less arbitrary manner. Along the western boundary it appears that boundary fluxes may in fact be leakage from Cold and Dry Creeks within the Hanford Site, in which case most of the flux should be apportioned to the upper part of the aquifer.

Table 3.2. Summary of Calculated Fluxes in Four Regions

Cold Creek Valley	2881.3 m^3/yr
Dry Creek Valley	1207.0 m^3/yr
Rattlesnake Hills	3104.4 m^3/yr
Surface natural recharge	8.47 x 10^6 m^3/yr

Addressing these issues will be part of a subsequent alternative conceptual model investigation.

The flow system is bounded by the Columbia River on the north and east and by the Yakima River and basalt ridges on the south and west, respectively. The Columbia River is assumed to represent a point of regional discharge for the unconfined aquifer system. The amount of groundwater discharging to the river is a function of the local hydraulic gradient between the groundwater elevation adjacent to the river and the river-stage elevation. This hydraulic gradient is highly variable because the river stage is affected by releases from upstream dams. To approximate the long-term effect of the Columbia River on the unconfined aquifer system in the existing SGM (i.e., the prior model), the CHARIMA river-simulation model (Walters et al. 1994) was used to generate the long-term, average river-stage elevations for the Columbia River.

In previous modeling efforts, the river itself is represented as a constant-head boundary in the uppermost nodes of the model at the approximate locations of the river's left bank and channel midpoint. Nodes representing the thickness of the aquifer below the nodes representing the mid-point of the river channel were treated as no-flow boundaries. This boundary condition is used to approximate the location of the groundwater divide that is assumed to exist beneath the Columbia River and that isolates the Hanford groundwater-river interactions from groundwater-river interactions that take place on the other side of the Columbia. The Yakima River was also represented as a specified-head boundary at surface nodes approximating its location. Like the Columbia River, nodes representing the thickness of the aquifer below the Yakima River channel were treated as no-flow boundaries, thus isolating Hanford groundwater-river interactions from those on the other side. The appropriateness of these assumptions as well as the selection of the areal extent of the domain boundaries of the Hanford Site model have been

questioned by the expert review panel and other reviewers (DOE-RL 2000). The expert review panel, in its discussion of alternative conceptual models, boundary conditions, boundary fluxes, and model implementation states that:

- *The locations and types of boundary conditions specified in 3D over time must be re-inspected. In general for large-scale applications to the Hanford site, the specified head boundary corresponding to rivers is adequate. However, the use of a specified head along the Columbia River may be inadequate for small-scale sites near the river or for short-term analyses potentially affected by the river. For example, the observed and predicted water levels for 1996 near the 100-B, C Area indicate flow directions that are at right angles to each other. In such cases, time-dependent heads and/or head-dependent fluxes should be considered. The specified head boundary along the Yakima River may be better represented by a head-dependent flux for some cases.*
- *Boundary conditions and values (e.g., inflows and their consistency with stream flow measurements, or impermeability of the lower boundary)*
- *The no-flow boundary between the basalts and the alluvial material at the base of the model may not be appropriate for areas of increased vertical permeability such as in the area northeast of the 200 East Area and in known or suspected fault areas. Further documentation of the justification for the treatment of the lower boundary throughout the domain needs to be provided. Such documentation should begin with the conceptual model and should include a water balance that accounts for flow in the basalts.*
- *The domain covered by the site-wide groundwater model must be better justified. The site-wide groundwater model simulates groundwater flow and contaminant transport only in the unconfined sedimentary aquifer in the Pasco Basin south and west of the Columbia River. The unconfined aquifer to the north and east of the river and the bedrock basalt aquifer are not represented in the site-wide groundwater model even though the major discharge area for both aquifers is the region adjacent to the Columbia River.*

These issues will be investigated in the planned alternative conceptual model investigations.

Because of the long-term nature of the calibration period, the previous design of the Columbia River boundary was modified to approximate the major annual changes in river flow that have occurred since the early 1940s. Appendix C describes the procedure followed to gather the necessary data and run the Modular Aquatic Simulation System 1D (MASS1) for the Hanford Reach of the Columbia River. Water-surface elevation boundary conditions were generated for the three-dimensional transient inverse calibration of the Hanford site-wide groundwater model in response to conditions between 1943 and 1996. The MASS1 model, developed at PNNL, is a one-dimensional, unsteady, hydrodynamic and water-quality model for branched river systems (Richmond et al 2000a, 2000b, 2000c). Appendix C also provides plots of input data, output data, and validation comparisons as well as the sources (web sites, agencies, etc.) from which data were gathered. Results of the MASS1 modeling provided the historical Columbia River stages that were averaged into six-month time steps to represent the gross annual and seasonal changes in river stage during the period. The effects of annual flow and stage fluctuations in the Yakima River during the calibration period were assumed to be less important than those of the Columbia River on the aquifer. Therefore, they were approximated using the annual average flow and stage used in previous modeling (e.g., Cole et al. 1997).

The uppermost units of basalt underlying the unconfined aquifer are assumed to represent the lower boundary for the Hanford Site unconfined aquifer system. The potential for interflow (recharge and discharge) between the basalt-confined aquifer system and the unconfined aquifer system is largely unquantified but was postulated to be very small relative to the other flow components estimated for the unconfined aquifer during the Hanford operational period. Therefore, the underlying basalt units were not included in the current three-dimensional model, and the bottom of the unconfined aquifer was treated as a no-flow boundary. The expert review panel and other reviewers (DOE-RL 2001) have indicated that the validity of this assumption is questionable, especially during the post-operational period. The expert review panel, in its discussion on improvements in the modeling framework, alternative conceptual models, and boundary fluxes, raised the following issues regarding interactions between the uppermost units of basalt and the unconfined aquifer:

- *The unconfined aquifer to the north and east of the river and the bedrock basalt aquifer are not represented in the site-wide groundwater model even though the major discharge area for both aquifers is the region adjacent to the Columbia River.*
- *The effects of larger-scale regional flow on the Hanford Site-Wide Groundwater Model domain, including flow through the basalt, flow through faults and fractures, and vertical flow through the lower boundary should be investigated.*
- *Boundary conditions and values (e.g., inflows and their consistency with stream flow measurements, or impermeability of the lower boundary)*
- *The no-flow boundary between the basalts and the alluvial material at the base of the model may not be appropriate for areas of increased vertical permeability such as in the area northeast of the 200 East Area and in known or suspected fault areas. Further documentation of the justification for the treatment of the lower boundary throughout the domain needs to be provided. Such documentation should begin with the conceptual model and should include a water balance that accounts for flow in the basalts.*

The first alternative conceptual model evaluation is examining the interactions between the unconfined aquifer system and the uppermost confined aquifer of the Columbia River Basalt system. Work on this alternative conceptual model began in FY 2000 with a study that prepared a bibliography and assembled other preliminary information on the intercommunication between the unconfined aquifer and the upper-most confined aquifer at the Hanford Site, which can be found in Appendix B of a companion report (*Uncertainty Analysis Framework—Hanford Site-Wide Groundwater Flow and Transport Model*) that discusses the uncertainty analysis framework being developed for the SGM.

3.1.3 Natural and Artificial Recharge

Natural recharge from precipitation falling on the Hanford Site is highly variable both spatially and temporally, ranging from near zero to more than 100 mm/yr, depending on climate, vegetation, and soil texture (Gee et al. 1992; Fayer and Walters 1995). Areas with shrubs and fine-textured soils like silt loams tend to have low recharge rates, while areas with little vegetation and coarse-textured soils, such as dune sands, tend to have high recharge rates. Recharge is also generally higher near the basalt ridges because of greater precipitation and runoff. Past estimates of recharge have been summarized in earlier status reports (Thorne and Chamness 1992; Thorne et al. 1993). Fayer and Walters (1995) developed a natural recharge map (Figure 3.2) for 1979 conditions to support the three-dimensional model. The

distributions of soil and vegetation types were mapped first. A recharge rate was then assigned to each combination on the basis of data from lysimeters, tracer studies, neutron probe measurements, and computer modeling. Estimated recharge rates for 1992 were found to range from 2.6 to 127 mm/yr, and the total volume of natural recharge from precipitation over the Hanford Site was estimated at 8.47×10^6 m^3/yr. This value is of the same order of magnitude as the artificial recharge to the 200-Area waste disposal facilities during 1992 and is about half the volume of discharge to these facilities during 1979 (Fayer and Walters 1995). The expert review panel, in their discussion of improvements in model implementation, alternative conceptual models, boundary fluxes, boundary conditions, and model implementation indicate that:

- *Spatial variability of recharge should be treated geostatistically to determine expected values, spatial correlation, and estimated uncertainties.*
- *Focused recharge* (should be investigated as an alternative conceptual model concept).
- *Assuming that the locations of lateral boundary fluxes are reasonable, there is an inadequate conceptual model of the existing boundary fluxes. Based on the map of recharge values used during calibration and the locations of Gable Butte and Gable Mountain, significant internal boundary fluxes apparently exist and are not considered in the active model domain. Similarly, fluxes along the western boundary are non-zero only along a small portion. Given the large drainage area in the Rattlesnake Hills and associated mountain area, some rationale must be supplied for assuming no-flow conditions, and/or those boundary fluxes must be reconsidered.*
- *Areal recharge is potentially the dominant source of water to the aquifer. The spatial distribution of recharge appears to have varied greatly in the past. As such, it is unclear how simulation of future events should represent this distributed water flux. The recharge map constructed by Fayer et al. (1996) is a good starting point to determine an average recharge map and a companion map of recharge uncertainty. Once available, this information can be used in identifying the range of model predictions (mentioned previously). In addition, the Panel recommends that experts at PNNL develop a strategy to represent the spatial distribution of recharge for a range of climatic conditions, consequent vegetation, and antecedent soil moisture conditions.*

These issues will be investigated as part of the alternative conceptual model investigations being planned.

The other source of recharge to the unconfined aquifer is artificial recharge from wastewater disposal. Over the past 50 years, the large volume of wastewater discharged to disposal facilities at the Hanford Site has significantly affected groundwater flow and contaminant transport in the unconfined aquifer (Figure 3.3). The volume of artificial recharge has decreased significantly during the past 10 years and continues to decrease. Wurstner et al. (1995) summarized the major discharge facilities incorporated in the three-dimensional model. Cole et al. (1997) summarized the major wastewater discharges from both past and future sources. There are significant uncertainties associated with the artificial discharges related to the spatial location and timing of their arrival at the water as well as the quantity that will be addressed in subsequent alternative conceptual model studies. This is an issue that has been raised by other reviewers (DOE-RL 2000).

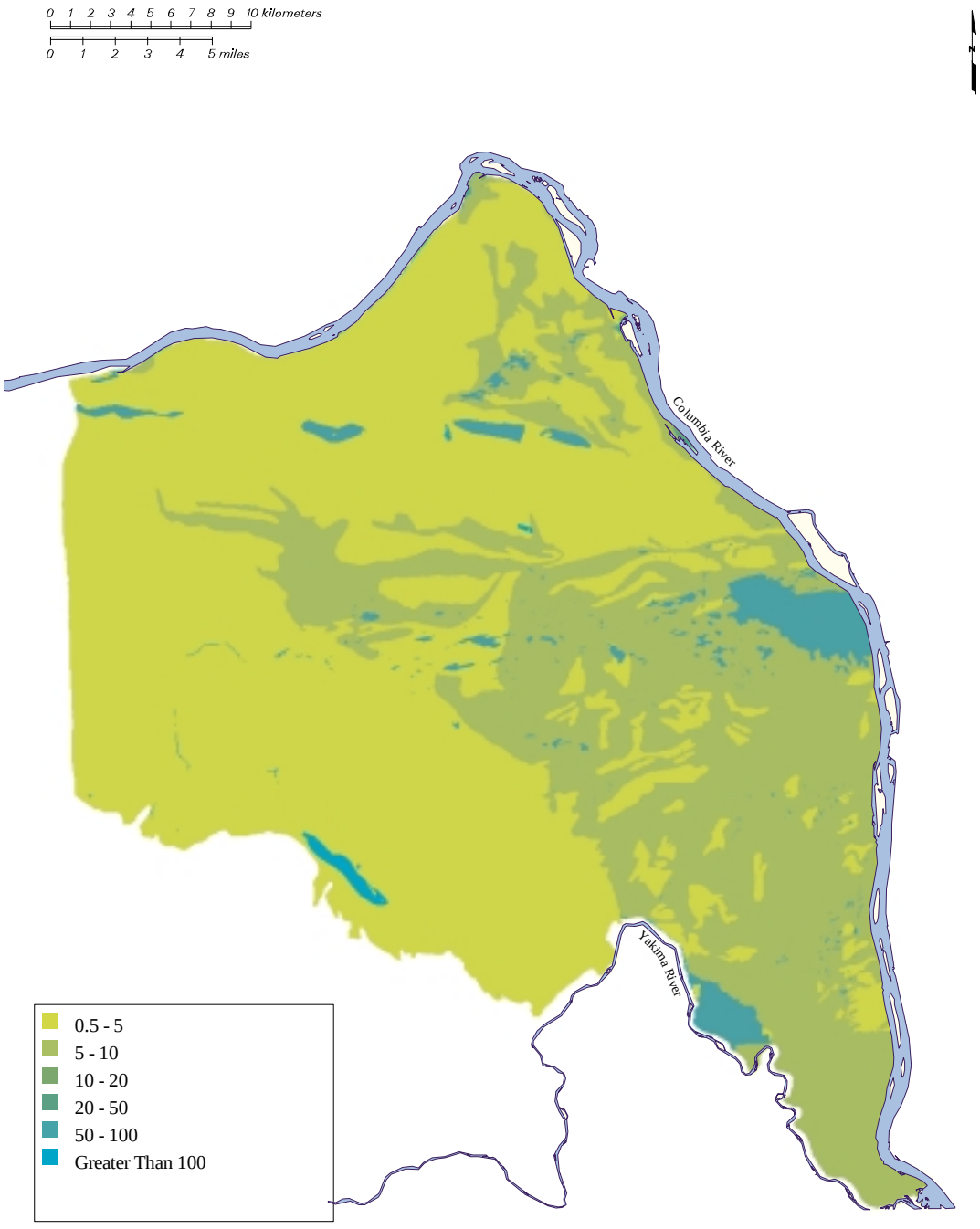

Figure 3.2. Estimates of Recharge for 1979 Conditions (Fayer and Walters 1995)

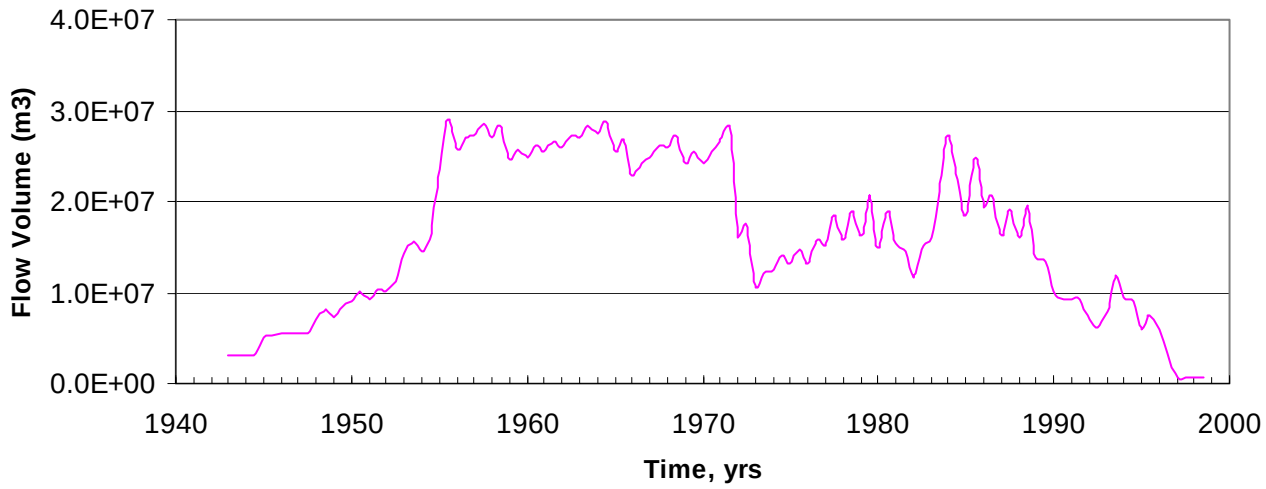

Figure 3.3. Artificial Discharges to Unconfined Aquifer from 1943–1998

3.1.4 Hydraulic Properties

Hydraulic properties important to the conceptual model include both horizontal and vertical hydraulic conductivities, storability, and specific yield. To apply a numerical model, the distribution of these parameters must be specified for each hydrogeologic unit. Hydraulic properties have been measured for the unconfined aquifer (considered as a simple hydrogeologic unit) mainly during aquifer pumping tests and from laboratory permeability tests. The results of these tests have been documented over the past 50 years and recently summarized (DOE 1988; Thorne and Newcomer 1992). As indicated in these documents, the quality of results from aquifer tests at the site is affected by both aquifer conditions and analysis procedures and varies widely. Thorne and Newcomer (1992) and Wurstner et al. (1995) reanalyzed the aquifer tests, many of which were single-well pumping tests, and they selected the set of aquifer transmissivity calibration data (Figure 3.4) used in the two-dimensional inverse model. Figure 3.5 shows the hydraulic conductivity distribution for hydrogeologic units that outcrop at the water table that are defined in the prior three-dimensional model. Figures 3.6(a,b) and 3.7(a,b) show the vertical distributions of the major hydrogeologic units and their associated hydraulic conductivities along the A-A' and B-B' cross-sections lines shown in Figure 3.4. Table 3.3 summarizes the range of hydraulic conductivities for each hydrogeologic unit.

The expert review panel discussions of improvements in model calibration indicate that with regard to hydraulic conductivity:

- *The calibration procedure for the current model is not defensible. Reasons include the insufficient justification for using a single snapshot of presumed steady-state conditions in 1979, over-parameterization of zonal transmissivities given an insufficient number of independent data, potential for incompatibility between pumping-test results and model representation of the aquifer, 2D model calibration for a 3D model, and use of interpolated head values.*

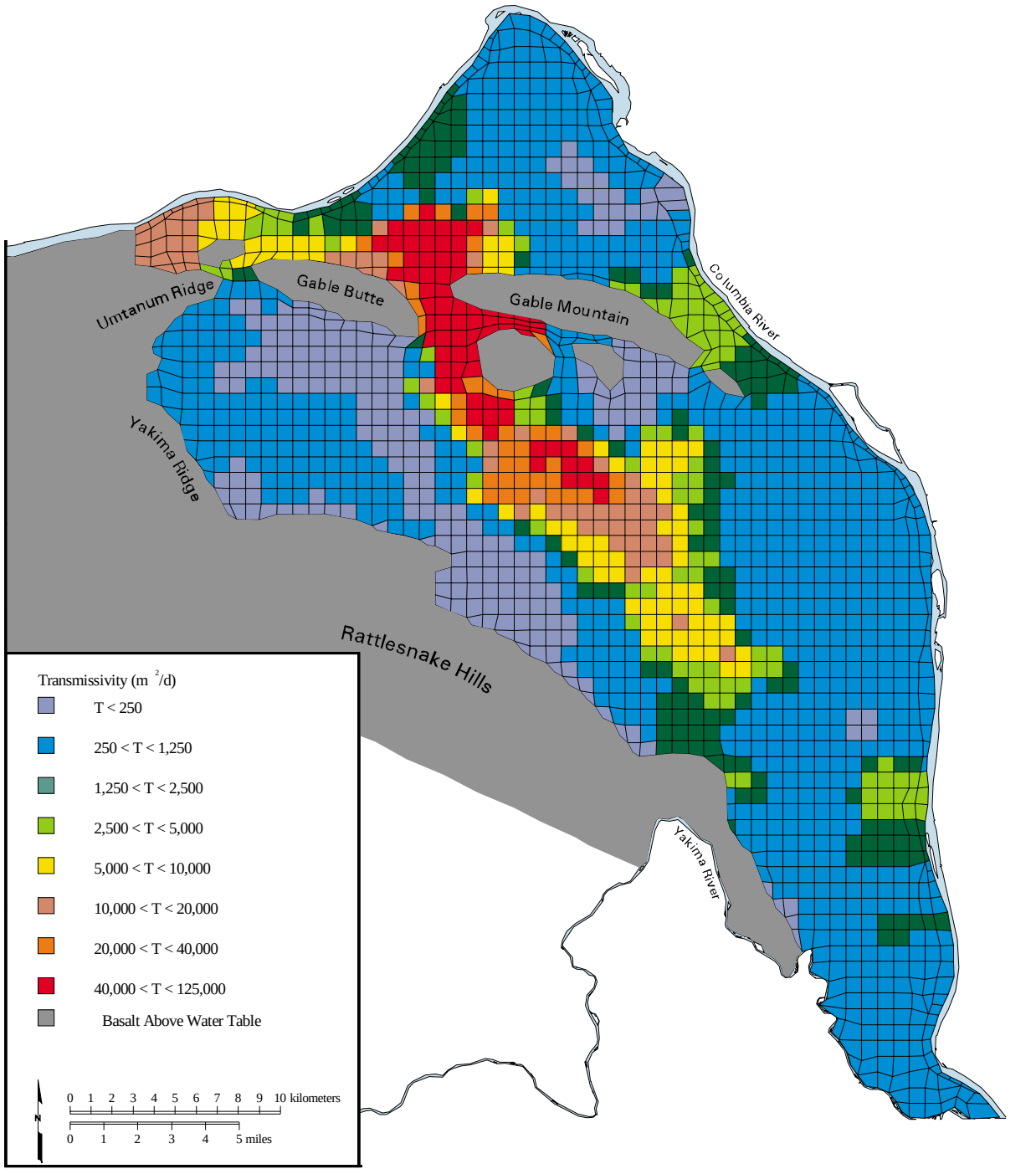

Figure 3.4. Transmissivity Data Used in Calibration

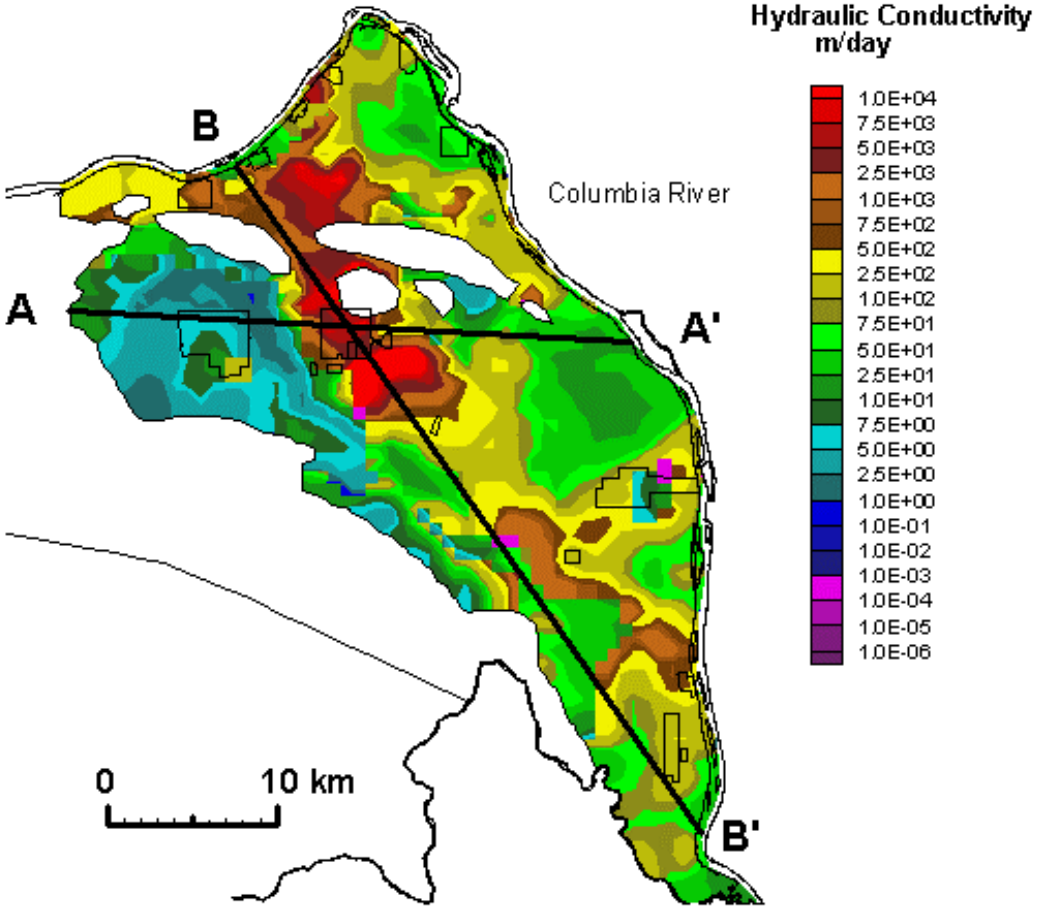

Figure 3.5. Hydraulic Conductivity Distribution

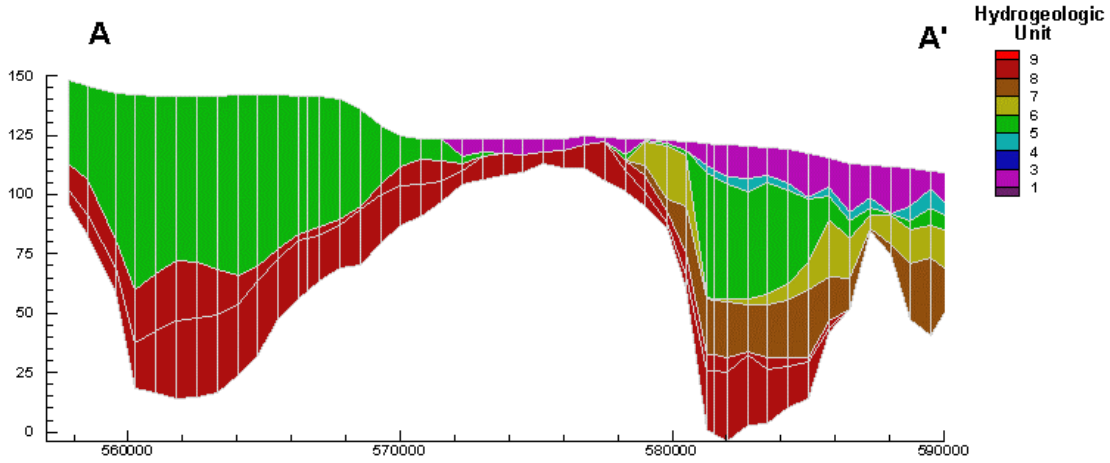

Figure 3.6a. Cross-Section A-A' Showing Distribution of Hydrogeologic Units

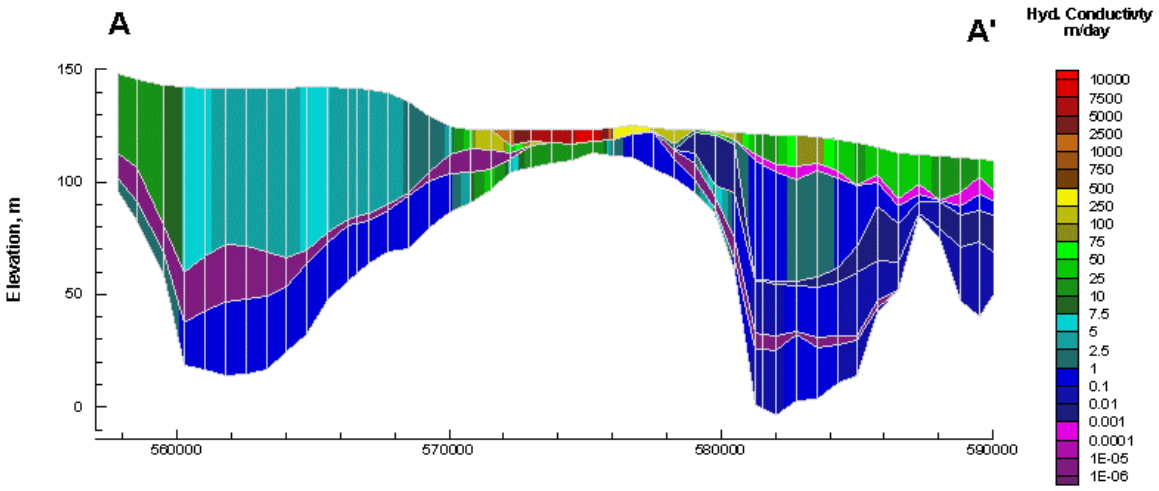

Figure 3.6b. Cross-Section A-A' Showing Distribution of Hydraulic Conductivity

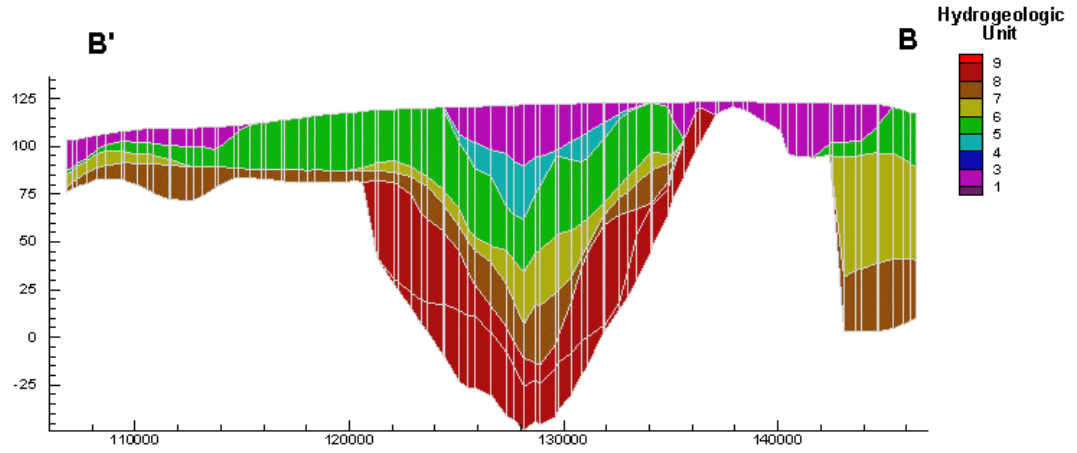

Figure 3.7a. Cross-Section B-B' Showing Distribution of Hydrogeologic Units

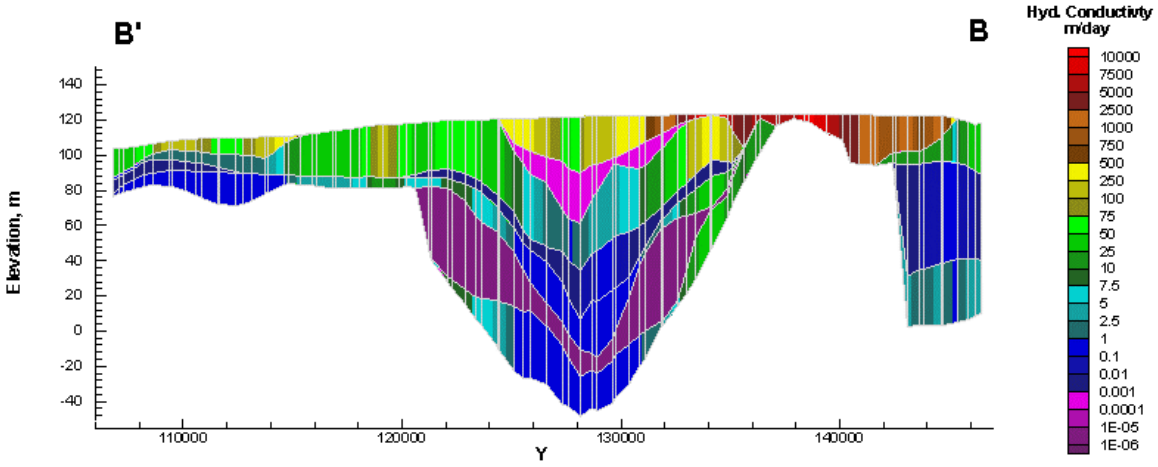

Figure 3.7b. Cross-Section B-B' Showing Distribution of Hydraulic Conductivity

- *The panel is also concerned about the effect of using transmissivities from wells that are partially screened in the aquifer to serve as observed transmissivities for the entire thickness of the alluvial aquifer.*

With regard to effective porosity and specific yield the panel states:

Although the values used for effective porosity and specific yield may sometimes be similar for a given aquifer material, there is no physical justification to base effective porosity values on measured specific yield values. There is considerable ambiguity in the literature regarding the term effective porosity. For purposes of the SGM, effective porosity is the quantity by which the seepage velocity must be multiplied to obtain the Darcy velocity. The seepage velocity is the average speed that water travels between two points due to advection. Specific yield is the drainable porosity, i.e., the volume of water that can be drained by gravity from a unit volume

Table 3.3. Range of Hydraulic Conductivities by Hydrogeologic Unit in Prior Model

	Range in Hydraulic Conductivity (m/day	
Hydrogeologic Unit	**Minimum**	**Maximum**
1	1.77	33122
3	14.2	190.5
4	5 e 10-3	5 e 10-3
5	0.04	1696.5
6	0.01	0.1
7	0.008	85.7
8	1 e 10-5	1 e 10-5
9	0.008	210.8

of initially saturated porous medium. In general, specific yield represents a much smaller fraction of total porosity than does effective porosity. Effective porosity values must be estimated, and the impact of their uncertainties must be assessed.

With regard to storage coefficient the panel indicates:

The error introduced by using wrong storage coefficient values may be responsible for some predictive errors. For example, hydrographs for Areas 5, 6, 7, 8, and 9 show an observed pulse of water. This pulse propagates through the subsurface faster and with a higher amplitude than does the simulated pulse of water. This comparison suggests that the storage parameter used in the simulation may be too high, or the hydraulic conductivity may be too small as the rate of propagation of the pulse is related to the ratio of hydraulic conductivity to the storage coefficient.

Some of these issues have been addressed as part of this initial three-dimensional inverse study and will be discussed later; others will be investigated as part of the planned alternative conceptual model investigations. The effective porosity issue will be addressed as alternative transport models are examined to address the expert panel concern regarding the potential importance of diffusive mass-transfer at the Hanford Site. Regarding this issue, the expert review panel states that:

It is noted that in almost all applications of groundwater transport models the simulated plume of a contaminant exhibits much less tailing (late arrival of mass) than is observed in the field. There are a number of processes that can explain the observed tailing, but in many instances the dominant process is diffusive mass-transfer from an immobile domain to a mobile domain. In alluvial sedimentary groundwater systems, the immobile domain may well correspond to zones of lower hydraulic conductivity, such as silt or clay lenses, within an aquifer unit. Experience suggests that, in any situation in which the effective porosity is significantly smaller than the total porosity, transfer to and from an immobile domain likely is important. In these cases, the immobile domain can be thought of as a functionally stagnant volume of water corresponding to the difference between the total porosity and the effective porosity.

The panel believes that tailing of contaminant plumes is likely to be significant in the unconfined aquifer at the Hanford Site. Therefore, the SGM will overestimate the rate at which contaminant plumes migrate and dissipate after a source has been removed because diffusive mass-transfer to and from immobile domains is not considered. The panel recommends that diffusive mass-transfer be addressed by modifying CFEST-96 to permit the option of including a mobile-immobile domain formulation.

4.0 Inverse Methodology and Computation Codes

This section presents the basic concepts used in the universal inverse modeling code UCODE (Poeter and Hill 1998), the inverse methodology selected for application to the Hanford Site SGM. Included are discussions of the objective function and the modified Gauss-Newton method used in UCODE to perform the nonlinear regression. Hill (1998) provides a full discussion of all of the other aspects of the UCODE and MODFLOW inverse model implementation. Much of the basic regression theory implemented is discussed in Cooley and Naff (1990). Other programs used by UCODE to test weighted residuals and calculate linear confidence and prediction intervals are discussed in Hill (1994).

This section also contains a brief discussion of the operational aspects UCODE, the recent enhancement to UCODE that allows parallel computation of the sensitivity coefficients through a parallel task manager MasterTasker (Appendix B) and the recent enhancement to CFEST (the forward model code used to model the Hanford Site unconfined aquifer system), which involved developing the CFUCODE and LP3UCODE modules that were developed to work directly with the enhanced version of UCODE.

4.1 Inverse Methodology

4.1.1 Inverse Modeling and the Objective Function

UCODE (Poeter and Hill 1998) performs universal inverse modeling using an indirect rather than direct approach in which the unknown parameters for the problem being solved are considered to be the dependent variables (Peck et al. 1988). In the indirect approach used by UCODE, the normal model equations (referred to as the forward equations) are solved and parameter estimates are sought that minimize a set of residuals (e.g., differences between observed and model predicted quantities). For the indirect groundwater flow inverse, the forward model can be any code that solves the standard groundwater flow equations for the dependent variable, hydraulic head. The UCODE concept of universal comes from the fact that UCODE solves the inverse problem through an indirect approach that is not directly linked or tied to any particular forward model and implementing computer code. MODFLOWP (Hill 1992), for example, is a specific inverse model that only applies to quasi-three-dimensional groundwater flow problems.

The nonlinear regression methodology implemented in UCODE and used in the initial three-dimensional transient inverse modeling discussed in this report is principally the same as that used in MODFLOWP, which was derived from Cooley and Naff (1990) (see Figure 4.1) and is the same as that discussed in the recommended guidelines for model calibration (Hill 1998). Other programs that were developed for MODFLOWP to test weighted residuals and calculate linear confidence and prediction intervals are discussed in Hill (1994) and are also implemented in UCODE with minor modifications.

The inverse problem of groundwater flow consists of estimating the vector of flow parameters, $\underline{b}$, being determined based on a number, ND, of field observations of dependent variables, y_i, (e.g., hydraulic head and flux to streams) and independent information about the number, NPR, of values for the parameters, P_p, themselves (i.e., prior information). If concentration and travel time observations/ estimates as well as information on prior estimates for additional transport parameters (e.g., dispersivity) are also

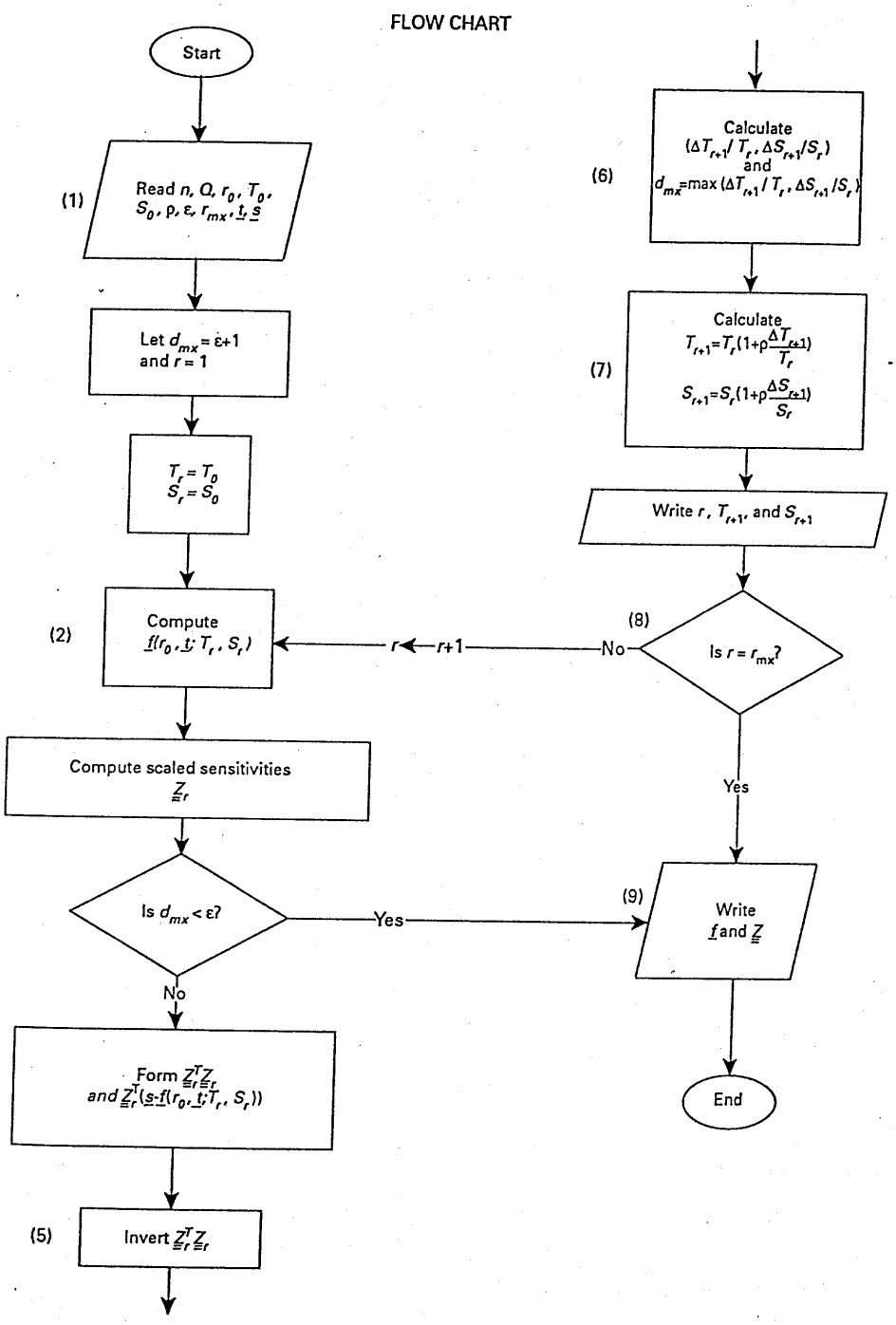

Figure 4.1. Flowchart Showing Linear Regression Algorithm (adopted from Cooley and Naff 1990, p. 65)

available, they can be added as additional observations and prior information, and a coupled inverse problem involving both flow and transport can be posed.

An objective function is a measure of the fit between simulated values and the observations that are being matched by the regression. The purpose of the regression is to calculate values of defined parameters that minimize the objective function. The resulting values are said to be "optimal," "optimized," or "estimated by regression." In UCODE, the objective function used in the nonlinear regression to determine the vector of unknown parameters, $\underline{b}$, being estimated is a weighted, ω, least squares function, $S(\underline{b})$ (e.g., the sum of the weighted square of differences between simulated, y_i, and observed values, y'_I). The regression consists of calculating the vector of parameter values, $\underline{b}$, that minimize the objective function, $S(\underline{b})$. The weighted least-squares objective function, $S(\underline{b})$, used in UCODE is

$$S(\underline{b}) = \Sigma^{ND}\ \omega_i [(y_i - y_i'(\underline{b})]^2 \ + \ \Sigma^{NPR}\ \omega_p [P_p - P'_p(\underline{b})]^2 \qquad (4.1)$$

where $y'_i(\underline{b})$ is the value simulated with the forward model that corresponds to the i-th observation (a function of $\underline{b}$); ω_i is the weight for the i-th observation; $P'_p(\underline{b})$ is the value simulated with the forward model that corresponds to the i-th observation (a function of $\underline{b}$), and ω_p is the weight for the p-th prior estimate.

As discussed in Hill (1998), the simulated values related to the observations are of the form $y'(\underline{b}) = f(\underline{b},\xi_i)$, where ξ_i are independent variables such as location and time. Additionally, the function may be nonlinear in $\underline{b}$ and ξ_i and for complex groundwater problems, like the Hanford SGM, a numerical solution is commonly required so that the function, f(), is actually a numerical model of flow (and transport) in three dimensions.

In UCODE, simulated values of prior parameter estimates, $P'_p(\underline{b})$, are restricted to be linear combinations of the components of the actual vector, $\underline{b}$, of parameters being estimated as shown in Equation 4.2.

$$P'_p(\underline{b}) = \Sigma\ a_{pj}\, b_j \qquad (4.2)$$

where a_{pj} are the constants that indicate what fraction of estimated parameter, b_j, is required to estimate the p-th simulated prior parameter estimate, $P'_p(\underline{b})$. As discussed in Hill (1998), most prior information equations have only one term with a coefficient equal to 1.0, so the contribution to the objective function is simply the prior information value of a parameter minus its estimated value. More terms are needed when the available prior information is related to more than one parameter value being estimated. In these cases, as discussed, it is restricted to be a linear combination of the estimated parameters (Equation 4.2). Hill (1998) discusses two examples. In the first example, additional terms are included when a groundwater inverse model estimates seasonal recharge rates, but the available prior information is related to annual recharge rates, so the P_p value equals the seasonal recharge rate, and the summation is a linear combination of the estimated seasonal recharge values that represents the annual recharge estimate that corresponds to the prior information. In the second example, the storage coefficient is being estimated separately for each of the two model layers, but the aquifer test conducted only provided an estimate for the combined storage coefficient. In this case, the P_p value equals the storage coefficient from the aquifer test, and an appropriate linear combination of the estimated layer storage coefficients is devised to represent the combined storage coefficient available as prior information.

The differences $[y_i - y'_i(\underline{b})]$ between the simulated and observed values (heads for the initial inverse efforts discussed in this report) and the differences $[P_p - P'_p(b)]$ between the prior parameter values and the new estimates are termed residuals, while $\omega_i^{1/2}$ $[y'_i - y'_i(b)]$ and $\omega_p^{1/2}$ $[P_p - P'_p(b)]$ are the weighted residuals and represent the fit of the regression in the context of how the residuals are weighted. The objective function increases linearly with the square of head residuals and the square of the differences of the parameter values from their prior estimates (this is why the second term in Equation 4.1 is referred to as a *penalty* function). It should be noted that prior information was not invoked in the initial inverse effort discussed in this report. From a simplistic point of view, these weights (ω) help in the fit of the regression; where information exists on measurement errors, correlations between heads, and covariance (i.e., variances and correlations) between parameters, the weights (ω) represent the inverse of the covariance matrices of heads (fluxes, and concentrations) and parameters.

In the objective function shown in Equation 4.1, a simple diagonal weight matrix was used to allow the equation to be written using summations instead of matrix notation. More generally, the weighting requires a full-weight matrix, $\boldsymbol{\omega}$, and Equation 4.1 can be written in matrix notation as

$$S(\underline{b}) = [\underline{y} - \underline{y}'(\underline{b})]^T \boldsymbol{\omega} [\underline{y} - \underline{y}'(\underline{b})] = \underline{e}^T \boldsymbol{\omega} \underline{e} \qquad (4.3)$$

where the vectors of observations, $\underline{y}$, and simulated values, $\underline{y}'(\underline{b})$, include terms for both the observations and the prior information, and $\underline{e}$ is a vector of residuals. While MODFLOWP allows use of a full-weight matrix for most types of observations and prior information, the weight matrix, $\boldsymbol{\omega}$, is considered a diagonal matrix in UCODE. Consequently, the resulting diagonal of weights in UCODE consists of $\underline{\omega}^{1/2}$.

The weighted least squares statistical framework used in UCODE is one of the three most widely used statistical frameworks for parameter estimation, which Peck et al. (1988) identify as

- Weighted Least Squares (Cooley 1977) used in UCODE (Hill 1998)
- Bayesian (Neuman and Yakowitz 1979)
- Maximum Likelihood (Carrera and Neuman 1986).

Regarding the Bayesian approach, Peck et al. (1988) indicate that a rigorous application of the Bayesian estimation criterion is not possible because the relationship between flow parameters and hydraulic head is not linear. Of the two other applicable and widely used alternatives for providing the required statistical framework, Hill (1998) argues that, in practical application, the maximum-likelihood objective function reduces to the weighted least-squares objective function. The following discussions summarized from Beck and Arnold (1977) that compare and contrast ordinary least squares (LS), maximum likelihood (ML), and other parameter estimation methods support Hill's argument. The discussion is presented because it illustrates that statistical assumptions regarding measurement errors are required to make statistical statements regarding the estimated parameters and as a result there is a need to carefully examine the nature of measurement errors as part of any estimation activity in order to be able to choose the most appropriate assumptions regarding measurement error so that the appropriate estimator can be chosen. The eight standard statistical assumptions (Beck and Arnold 1977), some of which are required to make statistical statements regarding the estimated parameters regardless of the approach, include

1. measurement errors are additive
2. observation errors have a zero mean

3. there are constant variance errors, homoskedasticity
4. uncorrelated errors exist
5. normality, i.e., the error vector, has a normal distribution
6. statistical parameters are known (covariance matrix of the observation errors completely known)
7. there are no errors in the independent variables (i.e., nonstochastic independent variables)
8. there is a constant parameter vector and no prior information.

Beck and Arnold (1977) indicate that, when all eight of the standard assumptions are valid (a condition which they denote as 11111111), the estimation problem of determining the "best" estimate values and their variance is less difficult. They indicate that a major difference in the LS and ML method is related to which of the eight standard statistical assumptions must be assumed to make statistical statements regarding the estimated parameters. While none of the standard statistical assumptions need to be valid to obtain parameter estimates using LS, some of these standard assumptions are required to make statistical statements regarding these estimates. Four of the standard assumptions (e.g., condition 11----11), for example, are required to show that the LS expected value estimator for the mean is unbiased. Additional standard assumptions of uncorrelated and constant variance measurement errors (i.e., covariance matrix of the observation errors is a diagonal matrix) are required to show that the ordinary least squares provides the minimum variance estimator (i.e., condition 1111--11). In contrast to LS, Beck and Arnold (1977) indicate that ML requires a great deal of assumptions regarding the measurement errors, including the assumption that the measurement errors have a normal probability density (i.e., condition 11--1111). However, for ML, correlated errors may be present and errors may have non-constant variance. When the third standard assumption, which assumes constant variance errors, is valid, then the covariance matrix of the observation errors, $\boldsymbol{\psi}$, must be a diagonal matrix, which means that $\boldsymbol{\psi} = \sigma^2 \mathbf{I}$. Under this additional assumption, the ML and LS estimators are exactly the same. This can be deduced by simply comparing Equations 4.4 and 4.5, which are Beck and Arnold's Equation 6.2.5 for the LS estimator ($\mathbf{b}_{LS}$) and 6.5.2 for the ML estimator ($\mathbf{b}_{ML}$) without defining all of the terms. As can be observed, the only difference is the presence of $\boldsymbol{\psi}^{-1}$, which is the identity matrix when this additional assumption is valid (i.e., $\boldsymbol{\psi} = \sigma^2 \mathbf{I}$).

$$\mathbf{b}_{LS} = (\mathbf{X}^T\mathbf{X})^{-1}\,\mathbf{X}^T\mathbf{Y} \tag{4.4}$$

$$\mathbf{b}_{ML} = (\mathbf{X}^T\boldsymbol{\psi}^{-1}\mathbf{X})^{-1}\,\mathbf{X}^T\boldsymbol{\psi}^{-1}\mathbf{Y} \tag{4.5}$$

In fact, Peck et al. (1988) indicate that the weighted LS approach is a special case of both maximum likelihood and Bayesian estimation procedures. The ML versus LS argument is more a philosophical argument since both methods will arrive the same place if the assumptions regarding measurement errors are appropriate. The main point is that measurement errors need to be carefully examined as part of each estimation activity in order to identify which of the standard assumptions are the most realistic since different assumptions lead to better and potentially different estimators.

While UCODE uses a weighted least squares approach the maximum-likelihood objective function is calculated and printed by UCODE because it can be used as a measure of model fit (Carrera and Neuman 1986; Loaiciga and Marino 1986). The value of the maximum-likelihood objective function, $S'(\underline{b})$, is calculated by UCODE as

$$S'(\underline{b}) = (ND+NPR)\ln 2\pi - \ln|\underline{\omega}| + (\underline{y} - \underline{y}')^T\underline{\omega}(\underline{y} - \underline{y}') \tag{4.6}$$

where $|\underline{\omega}|$ is the determinant of the weight matrix, and it is assumed that the common error variance equals 1.

4.1.2 Modified Gauss-Newton Optimization

As discussed in Hill (1998), the Gauss-Newton optimization method is an iterative form of standard linear regression that works well only when the regression is improved by adding a damping parameter and a Marquardt parameter (Marquardt 1963). The modified Gauss-Newton method presented in Hill (1998) closely follows that of Cooley and Naff (1990, Ch. 3). Hill notes that the modified Gauss-Newton method used in UCODE and MODFLOWP could also be termed a Levenberg-Marquardt method. There is some controversy in the name, Beck and Arnold (1977) indicate that a great many of algorithms have been proposed to improve the Gauss-Newton method (also know as Gauss, Newton-Gauss, or simply as the linearization method). Some of these are termed modifications and some of these others would call distinctly different methods. The Levenberg and Marquardt methods are in this latter category, but Beck and Arnold (1977) choose to treat them as modifications.

Parameter values that minimize the objective function are calculated using normal equations (4.7). One difference between linear regression and nonlinear regression is that, in linear regression, parameter values are estimated by solving the normal equations once, while nonlinear regression is iterative in that a sequence of parameter updates is calculated, solving the linearized normal equations once for each update. Thus, in nonlinear regression, there are parameter-estimation iterations. The normal equations and the iterative process in the modified Gauss-Newton optimization used in UCODE is as follows:

$$(\mathbf{C}^T \mathbf{X}^T_r \boldsymbol{\omega}\ \mathbf{X}_r \mathbf{C} + \mathbf{I}\, m_r)\mathbf{C}^{-1}\, \underline{d}_r = \mathbf{C}^T \mathbf{X}^T_r \boldsymbol{\omega}_h\, [\underline{y} - \underline{y}'(\underline{b}_r)] \tag{4.7}$$

$$\underline{b}_{r+1} = \rho_r \underline{d}_r + \underline{b}_r \tag{4.8}$$

where

r	=	parameter-estimation iteration number
$\mathbf{X}_r$	=	sensitivity matrix evaluated at parameter estimates $\underline{b}_r$, with elements equal to $\partial y'_i/\partial b_j$ (using forward or central difference numerical methods in UCODE)
$\boldsymbol{\omega}$	=	weight matrix (diagonal)
$(\mathbf{X}^T \boldsymbol{\omega} \mathbf{X})$	=	symmetric square matrix of dimension NP by NP (an estimate of the Fisher information matrix) used to calculate statistics
$\mathbf{C}$	=	diagonal scaling matrix with elements c_{jj} equal to $[(\mathbf{X}^T \boldsymbol{\omega} \mathbf{X})_{jj}]^{-1/2}$, which produces a scaled matrix with the smallest possible condition number (Forsythe and Strauss 1955; Hill 1998)
$\underline{d}_r$	=	vector with the number of elements equal to the number of estimated parameters used in Equation 4.8 to update the vector, $\underline{b}$, of parameter estimates for the next iteration.
$\mathbf{I}$	=	identity matrix with dimension NP by NP
ρ_r	=	a damping parameter
m_r	=	the Marquardt parameter (Marquardt 1963)

As Hill (1998) indicates, the Marquardt parameter, m_r, is used to improve regression performance for ill-posed problems (Theil 1963; Seber and Wild 1989). Beck and Arnold (1977) indicate that Levenberg (1944) tried to overcome the instability of "overshooting" in the Gauss-Newton method by introducing a

$\lambda \boldsymbol{\Omega}$ term that is in the same position as $\mathbf{I}\, m_r$ in Equation 4.7. Levenberg investigated values for the diagonal matrix $\boldsymbol{\Omega}$, including the diagonal matrix $\mathbf{I}$ and various approaches for determining values for λ, including having it be constant and allowing it to vary as the minimum solution is approached. This is because the basis of the method is to improve convergence when $(\mathbf{C}^T \mathbf{X}^T_r \boldsymbol{\omega} \mathbf{X}_r \mathbf{C})$ is poorly conditioned at the starting parameter vector and let it revert to strict Gauss-Newton as the minimum is approached. This is accomplished in Levenberg and Marquardt methods by selecting values for λ so that the $\lambda \boldsymbol{\Omega}$ term is large in comparison to $(\mathbf{C}^T \mathbf{X}^T_r \boldsymbol{\omega} \mathbf{X}_r \mathbf{C})$ early in the process and the parameter correction is close to that of the "method of steepest descent." Later, as the minimum solution is approached, the method must select values for λ such that the $\lambda \boldsymbol{\Omega}$ term is small compared with $(\mathbf{C}^T \mathbf{X}^T_r \boldsymbol{\omega} \mathbf{X}_r \mathbf{C})$ and the final iterations are close to the Gauss-Newton method. While Levenberg's methods, according to Beck and Arnold (1977), removed instability and reduced oscillations, they increased the number of iterations considerably.

In the UCODE implementation of the Marquardt method, Hill (1998) indicates that initially $m_r = 0$ for each parameter estimation iteration r. For iterations in which the vector $\underline{d}$ defines parameter changes that are unlikely to reduce the value of the objective function (as determined using the condition described by Cooley and Naff 1990, pp. 71–72), m_r is increased according to $m_r^{new} = 1.5\, m_r^{old} + 0.001$ until the condition is no longer met.

An additional damping parameter, ρ_r, is used in UCODE that can vary in value from 0.0 to 1.0 and the damping parameter modifies all values in the parameter change vector $\underline{d}_r$ by the same factor to preserve the direction of vector $\underline{d}_r$. The damping parameter, ρ_r, is used for two reasons: 1) to ensure that the absolute values of fractional parameter value changes are all less than a value specified by the user (i.e., the MAX-CHANGE value input by the user in the ".uni" file of UCODE) and 2) to damp oscillations that occur when elements in d_r and d_{r-1} define opposite directions (Cooley 1993). Fractional parameter value changes are calculated for each parameter as

$$(b_j^{r+1} - b_j^r) / |b_j^r| = d_j^r / |b_j^r| \quad j=1,NP \tag{4.9}$$

where b_j^r is the j-th element of vector $\underline{b}_r$, that is, the value of the jth parameter at parameter estimation iteration r, and b_j^r+1 is calculated with $\rho_r=1.0$ in Equation 4.8. If the largest absolute value of the NP values of Equation 4.9 is greater than MAX-CHANGE, ρ_r is calculated in many circumstances as

$$\rho_r = \text{MAX-CHANGE} / [\,|d_i^r| / |b_i^r|\,] \tag{4.10}$$

where the i for ρ_r is the parameter i for which Equation 4.9 has the largest absolute value.

In UCODE, as discussed in Hill (1998), the parameters in vector $\underline{b}$ of Equation 4.1 can be native values or the log-transform of the native values because log-transforming parameters can produce an inverse problem that converges more easily and prevents the actual parameter values from becoming negative (Carrera and Neuman 1986).

Hill (1998) indicates that upper and lower limits on parameters that constrain possible estimated values are commonly available in inverse models (for example, PEST) (Doherty 1994) and are suggested by, for example, Sun (1994, p. 35). While such limiting constraints on parameter values may, at first, appear to be necessary given the unrealistic parameter values that can be estimated through inverse

modeling, Hill et al. (1998), using a complex synthetic test case, demonstrate that this practice can disguise more fundamental modeling errors. Poeter and Hill (1996) use a simple synthetic test case to further demonstrate the concept, and in Anderman et al. (1996), unrealistic optimized values of recharge in a field problem revealed important model construction inaccuracies. "Guideline 5: Use Prior Information Carefully," of Hill (1998) indicates that unrealistic estimated parameter values are likely to indicate either that the data do not contain enough information to estimate the parameters or that there is a more fundamental model error. In the first circumstance, the best response is to use prior information on the parameter value, which will tend to produce an estimate that is close to the most likely value, instead of at the less likely value that generally constitutes the imposed upper or lower limit. In the second circumstance, the best response is to find and resolve the error. Hill (1998) indicates that neither UCODE nor MODFLOWP supports constraining limits on parameter values because a circumstance in which the use of such limits is the best way to proceed has not been identified. Our experience with the Hanford problem in estimating specific yields for the Ringold and Hanford sediments supports Hill's assessment. With a constraint option available, deficiencies in the conceptual model would likely not be identified.

Two convergence criteria are calculated if either of the user-defined tolerances is met or the modified Gauss-Newton iterative process is terminated. For the first criterion, convergence is achieved when none of the parameters being inversed are changing by less than a specified tolerance (e.g., TOL=0.01). The absolute value of the fractional change in each parameter value j, $| d_j^r / b_j^r |$, is compared to TOL, and when all values are less than TOL [i.e., $| d_j^r / b_j^r | <$ TOL for all j=1,NP], the iterative process is terminated. To find the largest absolute value of the d_j^r / b_j^r, j=1,NP is less than a user-defined convergence criterion TOL. The second convergence criterion, SOSR, often is useful early in the calibration process to avoid lengthy simulations that fail to improve model fit and terminate the nonlinear regression when the sum of squared objective function (Equation 4.1) changes less than the user-defined amount (i.e., SOSR) for three sequential iterations. Discussions by Cooley and Naff (1990, p. 70) indicate the modified Gauss-Newton optimization typically converges within "a number of iterations equal to five or twice the number of parameters, whichever is greater." Convergence will tend to occur sooner for well-conditioned problems and later for poorly conditioned problems. It is rarely fruitful to increase the number of iterations to more than twice the number of parameters, which can take large amounts of computer time. It generally is more productive to consider alternative models.

4.2 Computational Codes

4.2.1 UCODE

UCODE (Poeter and Hill 1998) performs universal inverse modeling using nonlinear regression. UCODE uses an indirect approach rather than the direct approach in which the unknown parameters for the problem being solved are considered to be the dependent variables. In the indirect approach used by UCODE, the normal model equations (often referred to as the forward equations) are solved, and parameter estimates are sought that minimize the residuals (e.g., differences between observed and model predicted quantities). For the indirect groundwater flow inverse, the forward model is the standard groundwater flow equations, and the dependent variable is hydraulic head. The UCODE concept of universal comes from the fact that UCODE solves the inverse problem through an indirect approach and is not directly linked or tied to any particular forward model and implementing computer code. MODFLOWP (Hill 1992), for example, is a specific inverse model that only applies to quasi-three-dimensional groundwater flow problems. The nonlinear regression theory (derived largely from Cooley

and Naff 1990) used in UCODE, programs for various statistical analyses (Hill 1994), and the recommended guidelines for model calibration (Hill 1998) are principally the same as that used in MODFLOWP.

In UCODE, an estimated parameter can be any quantity that appears in the input files of the application model(s) or that can be used in conjunction with user-defined functions to calculate a quantity that appears in the input files. Observations to be matched in the nonlinear regression performed by UCODE can be any quantity for which a simulated equivalent value can be produced. Simulated equivalent values are calculated using values that appear in the application model output files and a set of additive and multiplicative functions. Also in UCODE, prior or direct information on estimated parameters can be included in the regression.

UCODE solves the nonlinear regression problem by minimizing a weighted LS objective function with respect to the parameter values with a penalty term related to prior information (Equation 4.1) using the modified Gauss-Newton method discussed in Section 4.1.1. Parameter sensitivities needed for the nonlinear regression method are approximated numerically using forward or central differences methods.

UCODE is designed to operate in phases. As indicated in the UCODE manual, it is useful to begin with PHASE=1 and proceed to 2 or 22, and then 3. Runs with PHASE=33, 44, and 45 generally are run only using a satisfactorily calibrated model. Phase 11 produces values that can be used to create a sum-of-squared, weighted residuals contour graph and may never be used in some circumstances. The function of each PHASE is described in Table 4.1 (from the UCODE manual).

UCODE allows for restarts and provides a variety of printed results (at each iteration if desired) and also results in the form of files to allow for graphical analysis and to allow the user to analyze the regression results. Parameter statistics reported allow the user to learn of the optimal parameter values and evaluate the quality of the calibration. Parameter statistics reported include scaled sensitivities, composite-scaled sensitivities, parameter covariance matrix, parameter values, standard deviations, coefficients of variation, 95% linear individual confidence intervals, and correlation coefficients. Regression performance measures reported include Marquardt parameter, the parameter that changed the most, and the amount of change. A fairly comprehensive set of statistical results is made available to check fit and residual statistics in the regression results. These include:

- table of observations, simulated values, residuals, and weighted residuals
- maximum, minimum, and average weighted residuals
- number of residuals >= 0 and number of residuals < 0
- number of sequences of residuals with the same sign (+ or -)
- least-squares objective function with and without the prior information
- weighted least-squares objective function value
- calculated error variance
- standard error of the regression
- square root of the calculated error variance,
- correlation coefficients with prior information
- maximum likelihood objective function, and the AIC and BIC statistics

Table 4.1. Brief Description of the Various UCODE Phases and Their Function (Poeter and Hill 1998)

PHASE	Function
1	Parameter substitution and forward modeling using the starting parameter values specified in the prepare file.
11	Substitutes parameters, performs a forward model run, and calculates the sum-of- squared, weighted residuals objective function for many sets of parameter values. PHASE=11 produces data sets from which objective-function contour graphs can be produced. Execution of PHASE=11 requires modification of the search-string lines of the *fn.pre* file.
2	Sensitivities at starting parameter values
22	Sensitivities and parameter variances, covariances and correlations at starting parameter values. Execution time for 22 is about twice that of 2 because central differences, rather than forward differences are calculated.
3	Perform regression.
PHASE=33 and 44 need to be preceded by executing UCODE with PHASE=3 and GRAPH=1 in the same directory (GRAPH is a variable in *fn.uni*).	
33	Calculate the modified Beale's measure of model linearity using methods discussed by Cooley and Naff (1990) and Hill (1994).
44	Calculate predictions and their linear confidence and prediction intervals. Only PHASE is read from the universal file.[1]
PHASE=45 needs to be preceded by executing UCODE with PHASE=44 in the same directory.	
45	Calculate differences and their linear confidence and prediction intervals. Only PHASE is read from the universal file.[1]

1. Calculated using a slightly modified version of the computer program YCINT (Hill, 1994).

- ordered weighted residuals
- correlation between ordered weighted residuals and normal order statistics.

The set of files made available by UCODE for use in the recommended graphical analysis to evaluate fit and residual statistics includes:

- "< >._os" - contains a list of the observed versus simulated values
- "< >._ww" - contains the weighted observed versus weighted simulated values
- "< >._ws" - contains a list of the weighted residuals versus weighted simulated values
- "< >._r" - contains the residuals
- "< >._w" - contains the weighted residuals
- "< >._nm" - contains data for making the normal probability graph of the weighted residuals (probability values are transformed so that they plot on an arithmetic scale)
- "< >._rd" - contains data for comparing model fit to a normal probability graph of random numbers
- "< >._rg" - contains data for comparing model fit to a normal probability graph of correlated random numbers.

The "< >._?" in this list represents the particular file that contains the described information following the UCODE file-naming convention described below, in which "< >" is the user-defined prefix used by UCODE in naming files (note that all UCODE output files have a "._?" format for their file extension).

UCODE is intended for use on any computer operating system. It consists of algorithms programmed in Perl[(a)] to facilitate the text manipulation for interfacing with any type of forward model and Fortran90, which performs numerical calculations efficiently.

UCODE is designed to allow inversion for any type of problem for which nonlinear regression is applicable for inverse modeling by allowing the user to use their existing forward model(s) as long as they use numerical (ASCII or text only) input or produce numerical output and the forward model can be executed in batch mode. Summarizing then, UCODE was developed to

1. manipulate application-model input files and read values from application-model output files
2. compare user-provided observations with equivalent simulated values derived from the values read from the application-model output files to compute a weighted LS objective function
3. use the modified Gauss-Newton method described in Section 4.1.1 to adjust the value of user-selected input parameters in an iterative procedure to minimize the value of the weighted LS objective function, which can include a penalty function related to prior information
4. report the estimated parameter values
5. calculate and print statistics to be used to
    a. diagnose inadequate data or identify parameters that probably cannot be estimated
    b. evaluate estimated parameter values
    c. evaluate how accurately the model represents the actual processes
    d. quantify the uncertainty of model simulated values.

A UCODE flowchart from Poeter and Hill (1998) is shown in Figure 4.2. UCODE uses a variety of input files, some of which have file-naming restrictions. Most UCODE input and output files are named with a user-defined prefix (referred to as fn. in the UCODE documentation and < > in this report) that is specified on the UCODE command line. For the required input files, which include the universal, prepare, and extract input files and the optional "function" and "XYZT" files, all of which are presented below, UCODE requires the specified extensions or suffixes of ".uni," ".pre," ".ext," ".fnc," and "xyz," respectively. If the UCODE command line argument is "hanford," then "hanford.uni," "hanford.pre," "hanford.ext," "hanford.fnc," and "hanford.xyz" would be the files UCODE would be expecting as input for the universal, prepare, and extract input files and the optional "function" and "XYZT" files, respectively. The graphical analysis output file,"< >._os." for the list of the observed versus simulated values would be "hanford.os."

(a) Perl (Practical Extraction and Report Language) is a freeware language originally designed as a glue language for UNIX operating systems. It is now one of the most portable languages available. The latest binaries are available from http://www.ActiveState.com/Products/ActivePerl/index.html.

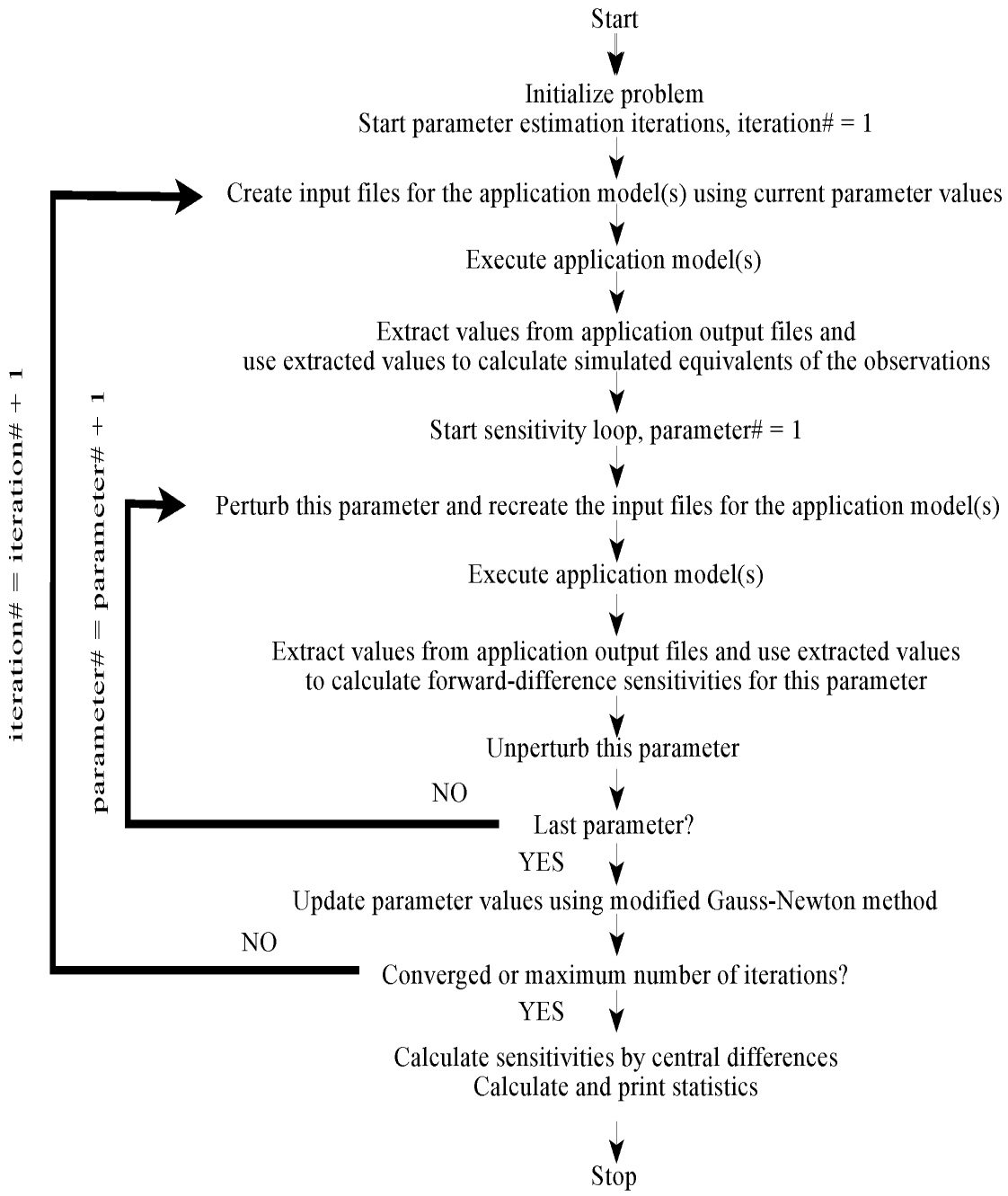

Figure 4.2. Flowchart for Estimating Parameters with UCODE

The UCODE manual (Poeter and Hill 1998) and the various guidance documents that come with the UCODE distribution contain the complete descriptions needed to actually run UCODE. The UCODE input files include the universal file, the prepare file, and the extract file (one of each is needed for each UCODE run), the optional function and XYZT files (one may be used for each UCODE run), and at least one template file (one or more are used for each UCODE run, depending on the number of separate input files that must be prepared for a forward model run). The application or forward model(s) executed by UCODE can include only one process/simulation model, a sequence of such models, or any combination

of pre-processors, process/simulation models, and post-processors, but each application model needs to be set up to run in batch mode. The following briefly describes the basic purpose and a brief description of the content of the various input files used by UCODE:

1. The universal file, < >.uni, is a required input file with naming restrictions that contains control parameters for the regression and printing and observation information.
2. The prepare file, < >.pre, is a required input file with naming restrictions that specifies whether a ".fnc" function file, < >.fnc, is used to define parameters. In addition, the prepare file supplies names for the template files, the application model input files that the template files are used to create with the defined parameters, the starting parameter values, reasonable minimum and maximum values of the parameters, perturbation size, a unique search string and formatting information for use in locating and substituting parameter values into template files, and prior information on the parameters.
3. The extract file, < >.ext, is a required input file with naming restrictions that names the application model output files, describes how to extract values from the model output, and defines how to use the extracted information to calculate simulated equivalents of the observations.
4. The XYZT file, < >.xyz, is an optional input file with naming restrictions that is read only if it exists in the directory where UCODE is launched that contains the observation names and their x, y, z, t coordinates.
5. The function file, < >.fnc, is an optional input file with naming restrictions that is read only if indicated in the prepare file, which allows functions of the parameter values to be used as input to the application model.
6. The template files (any file names and extensions can be used since these file names are specified in the ".pre" file above, but we have found it convenient to use a ".tpl" extension as is done in the UCODE literature) are actually copies of application model input files, edited such that the "search strings" specified in the prepare file discussed above can be used to replace values derived from the defined parameters. The prepare file describes how the search strings are to be replaced with numbers to create input files for the application model.

Parallel Operation and Other Enhancements

UCODE Version 1.09, which existed at the beginning of this effort in FY 2000, required that all forward model runs be executed in a serial fashion, even though all the forward model runs associated with approximating parameter sensitivity coefficients by forward or central differences methods at each step in the iterative regression process could in fact be executed in parallel. Initial estimates for the computational effort, based on conservative estimates for the number of parameters and simulation run times for the transient (1943–1996) three-dimensional model of the Hanford Site, indicated a serial computational approach for a coupled flow and transport inverse would require more than a year of computational effort for each inverse run. As a result, a joint effort between PNNL and Dr. Eileen Poeter (IGWMC), a primary author of UCODE, was initiated to develop an enhancement that would allow the user to use either a serial (i.e., on a single computer) or parallel mode and to make other enhancements that would make the processing of our problem easier (e.g., dealing with large numbers of observations). In the recently developed parallel mode, UCODE communicates with a parallel task manager to propagate the multiple simulation tasks (i.e., the forward model runs), required to approximate parameter sensitivity coefficients, to a networked group of computers for simultaneous computation. While the

newly developed approach works on a heterogeneous network of computers with different resources and speeds, the full speedup is only obtainable on an isolated network of dedicated computers of nearly identical speed and resources. As a result, the Hanford inverse model runs were carried out on an isolated network of 23 identical 600 MHz Windows NT computers.

The newly developed parallel enhancement of UCODE (Version 3.03) was designed to operate with MasterTasker (Appendix B) developed by Micki McKinley (PNNL). MasterTasker, a Perl program for management of distributed tasks, is designed to allow multiple tasks (e.g., forward model runs) to be distributed for simultaneous execution on any of the networked systems of computers having access to a common shared disk. One of the computers must be running MasterTasker as a background "master" process and the rest of the computers must be running MasterTasker as a background "slave" process. Inter-computer and inter-task communications occur by reading and writing files with specified content and/or names on the common shared disk to predetermined locations specified in the MasterTasker initialization files, "< >.ini" (Appendix B) and in the UCODE universal files, "< >.uni". UCODE prepares a separate folder containing all the information and programs needed to complete each forward model run that can be computed in parallel. The information on the folder names as well as the name of the script file (e.g., a ".bat" file) that must be forked by a MasterTasker "slave" to initiate a forward model run on the slave computer is written by UCODE into a "parameters" file that is processed by the "master" MasterTasker, as described in Appendix B. The parallel version of UCODE and MasterTasker are in the public domain and available for download through the International Groundwater Modeling Center website, which, at the time of this report, is: http://www.mines.edu/igwmc/freeware/ucode/.

The other enhancements made as part of this effort are discussed in an Adobe Acrobat® file, "< >.pdf" d file (UCODE3.0_MANUAL_ADDITIONS.pdf) distributed with UCODE. There is not a separate parallel version of UCODE. If the parallel option is not specified, UCODE operates in its normal serial mode, including the new enhancements. In this section, only a brief description of the new parallel processing option is presented. Invoking the parallel option requires that the companion MasterTasker code discussed above must be operating on the master and slave computers to use the parallel option in UCODE. Additionally, the available computers must be of similar computing capacity, and the number available must be sufficient to gain a computing advantage. At each iteration of an inverse problem with N parameters, there are N+1 forward model runs that can be run in parallel, so the speedup (assuming all N+1 model runs take approximately the same time to compute on the available computers) is the integer part of (N+1)/M, where M is the number of computers running MasterTasker as a slave process. If N=19 and M=20, the inverse computation speedup would be ~20 times.

The specification of parallel operation in UCODE 3.0 is made in the universal input file just before the place where the END would normally be if a normal serial computational approach were being used. The parallel option is invoked by additions to the universal input file. Rather than having the normal END, a line with the PARALLEL is input, followed by the information required to undertake the parallel solution, as delineated below. As discussed above (and in more detail in Appendix B), the MasterTasker code manages a "master" computer and "slave" computers that have been set up to accept parallel executions of the application codes. For parallel operation of UCODE, all the batch files, which in parallel mode will be executed as a "forked process" on slave machines, must include a final line as follows:

```
echo process_complete>%1
```

if forward model run was successfully completed or

```
echo processing_failed>%1
```

if the forward model was not successfully completed.

This is because the MasterTasker slave process needs to be informed when the forward model run is completed and whether it was successful so it can be reported to UCODE. UCODE universal file additions include

PARALLEL (not case-sensitive)—Indicates the parallel option of UCODE will be invoked. If "PARALLEL" is specified, each of the items, delineated in bold below, must be specified in the following order and start on a new line:

PATH_to_MASTER—path to the directory, including a slash at the end (e.g. D:\master\), where the subdirectories that will hold the information for the parallel executions will be created.

PATH_to_DYNAMIC_FILES—path to the directory, including a slash at the end, where files that change throughout the parallel execution are stored and the application codes, batch files, and input files are held. Everything in this directory will be copied to the slave machines. These files need to be all that is needed to conduct a forward model run when the proper command is given, except the input files that are created by UCODE substitution of values into template files.

PATH_to_STATIC FILES—path to the directory where files that will not change throughout the parallel execution. Everything in this directory will be copied to the slave machines only once at the beginning of the parallel operation. The setup of the application must reference these files with the correct relative directory references.

REPEAT_TIME—frequency with which to check "slave" machines where one forward execution of the application model(s) is performed to obtain information for calculation of sensitivities.

REPEAT_TIME_UNIT—time unit for REPEAT_TIME (s=second, m = minutes, h = hours).

DEFAULT_SPACE—minimum amount of disk space required to hold the application software, data files, and output.

SPACE_UNIT—unit for DEFAULT_SPACE (M = megabytes, G = gigabytes).

DEFAULT_SPEED—speed assumed for slave computer to perform one forward execution of the application model(s) in the default_time (see below).

SPEED_UNIT—unit for DEFAULT_SPEED (Mhz = megahertz).

DEFAUL**T_TIME**—estimated time required to perform one forward execution of the application model(s) on a machine with the default speed.

TIME_UNIT—time unit for DEFAULT_TIME (s=seconds, m = minutes, h = hours).

DEFAULT_LAUNCH—command to launch one forward execution of the application model(s) when executed within the directory indicated as PATH_to_APPLICATION_FILES

TIME_OUT—time limit for waiting for short term responses (e.g. ,acknowledgement that requests have been made or status has been updated) from the Master (secs) before terminating

ABORT_CONDITION—condition to leave slave machines in after the program terminates unnaturally specify, either CLEAN or LEAVE**.**

After all optional features have been specified, the < >.uni file ends with:

END (not case-sensitive)—Indicates the end of the data input for this file.

Figure 4.3 is a list of the ".uni" file for the parallel test problem that comes with the UCODE distribution.

4.2.2 CFUCODE

A version of CFEST called CFUCODE was developed to work directly with the parallel version of UCODE. CFUCODE provides an interface between CFEST and UCODE that streamlines the calibration process. A detailed description of CFUCODE is provided in Appendix A, and a summary is provided here. CFUCODE performs two major functions:

- Processes observed data and parameter definitions and creates all the required UCODE input files (i.e., the universal, prepare, and extract input files and the optional "function" and "XYZT" files)
- Extracts CFEST simulation results and provides them to UCODE in an efficient manner so that extensive extraction parsing would not have to be processed from the large ASCII files from the normal version of CFEST.

4.2.2.1 Observed Data and Parameter Definitions

As discussed, UCODE requires several input files. CFUCODE generates these files in the format required by UCODE from information provided to CFUCODE in a series of simpler data and control files. The user is required to specify the format of the files input to CFUCODE, providing flexibility in data input. Observation locations are provided in user-defined x, y, z, locations and time units, keeping the observation data independent of the model grid and temporal specifications. CFUCODE then translates the observation locations to the internal element numbers, local coordinates, and appropriate time step. Only the observations that are within the currently specified simulation time period are selected as input to UCODE. This allows shorter time periods to be investigated when doing the initial debug runs to speed up the process without requiring new time period-specific data files to be prepared. In addition, each well name is converted to an internal sequential identification code for use by the UCODE interface, and an additional file that includes the display of simulated and observed data for easy comparison is generated.

CFUCODE also uses sets of model parameter variables referred to as paracodes that have been designed for use in the CFEST-UCODE tools to address the need to analyze parameter sensitivities. The paracodes developed for this initial inversing are grouped in the following broad categories: 1) hydraulic properties (i.e., hydraulic conductivities in the x, y, and z direction, specific yield, specific storage, porosity), 2) surface recharge, 3) stream and river properties, 4) boundary flux, and 5) well flow rates. A more detailed description of the use and implementation of paracodes is provided in Appendix A. Codes for model parameter variables do not currently include transport-related parameters (e.g. longitudinal and lateral dispersivities, molecular dispersion, retardation, and other transport related parameters). However, transport-related parameters will be added in the next update after needed experience has been gained with the flow parameters.

```
# phase
  3 #phase
# # of parameters
   1                # forward differencing (1 forward 2 central)
   0.01             # tolerance parameters
   0.0              # tolerance sosr
     0              # option to use quasi-newton updating = 0 i.e., NO
    10              # maximum # of iterations
   2.0              # maximum fractional parameter change
..\bin\mrdrive      # name of inverse code
    1               # number of application models
test\m              # name of application model (batch code for mf96)
    3               # sensitivity scaling  (0 none 1 scaled)
    0               # printing intermediate residuals & sensitivities(0 no print, 1 print)
    1               # graphing & postprocessing (0 no, 1 yes)
   10               # # of residual sets
# this well is trouble
1A          100.2 0.5 0 1
1B          121.5 0.5 0 1
# ?location
1C          141.0 0.5 0 1   # the location of this well is uncertain
1D          120.2 0.5 0 1
#etc
1E          128.9 0.5 0 1    # etc
2A          121.8 0.5 0 2
2B          126.8 0.5 0 2
2C          101.5 0.5 0 2
2D          144.2 0.5 0 2
2E          156.1 0.5 0 2
flow          2.  0.8 1 0
PARALLEL
C:\wrdapp\ucode3.0\master-tasker\shared\     #path to directory where parallel job will be run
C:\wrdapp\ucode3.0\test-parallel2\test\      #path to directory where variable files are held
C:\wrdapp\ucode3.0\test-parallel2\static\    #path to directory where unchanging files are held
5                              #rpttime 1 sec
s                              #rptunit
1                              #default space 1M
m                              #space unit
600                            #default speed
m                              #speed unit MHz
5                              #default time 2 min
m                              #time unit
m                              #default launch for (model.bat)
1500                           #time out (secs)
leave                          #leave files on slaves in case of abort
# end of parallel processing specifications and start of optional feature specification
RESTART_SAVE
ALTERNATE_PRINT
HEADER_PRINT
GROUP_STATS
END
```

Figure 4.3. Listing of UCODE Parallel Test Problem Universal (.uni) file

4.2.2.2 Extraction of Simulation Results for Residual Analysis

UCODE requires an ASCII output file and instructions to determine how it should extract simulation results from that file for each observation. Output files for simulations having tens of thousands of nodes and hundreds of time steps, such as in the implementation of the Hanford Site-wide model, are much too large for UCODE to process efficiently. Not only is the volume of data in the output file unwieldy, but

describing accurately the instructions for finding the few values within the output file that are needed by UCODE for each interpolation is subject to error for such complex files. Additionally, the files are time-consuming to generate. CFUCODE streamlines this process by generating a simple ASCII output file that includes only the simulated information for the observations listed in the UCODE universal file. Each line of this file contains simulated output for each observation line in the universal file, written in the same sequence as the observations are written in the universal file. CFUCODE also creates the required companion extraction file that tells UCODE how to process the simple result file just described. CFUCODE also generates additional ASCII output files that include simulation results, observation data, and other relevant information for the observations. This output is printed in a format that is easy for the user to review and plot. A more complete description of these input and output files is provided in Appendix A.

5.0 Implementation of Transient Inverse Approach for Prior Conceptual Model

This section summarizes the approach used in setting up the transient inverse recalibration of the prior site-wide conceptual model for the historical period of Hanford Operations (1943–1996). In this discussion and in Section 7, the context of the numerical implementation of this conceptual model and its evaluation will follow the "Guidelines for Effective Model Calibration" (Hill 1998) as outlined in the UCODE distribution documentation. The discussion will present each of the guidelines complete with its short description as it appears in Table 2 of Hill (1998) and then explain how this guideline was addressed or not addressed and why. The appropriateness and use of these guidelines are recommended in Neuman (1999) in his draft report on "Methodology to Identify and Evaluate Conceptual Models and Uncertainty Related to Groundwater Transport at Nuclear Facilities and Sites" for the U.S. Nuclear Regulatory Commission.

This section deals with Guidelines 1 through 6 presented in Table 5.1 and adapted from Hill (1998, Table 2). As part of these discussions, the rationale and purpose of the effort, rationale for the time period, the model parameters considered, the approach used in weighting observations, and the initial evaluation of parameter sensitivities are presented.

5.1 Implementation Strategy and Rationale

Guidelines 1 (principle of parsimony) and 2 (use of a broad range of information to constrain the problem) were only indirectly addressed because the scope of this effort was to recalibrate the prior SGM model using all available information on the historical impacts to the aquifer system between 1943 and 1996. The decision to initially apply this newly developed technical approach to the recalibration of the prior conceptual model was more practical than technical. Use of the prior model in this initial application allowed the emphasis to be placed on the development and testing and not on model building. It also allowed the activities such as gathering the information necessary to evaluate potentially key components missing from the prior conceptual model that were identified as part of the various reviews of the Hanford site-wide groundwater model (DOE-RL 2000) to be developed, and these alternative conceptual models and numerical implementations to proceed in parallel. These activities, which are still ongoing, include development and evaluation of basalt interactions to implement this understanding into an alternative conceptual model, which is the focus of the FY-2001 transient inverse investigations, development of alternative interpretations for the areal extent and thickness clay units, and development of regional scale zones for hydraulic conductivity and specific yield based on available data on facies according to Guidelines 1 and 2. Recalibration of the prior conceptual model also provided for development of baseline model fit information on the transient inverse calibrated prior model for comparing with subsequent alternative models and provided additional information about flaws and/or weaknesses in the prior conceptual model that had not yet been identified.

Table 5.1. Guidelines 1–6 for Effective Model Calibration (adapted from Table 2 of Hill 1998)

Guideline	Description
1. Apply the principle of parsimony	Start simple and add complexity as warranted by the hydrogeology and the inability of the model to reproduce observations.
2. Use a broad range of information to constrain the problem	For example, in groundwater model calibration, use hydrology and hydrogeology to identify likely spatial and temporal structure in, for example, areal recharge and hydraulic conductivity, and use this structure to limit the number of parameters needed to represent the system. Do not add features to the model to attain model fit if they contradict other information about the system.
3. Maintain a well-posed, comprehensive regression problem	a) Define parameters based upon their need to represent the system, within the constraint that the regression remains well posed. Accomplish this using composite scaled sensitivities (css_j) and parameter correlation coefficients. b) Maintain a comprehensive model in which as many aspects of the system as possible are represented by parameters, and as many parameters as possible are estimated simultaneously by regression.
4. Include many kinds of data as observations in the regression	Adding different kinds of data generally provides more information about the system. In groundwater flow model calibration, it is especially important to provide information about flows. Hydraulic heads simply do not contain enough information in many circumstances, as indicated by the frequency with which extreme values of parameter correlation coefficients occur when using only hydraulic heads.
5. Use prior information carefully	a) Begin with no prior information to determine the information content of the observations. b) Include insensitive parameters (parameters with small composite scaled sensitivities) in regression using prior information to maintain a well-posed problem, but during calibration it often is advantageous to exclude them from the regression to reduce execution time. Include these parameters for Guidelines 13 and 14. c) For sensitive parameters, do not use prior information to make unrealistic optimized parameter values realistic.
6. Assign weights that reflect measurement errors	Initially assign weights to equal $1/\sigma_i^2$, where σ_i^2 is the best available approximation of the variance of the error of the i_{th} measurement (this is for a diagonal weight matrix; see text for full-weight matrix.)

The rationale for the selected period of calibration (1943–1996), even though little observational data are available early in the time period, is that the groundwater system prior to this time period was in a nearly unstressed state and thus could be assumed to be in equilibrium with natural recharge and mean river flow conditions. This means that when performing a forward model run for a new set of parameters as part of the transient inverse, initial conditions at the start of each new forward model run could be determined by calculating them from a steady-state model run that considered only natural recharge and mean annual river flow stage for river boundary conditions. Preparing for simulation of this extended calibration period required developing a complete set of head observations for the entire extended

calibration period from 1943 to 1996, developing estimates of river stage variations for the entire extended period (see Appendix C), and developing estimates of the artificial discharges resulting from Hanford operations prior to 1979 to supplement the existing 1979–1996 data set prepared as part of the earlier calibration efforts for the prior model (Cole et al. 1997). Existing recharge estimates were already a best long-term estimate based on conditions prior to Hanford operations.

5.2 Parameterization

In developing the parameterization all model parameters, which include 1) parameters that represent the hydraulic properties of all the hydrogeologic units and 2) parameters that represent all the specified boundary conditions except for artificial discharges associated with Hanford Operations and pumping well rates, were considered for inclusion in keeping with Guideline 3. Artificial discharges and pumping were excluded because there are reasonable records for these model inputs. Hydraulic property parameters for the saturated fined grained mud and clay units (unit numbers 4, 6, and 8 of Table 3.1) were not included in this inverse analysis because it was believed that this would be better addressed by the ongoing activity to develop alternative model structure interpretations by examining the interpretational uncertainty in the areal extent and thickness of the various fine-grained mud and clay units, as discussed earlier. Additional hydraulic properties not considered are the specific yields of those hydrologic units below the water table, which includes the specific yield of hydrologic units 7 and 9. Of the various specified model boundaries, only the specified boundaries used to represent the Columbia River and the Yakima River stages were not specifically addressed in this initial inversing process. This is because they are treated as constant-head boundary conditions that vary in time and space, and the estimates for these spatial and temporal variations are a function of historical river flows and channel configuration that are reasonably well understood (Appendix C). The intensive computational requirements of the inversing process provides additional motivation to give careful consideration to Guideline 1's principle of maintaining simplicity and to Guideline 2's suggestion to use hydrogeological and hydrological evidence to identify likely spatial structure to limit the number of parameters needed to represent the system.

Hydraulic parameters in this initial inverse were limited to the conductive units. The final list of model parameters that were initially evaluated in the inversing process by performing a UCODE phase 22 analysis to examine parameter correlations and sensitivity coefficients were limited to

- Hydraulic conductivity of conductive units: Hanford formation (Unit 1), Ringold Formation (Unit 5), Ringold Formation (Unit 7), and Ringold Formation (Unit 9),
- Specific yield of the units containing the water table: Hanford formation (Unit 1) and Ringold Formation (Unit 5)
- Specific storage of all the conductive units: Hanford formation (Unit 1), Ringold Formation (Unit 5), Ringold Formation (Unit 7), and Ringold Formation (Unit 9)
- Boundary Fluxes: Cold Creek Valley, Dry Creek Valley, Rattlesnake Hills, and Surface Natural Recharge
- Single anisotropy ratio applied to all conductive hydrogeologic units.

The UCODE phase-22 analysis results (Section 5.4) provide information to determine whether some parameters should be omitted from early phases of the regression to maintain a well-posed regression

problem (part b of Guideline 3). Since parameter sensitivity coefficients change throughout the regression process as a result of nonlinearities, the sensitivities need to be reinvestigated as the regression proceeds and parameters move away from their starting values.

In UCODE, parameters are defined by specifying how the parameters are to be substituted into the model input file(s). Selected parameter values can be input directly or input after being manipulated by multiplication, division, and a wide range of algebraic functions, as described in Poeter and Hill (1998). Parameters can also be defined as a scaling factor to allow for the perturbation of hydraulic properties or parameters that are spatially distributed over individual model layers or for boundary fluxes that may be defined over several node locations.

Table 5.2 summarizes all the model parameters that were initially evaluated (UCODE phase 22) in the inversing process. As can be seen, the number of parameters to be inversed was minimized to 15 by using the UCODE scaling approach that allowed the spatial variability in both the boundary fluxes of the prior model and in its hydraulic conductivity distribution to be maintained (Section 5.2.1). With the exception of the approach used to deal with the implicit spatial variability of the prior model's hydraulic conductivity distribution and boundary fluxes, all the other model parameters were treated as single model parameter multipliers for each hydrologic unit considered for that parameter type, as discussed above.

Table 5.2. Initial Model Parameter Definitions and Initial Values Used in the Inversing Process for the Prior Hanford Model

Parameter Type	Description	Initial Value	Type of Distribution
	Hydraulic Conductivity Multiplier		
Multiplier	Hanford formation (Unit 1)	1	Variable Distribution
Multiplier	Ringold Formation (Unit 5)	1	Variable Distribution
Multiplier	Ringold Formation (Unit 7)	1	Variable Distribution
Multiplier	Ringold Formation (Unit 9)	1	Variable Distribution
	Specific Yield		
Multiplier	Hanford formation (Unit 1)	0.25	Constant
Multiplier	Ringold Formation (Unit 5)	0.1	Constant
	Specific Storage		
Multiplier	Hanford formation (Unit 1)	1 e 10^{-6}	Constant
Multiplier	Ringold Formation (Unit 5)	1 e 10^{-6}	Constant
Multiplier	Ringold Formation (Unit 7)	1 e 10^{-6}	Constant
Multiplier	Ringold Formation (Unit 9)	1 e 10^{-6}	Constant
	Boundary Fluxes		
Multiplier	Cold Creek Valley	2881.3 m^3/yr	Distributed Flux
Multiplier	Dry Creek Valley	1207.0 m^3/yr	Distributed Flux
Multiplier	Rattlesnake Hills	3104.4 m^3/yr	Distributed Flux
Multiplier	Surface Natural Recharge	8.47 x 10^6 m^3/yr	Distributed Flux
Multiplier	Anisotropy of all hydrogeologic units	0.1	Constant

5.2.1 Hydraulic Properties

As outlined above, the hydraulic properties considered in the initial evaluation of the prior Hanford Site focused on the model characteristics of the major permeable hydrogeologic units identified in the model (model units 1,5,7, and 9). The properties evaluated included the hydraulic conductivities and the anisotropy of all major permeable units, the specific yields of the units that are found at the water table, and the specific storage of the units found below the water table.

Parameters that could be used to represent the hydraulic conductivities of the major units were evaluated first. Hydraulic conductivities of the major hydrogeologic units that were developed from the prior model-calibration process are highly variable (see discussion in Section 3.1.4). Without some simplification of these spatial distributions into a distribution of hydraulic conductivity zones, the number of parameters to consider in the inversing process could potentially be as large as the number of elements in the model. Using the required large number of parameters to represent the aquifer system as currently represented in the inversing process would not be practical, feasible.

Because this initial inversing was primarily done to demonstrate the feasibility of using the UCODE non-linear regression approach at the Hanford Site, a simpler alternative approach for parameterization was developed that make use of the scaling-factor capability in UCODE. In this approach, the spatial distribution of hydraulic conductivity in each conductive model layer is perturbed in the inversing process using a single multiplication factor for each model layer. Four multiplication factors in all were used to scale were used to scale the prior models spatial distribution of hydraulic conductivities in the conductive units which include the Hanford formation (Unit 1) and all permeable model units of the Ringold formation (Units 5, 7, and 9). This type of parameterization has been used successfully by others (D'Agnese et al. 1997 in evaluating conceptual models of regional flow systems and is recommended in parameterization approaches within UCODE (Poeter and Hill 1998) and MODFLOWP (Hill 1992) for model parameters that are spatially distributed. A single multiplier was used to parameterize (Table 5.2) and hence to perturb the anisotropy of all major permeable units from their initial ratio of 0.1 used in the prior model.

The specific yields of the uppermost hydrologic units were considered to be homogeneous in the prior model and as a result only one parameter value was needed to describe the specific yield for the Hanford formation (Unit 1) and one for the Ringold Formation (Unit 5), which are the dominant hydrogeologic units found at the water table. The initial specific yield was set at that of the prior model, or 0.25 for the Hanford and 0.1 for the Ringold Formation. As indicated in Table 5.2, a multiplier parameter type was used to perturb these parameters.

The specific storage of all units found at or below the water table were considered to be homogeneous and as a result required only four additional parameters were required, one for the Hanford and one for each of the three permeable Ringold Formation units. The initial specific storage was assumed to be 1×10^{-6} for all model units considered, and a multiplier parameter type was used to perturb these parameters (Table 5.2).

5.2.2 Boundary Conditions

Three multiplier type parameters (Table 5.2) were used to perturb each of the major boundary flux distributions within in the prior Hanford model because they are all spatially variable. This includes one parameter each for the prescribed upgradient spatially distributed fluxes at the Cold Creek Valley, Dry Creek Valley, and the Rattlesnake Hills Springs model boundaries, respectively. An additional multiplier-type parameter was used to perturb the spatial distribution of natural recharge fluxes infiltrating at the water table boundary of the model.

5.3 Well Observation Processing and Weighting for Use in Regression

According to Guideline 4 (Table 5.1) as many different kinds of observational data as possible need to be included in the regression to provide the variety of constraints need to break intrinsic correlations (e.g., recharge and hydraulic conductivity). Inclusion of flow observations (e.g., base stream flow contributed by the groundwater system) generally provides more information about the hydrologic system for groundwater flow model calibration. For the Hanford Site, the incremental river fluxes related to the Hanford groundwater system contributions cannot be measured. However, while there are no observational flux measurements available, the quantities of water discharged as a result of Hanford operations is relatively well documented, although still uncertain. These specified artificial discharges, whose magnitude is estimated to be much greater than natural recharge, may play a similar role relative to breaking the typical correlation between hydraulic conductivity and recharge.

The only observations available for inclusion in the regression are measured hydraulic heads made in wells completed in the unconfined aquifer. For the period between 1943 and 1996, a total of about 76,000 individual measurements were used in the calculations, with the initial measurements becoming available in 1948. These 76,000 observations were used by UCODE during the parameter-estimation process to define the objective function for the period of model simulation.

For purposes of the inversing process, each measurement of head had the following quantities specified: measured head value, well location, the principal contributing model unit represented by the measurement, and a statistical measure that would be representative of the potential measurement error in the observation. In UCODE, the measurement error is specified as either using a specified standard deviation and calculating the variance or specifying the variance for use in the residual head analysis.

Guideline 6 (pp. 45-46) was followed in the assignment of measurement errors; however, additional weighting following standard statistical approaches from Isaaks and Srivastava (1989) was also developed to remove sampling bias in the measurements related to the large variations in the spatial and temporal distributions of observation well measurements. Without proper weighting, the final calculation of the sum of squared residuals and the associated estimates of model parameters would be incorrectly influenced by measurements that are closely spaced in space and time.

5.3.1 Measurement Error

At the Hanford Site, nearly all wells under consideration have been carefully surveyed and measured using a steel tape or other relatively precise measuring device. Thus, the measurement error associated with each of the observations was assumed to be relatively small for most wells. This would generally

hold true for all measurements made in wells outside of the influence of the effects of transient river-stage fluctuations of the Columbia River. The measurement error variance was assumed to be a constant value of 0.03 m (0.1 ft).

For other wells located near and influenced by the Columbia River, a different specification of the measurement error was developed to deal with the temporal effects related to bank storage not part of the Hanford SGM. Because of the time-stepping used in the inversing process, simulated heads at wells located near and influenced by the Columbia River are more representative of the annual averages of head predicted near the river, and the Hanford SGM is not capable of predicting the large fluctuations in heads in wells near the river that respond to the highly transient behavior in river-stage fluctuations at time scales not simulated. Thus, the measurement error was developed by calculating the variance in measurements in wells near the river made before and after a specific measurement under consideration over a six-month period. As a result, wells near the river with larger head fluctuations would be assigned a larger measurement error and given relatively less weight in developing the objective function than wells away from the river.

5.3.2 Time Weighting

For time declustering, a one-dimensional version of polygonal declustering, as described in Isaaks and Srivastava (1989), was used where the weight attached to each observation is proportional to the period of time it represents within a given model time step. The period represented by each observation *i* is

$$\Delta t = \frac{\frac{t_{i+1} - t_i}{2} + \left(t_i - t_{i-1}\right)}{2} \tag{5.1}$$

except for first and last observations within a time step, where there is no division by 2. Consequently, the weight for observation *i* is

$$w_i = \frac{\Delta t}{T} \tag{5.2}$$

where T is the total time step.

A computer code was developed to read head observations, observation times, and well data from many data files that contain this information (a file for each well). Weights for all observations are calculated, and the times and dates are converted to model time starting at the beginning of the Hanford operation in 1940. There is also a provision for calculating measurement weight, inversely proportional to depth, but due to lack of rigorous relationships between depth to water table and accuracy, this option was not used.

5.3.3 Spatial Weighting

In some cases, more than one observation is located within the same time step and within the same element of the finite-element grid used for the inverse calibration. This produced a need for weighting the observations based on their location, that is, declustering based on spatial distribution. In this initial inversing effort, several unsuccessful attempts were made to use the polygonal declustering methods available through ARC-INFO to perform spatial declustering of the observation-well measurements directly using the finite element grid being used in the simulations. As a consequence, the method used for spatial declustering in this initial inverse was a simple weighting calculation based on the number of observations within an element during each time step. Observations falling singly within an element were assigned a weight of 1.0. If more than one observation fell within the same time step and element, the weight assigned to the observation was equal to 1.0 divided by the number of well observations.

Other applicable spatial declustering methods that implement similar global spatial declustering methods described in Isaaks and Srivastava (1989) will be identified and investigated for use in future inversing efforts.

5.3.4 Final Measurement Error Statistic

For purposes of this analysis, the resulting statistic used was the standard deviation calculated from the combination of the measurement error and the space and time weights using the following equation:

$$\sigma = \sqrt{ME^2 / 1.96^2 * SW * TW} \tag{5.3}$$

where

σ	is the standard deviation
ME	is the measurement error
SW	is the space weight
TW	is the time weight.

As indicated previously, for wells generally believed to outside the influence of the river stage changes in the Columbia River, the measurement error was assumed to be 0.03 m (0.1 ft). For wells influenced by river-stage fluctuations, the measurement error was calculated as the variance in measurements in wells near the river over a six-month period made before and after a specific measurement under consideration.

5.4 Initial Evaluation of Parameter Sensitivities

Following Guideline 3 (Table 5.1), the parameter sensitivities need to be examined at the outset to start the regression with a well-posed problem, which means that only parameters whose composite scaled sensitivities are within a factor of 100 of the maximum can generally be determined. The evaluation of which of the 15 identified parameters (Table 5.2) could be determined by regression was accomplished by making a UCODE phase 22 analysis, which performs an evaluation of the 1 percent scaled sensitivity coefficients for all parameters being considered. The scaled sensitivity coefficient for each parameter is calculated by evaluating the change in the objective function (i.e., residual between

simulated and observed heads) to a 1 percent change in each of the parameters evaluated. The results are summarized in a scaled composite sensitivity coefficient of each parameter that is formed by summing up the calculated residuals for all observations and dividing that quantity by the total number of observations.

Results of the calculated scaled composite-sensitivity coefficients for all parameters (Table 5.2) initially considered in this calibration effort are provided in Table 5.3. Results of this analysis were used to identify the most sensitive and least sensitive parameters and provided the basis for selecting the parameters to move forward within the initial calibration runs. Based on the results of this analysis, the calibration effort primarily focused on estimating the 10 parameter values listed in bold lettering in Table 5.3. These included:

- Scaling factors on hydraulic conductivity distributions within
  - Hanford formation (Unit 1)
  - Ringold Formation (Unit 5)
  - Ringold Formation (Unit 9)

- Scaling factors on the specific yield of the
  - Hanford formation (Unit 1)
  - Ringold Formation (Unit 5)

- Scaling factors for the
  - Cold Creek Valley boundary fluxes
  - Dry Creek Valley boundary fluxes
  - Surface recharge from natural sources

The following list of parameters, because of their associated low-sensitivity coefficients (i.e., a factor of at least 100 small than the maximum sensitivity coefficient of 126), were not given further consideration and were fixed during most of the inverse calibration.

- Scaling factors on hydraulic conductivity distributions within
  - Ringold Formation (Unit 7)

- Scaling factors on the specific storage values of the
  - Hanford formation (Unit 1)
  - Ringold Formation (Unit 5)
  - Ringold Formation (Unit 7)
  - Ringold Formation (Unit 9)

- Scaling factor for the anisotropy of all hydrogeologic units.

Some parameters (e.g., storage) were never revisited to see if their sensitivities changed while others were occasionally revisited to see if they had become sensitive near the optimized values for the other parameters being regressed. Parameter correlation coefficients and their influence on how well posed the regression is will be discussed in Section 7.

Table 5.3. Summary of Composite Scaled Sensitivity Coefficients from UCODE Phase-22 Analysis for Parameters and Factors Considered in Inverse Calibration (Table 5.2)

Model Parameter Factors	Scaled Composite Sensitivity Coefficient
Hydraulic conductivity distributions within	
Hanford formation (Unit 1)	**49.7**
Ringold Formation (Unit 5)	**126**
Ringold Formation (Unit 7)	**0.94**
Ringold Formation (Unit 9)	**5.10**
Specific yield of the	
Hanford formation (Unit 1)	**23.7**
Ringold Formation (Unit 5)	**72.7**
Specific storage values of the	
Hanford formation (Unit 1)	0.00
Ringold Formation (Unit 5)	8.28E-02
Ringold Formation (Unit 7)	5.52E-02
Ringold Formation (Unit 9)	0.24
Boundary Fluxes	
Cold Creek Valley	**30.8**
Dry Creek Valley	**20.4**
Rattlesnake Hills	**8.3**
Surface natural recharge	**28.0**
Anisotropy of all hydrogeologic units	3.91E-03

***** The 10 parameters in bold are those selected for consideration in the initial inverse calibration.

6.0 Results of Transient Simulation for the Period of Hanford Operations (1943 to 1996) Using the Prior Model

This section discusses results from using the prior SGM model to simulate the historical time period from just before the start of Hanford Site operations in 1943 through 1996. This extended calibration time period (1943 to 1996) was selected for the initial three-dimensional transient inverse modeling discussed in the Section 7. Preparing for simulation of this extended calibration period required development of estimates of the artificial discharges and river stage variations (Appendix C) before 1979 to supplement the 1979 through 1996 data set prepared as part of the calibration efforts for the prior model (Cole et al. 1997). In addition, a complete set of head observations for this extended calibration period (1943 to 1996) had to be prepared.

The rationale for this time period, even though few observational data are available early in the time period, is that the groundwater system prior to this was in a nearly unstressed state and thus it could be assumed to be in equilibrium with natural recharge and mean river flow conditions. This means that when performing the transient inverse, the conditions at the start of each new forward model run could be determined by calculating them from a steady state model run with only natural recharge and mean annual river flow stage for boundary conditions.

Modeling the same historical time period with the prior model using the parameter estimates from the prior model calibration efforts (Cole et al. 1997) provides a useful baseline set of the model fit measures for comparison with results from the three-dimensional transient inverse calibration of the prior conceptual model discussed in this report (Section 7). Measures summarized for the prior model simulation include spatial and temporal distribution of residual error and residual error statistics.

6.1 Spatial and Temporal Distributions of Residuals

Several graphical visualizations were used to examine spatial and temporal changes in model results and residuals over the historical period from just before the start of Hanford Site operations in 1943 through 1996 to evaluate the goodness of fit obtained with the previously calibrated prior model discussed in Cole et al. (1997). For comparison purposes, these same graphical visualizations are presented in Section 7 for the results predicted by the newly calibrated prior conceptual model based on the parameters determined from our initial application of the three-dimensional transient inverse calibration approach. Changes in the spatial distribution of the simulated water-table elevation contours and in the spatial distribution of residuals (i.e., the difference between simulated and measured heads in wells) were examined over the operational period from 1943 to 1996 to investigate the overall trends. The simulated configuration of the water table elevation contours at the start of the simulation in 1943 is shown in Figure 6.1 and the predicted change in water-table elevation contours is presented in five-year increments between 1950 and 1996 in a series of plots found in Appendix D (Figures D-1a through D-1k). Figures D-2a through D-2j also contain color-coded plots illustrating how the spatial distribution of measurements, their location, and the associated head residuals (i.e., difference between simulated and measured head values) vary for the same for the same five-year increments between 1950 and 1996 for which water table elevation contours are shown. Additionally Appendix D (Figures D-3a through D-3j) shows scattergrams of simulated versus measured heads for the same five-year increments between 1950 and 1996 for which water table elevation contours and spatial distributions of residuals are shown.

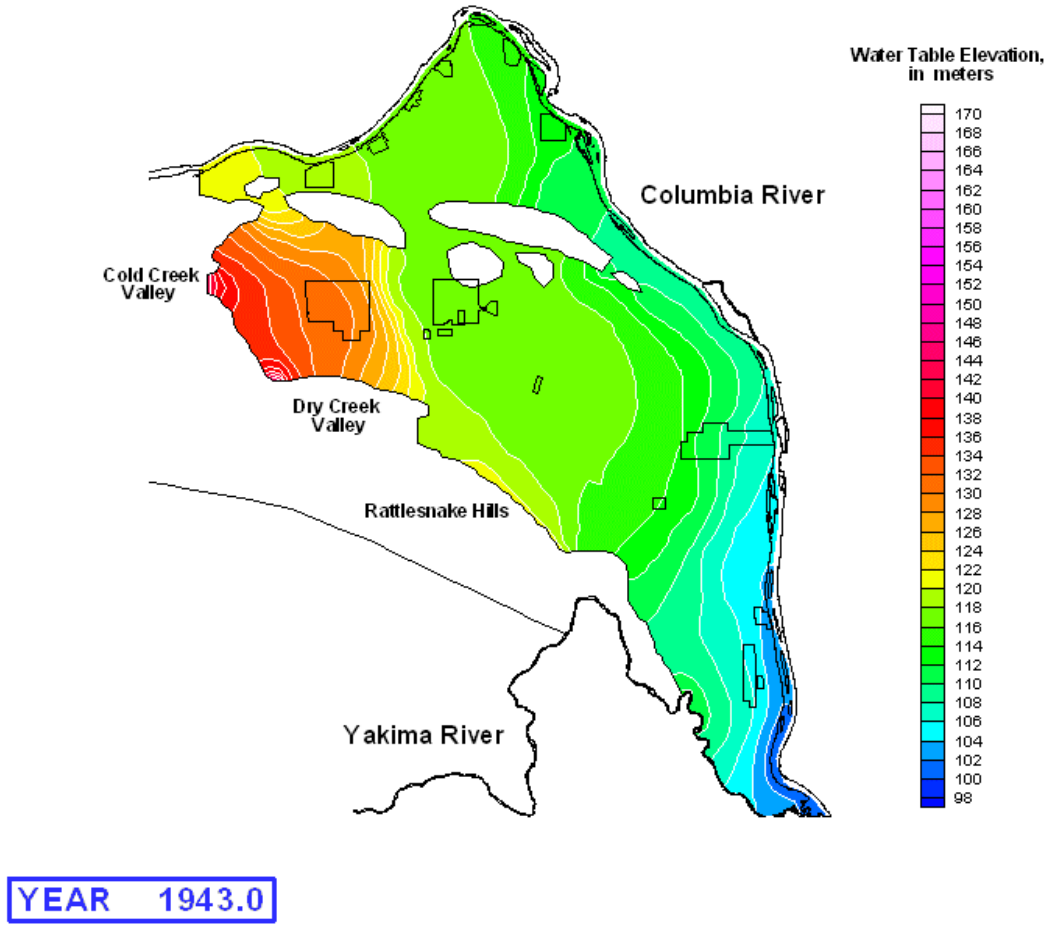

Figure 6.1. Steady-State Water Conditions for 1943 Using Prior Model

The scattergrams presented contrast the scatter of residual error points with a 45-degree line that would represent a perfect match between model and data. The individual plotted points are also color-coded to indicate the magnitude of the residual error

The series of water-table contour plots illustrates the overall growth and decline of major groundwater mounds beneath major wastewater discharge facilities in the operational areas. The most notable mound growth across the site relates to two large mounds associated with discharge facilities in the 200 East and 200 West Areas.

Near the 200 East Area, large volume discharges, primarily at B-Pond and Gable Mountain Pond, caused the predicted water table to rise several meters near each facility between the early 1960s and the early 1990s. A maximum water-table rise of about 14 m was simulated beneath B-Pond in 1988. A maximum water table rise of about 6.5 m was simulated beneath Gable Mountain Pond in 1986. Examination of calculated residuals indicates that predicted heads during this period of buildup (between 1960 and 1990) were within 1 to 3 m of observed water levels at individual wells within the 200 East Area.

In the 200 West Area, groundwater mounds are associated with large-volume historical discharges near T-Plant, U-Pond, and S-Pond that occurred from the late 1940s through the mid-1980s. Between 1943 and 1959, the simulated water table rose about 40 to 43 m beneath T-Plant and U-Pond. Between 1943 and 1965, the simulated water table rose about 38 m beneath S-Pond. The highest rise in the water-table position was simulated near U-Pond in 1959. Within 200 West Area, the predicted water-table position was significantly higher and broader than was observed in wells during the periods of highest discharges and mound building. Some of the calculated residuals were as much as 16 to 17 m higher than values observed in wells near the major discharge areas.

An overall comparison of simulated and measured water levels for the entire calibration period is shown in Figures 6.2 and 6.3. The large residuals associated with predicted and observed heads near large-scale discharges in the 200 West Area are noteworthy on the upper end of the overall scale where

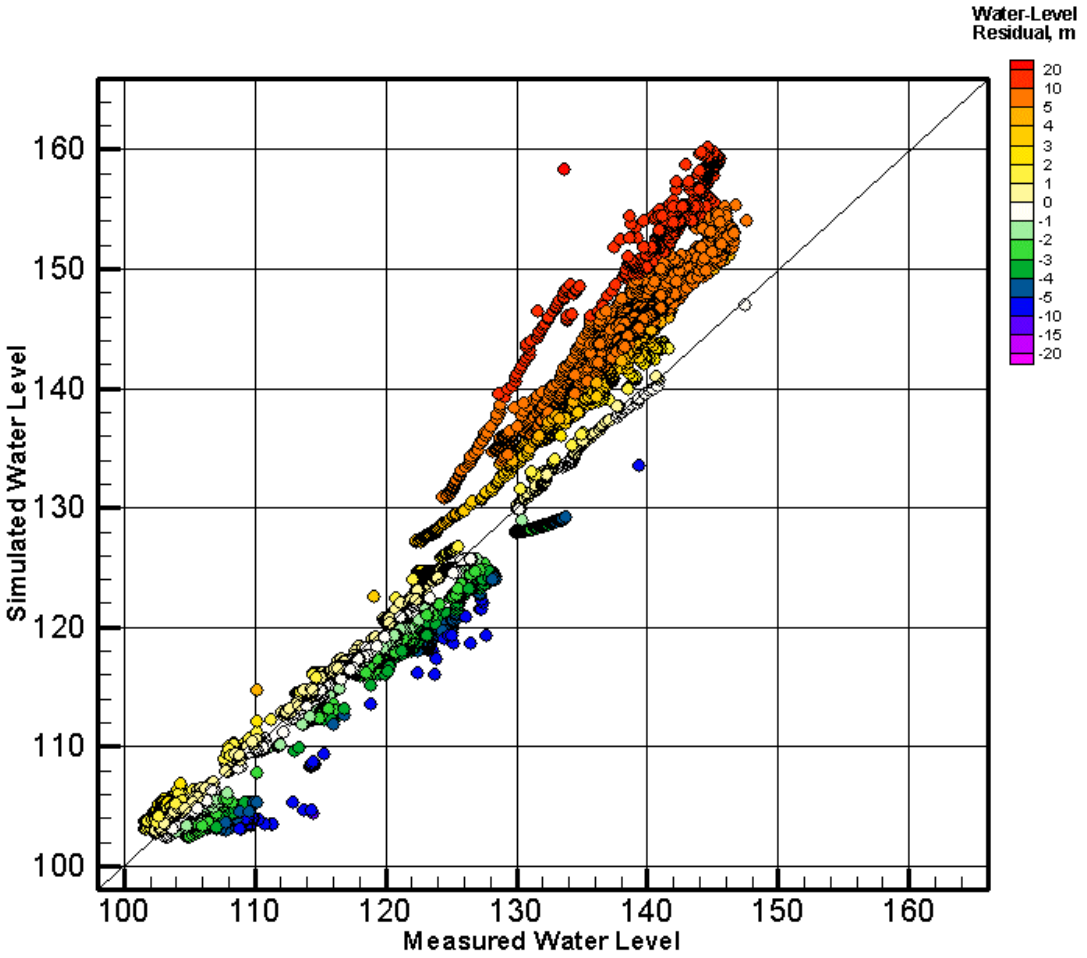

Figure 6.2. Simulated Versus Observed Heads for Prior Model for All Observations Through Time

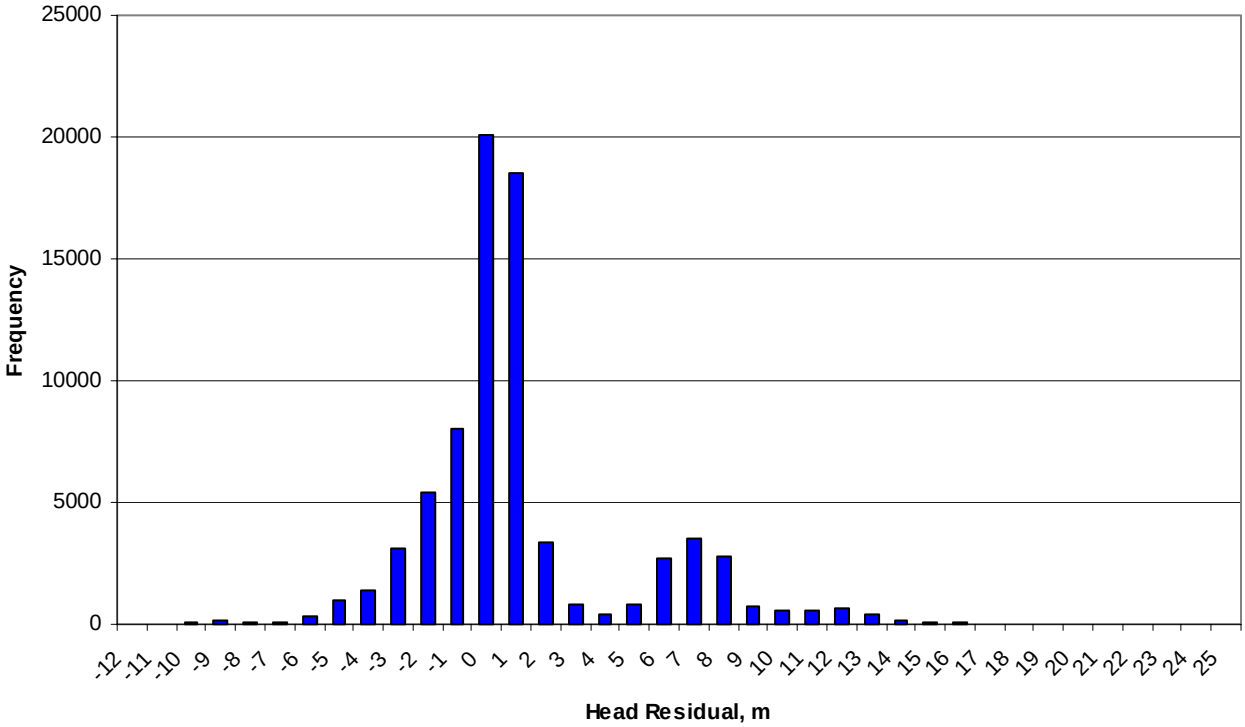

Figure 6.3. Histogram of Head Residuals for Prior Model for All Observations Through Time

simulated head values range to just over 160 m at locations where observed water levels are on the order of 145 m. As a result, the scattergram of simulated versus observed heads makes a significant departure above the 45-degree line that would indicate a perfect match between model and observations. Figure 6.3 shows a residual frequency histogram over the calibration time period that plots the number of residuals in each residual category from –12 m to +25 m in 1m increments. The residuals error frequency plot for this same time period distinctly shows a bimodal distribution of residuals. The main peak is the smaller one nearly centered around zero, and the second peak occurs at a residual error of 8 m.

6.2 Residual Error Statistics

Table 6.1 contains the various on the head residuals for the prior model. The graphical comparisons discussed above and shown in Appendix D and summary statistics of Table 6.1 illustrate that a significant number of simulated values were within a few meters of measured values. The overall mean residual was about 2.3 m (1.6 m for negative residuals and -3.1 m for positive residuals). The residuals ranged from -11.6 m to 24.8 m. The sum of squared residuals was calculated to be 1.01×10^6. About 51 percent of all values were between ±1 m of measured values and more than 85 percent of all simulated heads were between ±5 m of measured values.

Simulation results with the prior model indicate its limitation in duplicating overall historical trends in mound building during the 1943–1996 period of Hanford Site operations. The most noteworthy limitation of the prior model is that it overpredicts the historical growth and decline of groundwater mounds in the 200 West Area that were not considered in its calibration (i.e., the period from 1979

Table 6.1. Summary of Residual Error Statistics of Prior Model

Residuals		Number of Observations
Positive Residuals		
Mean (m)	3.101	36235
Standard deviation (m)	3.568	
Min (m)	5.5E-09	
Max (m)	24.830	
Negative Residuals		
Mean (m)	-1.578	39800
Standard deviation (m)	1.605	
Min (m)	-11.591	
Max (m)	-0.100	
Overall mean (m)	2.30	76035
Standard deviation(m)	3.588	
Sum of squared residuals (m2)	1.01E+06	
Residual Range	**Percent of Total**	
Between 1 and -1	51	
Between 2 and -2	66	
Between 3 and -3	77	
Between 4 and -4	82	
Between 5 and -5	85	
Greater than 5 m or less than -5 m	15	

through 1996). Simulated values of heads were between 10 to 15 m higher than were observed in wells during the period of maximum discharges and mound building that took place during the 1950s, 1960s, and 1970s.

7.0 Results and Evaluation of Transient Inverse Simulation of Prior Conceptual Model

The results of the transient inverse simulation of the prior conceptual model for the historical period of Hanford operations (1943–1996) are discussed in Section 7.1 in the same manner as for the prior SGM model in Section 6.

Discussions in Section 7.2 evaluate the regression results by following the "Guidelines for Effective Model Calibration" Hill (1998) and the associated outline in the "UCODE_CHECK_LIST.pdf" file provided as part of the UCODE distribution documentation in the same way as done in Section 5. The remaining guidelines (7–14), as applicable, are presented with the short description that appears in Hill (1998, Table 2) and an explanation of how this guideline was addressed or not addressed and why.

7.1 Transient Inverse Simulation Results

Recalibrating the prior SGM model significantly improved the capability of the resulting model to simulate historical trends in water-table changes over the entire site. However, some parameters took on unrealistic values, indicating that parameter zonation or other conceptual model improvements are needed, as is discussed in Section 7.2. The comparisons, however, are still useful in indicating the kinds of improvements that are possible. Comparisons, measures, and discussions in this section parallel those summarized for the prior model simulation results (Section 6) so results of the two models can be directly compared. The discussions in this section include spatial and temporal distribution of residual errors and residual error statistics.

Spatial and Temporal Distributions of Residual Errors

Several graphical visualizations were used to examine spatial and temporal changes in model results and residuals over the historical time period from just before the start of Hanford Site operations in 1943 through 1996 to evaluate the goodness of model fit obtained for the prior conceptual model using the inversed set of parameter values. Changes in the spatial distribution of the simulated water-table elevation contours and the changes in the spatial distribution of residuals (i.e., the difference between simulated and measured heads in wells) were examined over the operational period from 1943 to 1996 to investigate the overall trends. The simulated configuration of the water table elevation contours at the start of the simulation in 1943 is shown in Figure 7.1, and the predicted change in water-table elevation contours is presented in five-year increments between 1950 and 1996 in a series of plots found in Appendix E (Figures E-1a through E-1k). Appendix E (Figures E-2a through E-2j) also contains color-coded plots illustrating how the spatial distribution of measurements, their location, and the associated head residuals vary for the same for the same five-year increments between 1950 and 1996 for which water table elevation contours are shown. In addition, Appendix E (Figures E-3a through E-3j) shows scattergrams of simulated versus measured heads for the same five-year increments between 1950 and 1996, for which water table elevation contours and spatial distributions of residuals are shown. The scattergrams contrast the scatter of residual error points with a 45-degree line that would represent a perfect match between model and data. The individual plotted points are also color-coded to indicate the magnitude of the residual error. All of the plots presented in Appendix E are equivalent to the plots for the prior model presented in Appendix D.

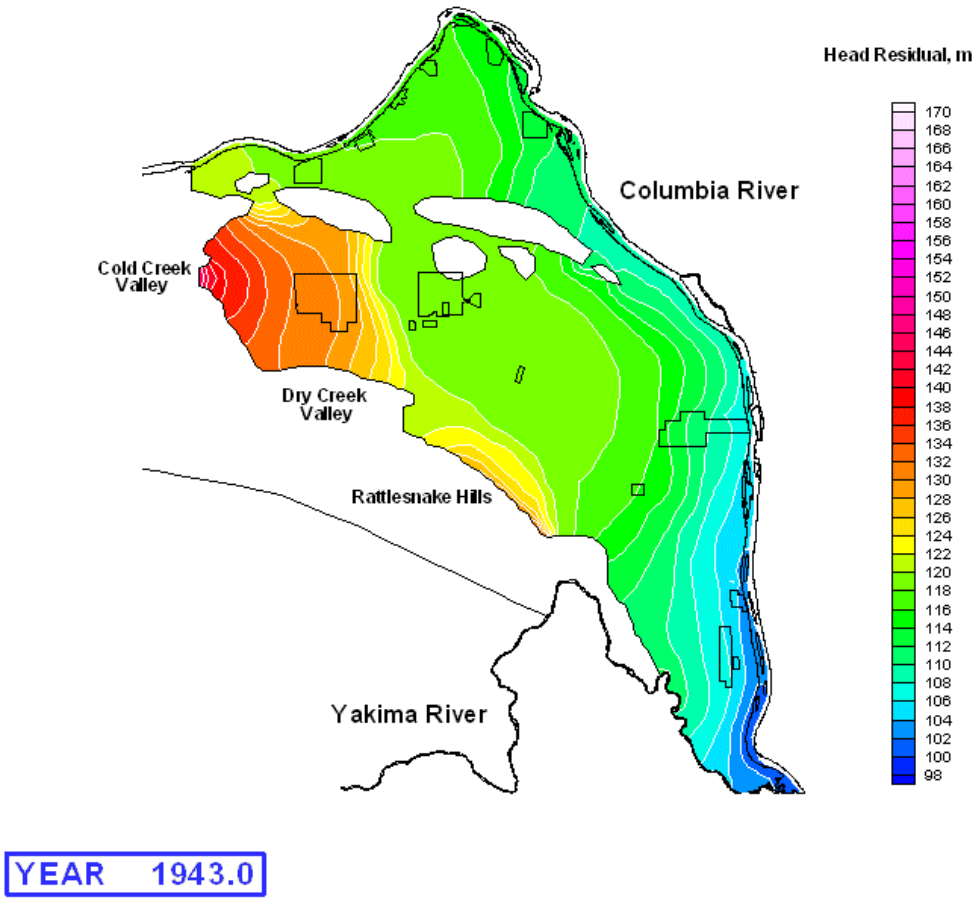

Figure 7.1. Steady-State Water-Table Conditions for 1943 Using Best-Fit Inverse Model

The series of water-table contour plots illustrates the overall growth and decline of major groundwater mounds beneath major wastewater-discharge facilities in the operational areas. The most notable mound growth across the site relates to two large mounds associated with discharge facilities in the 200 East and 200 West Areas.

Near the 200 East Area, large-volume artificial discharges at B-Pond and Gable Mountain Pond dominated the impacts to the water table. Artificial discharges at these locations caused the predicted water table to rise locally on the order of 10 m near B-Pond and about 7 to 8 m near Gable Mountain Pond. Simulations showed the highest water-table positions at these locations occurring in 1988. A comparison of simulated water table changes from the prior and best-fit inversed models for 1988 conditions are provided in Figure 7.2 to contrast the simulated water-table rises derived from each model prediction. Heads predicted by the model for the period of operation in the 200 East Area were generally within 1 to 3 m of measured water levels based on calculated residuals at individual wells, which were generally comparable the residuals derived using the prior model.

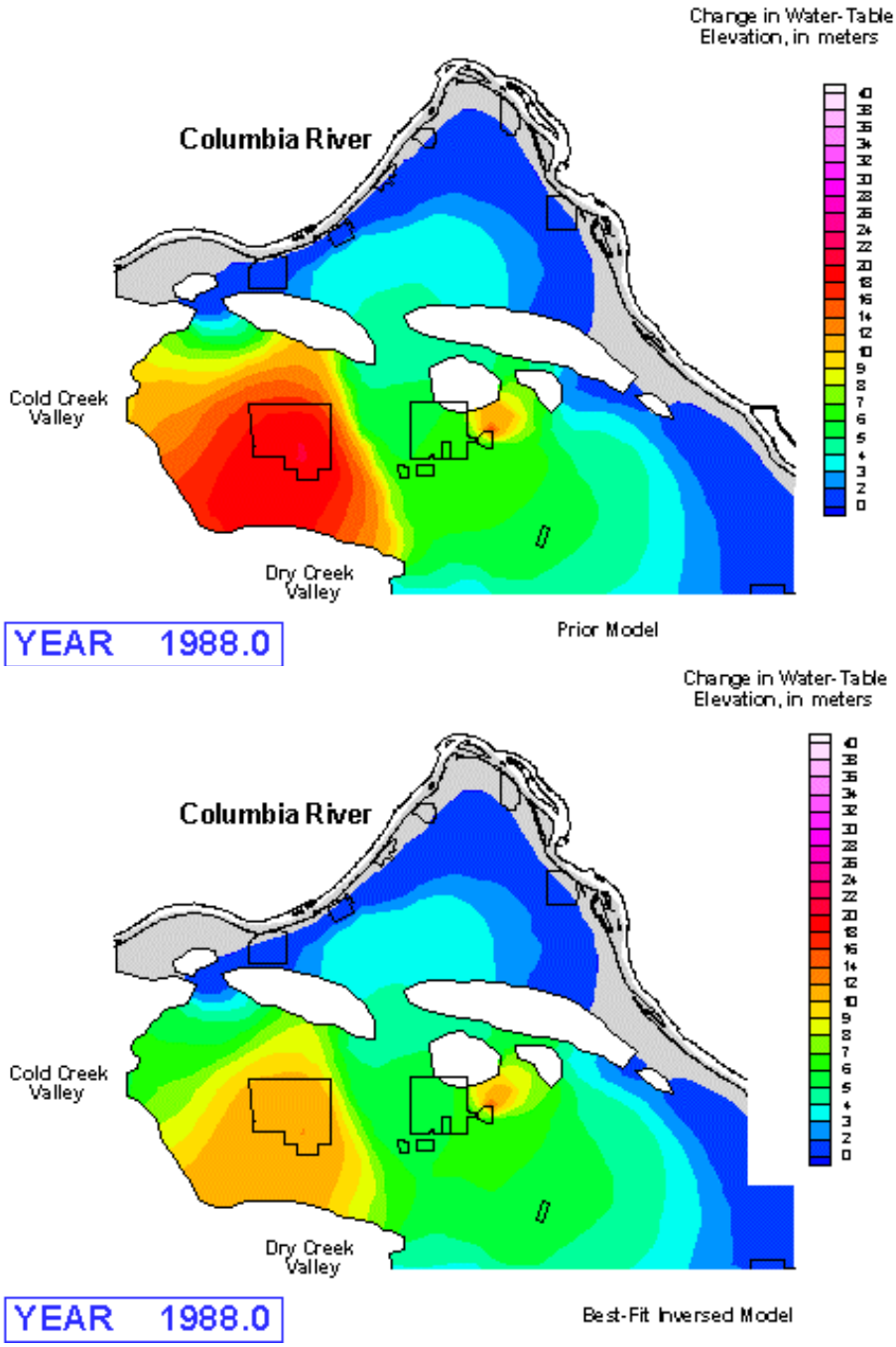

Figure 7.2. Water-Table Rise in Unconfined Aquifer Between 1943 and 1988 Using Prior and Best-Fit Inversed Models

In the 200 West Area, groundwater mounds were associated with large-volume historical discharges near T-Plant, U-Pond, and S-Pond that occurred between the late 1940s and mid-1980s. Using the best-fit model, simulated artificial discharges at these locations caused the predicted water table to rise locally on the order of 20 and 22 m near T-plant and U-Pond between 1943 and 1959 and about 20 m near S-Pond between 1943 and 1965. These predicted water table positions were significantly lower that those

predicted in the prior model simulation (e.g., 40 to 42 m near T-Plant and U-Pond between 1943 and 1959; 38 m near S-Pond between 1943 and 1965). The comparison of the simulated rises in the water table between 1943 and 1959 (Figure 7.3) and between 1943 and 1965 (Figure 7.4) illustrates the differences between the two model predictions. As in the prior model simulation, the highest water table in 200 West was simulated near U-Pond in 1959.

The water-table residuals calculated in 200 West, which were generally in the range of 1 to 3 m lower that measured water levels, were a considerable improvement over the calculated range of residuals resulting from the prior model simulation. Those were 15 to 17 m higher than measured water levels near major discharge areas.

An overall comparison of simulated and measured water levels for the entire calibration period (1943 to 1996), with the first observations starting in 1948, is shown in Figure 7.5, as is a scattergram plot of simulated head from the inversed model plotted versus the measured head. The scattergram also contrasts the scatter of residual error points with a 45-degree line that represents a perfect match between model and data with the individual plotted points color-coded to indicate the magnitude of the residual error. Figure 7.6 shows a residual frequency histogram over this same calibration time period that plots the number of residuals in each residual category from -12 m to +12 m in 1-m increments. The associated statistics for these comparisons are found in Table 7.1.

The graphical comparisons and summary statistics illustrate that, over the entire prediction period, 62 percent of the simulated values were within ±1 m of measured values and 98 percent were within ±5 m, compared to 51 and 85 percent, respectively, for the prior model. The mean residual was 1.1 m (-1.34 m for negative residuals and 0.91 m for positive residuals), which compares to 2.3 m (1.6 m for negative residuals and -3.1 m for positive residuals) for the prior model. The residuals ranged from -11.1 m to 11.8 m, which compares to –11.6 m to 24.8 m for the prior model. The sum of squared residuals was calculated to be 2.0×10^5 while the sum of squared residuals for the prior model was 1.01E+06.

Nearly 62 percent of all values were between ±1 m of measured values and, about 98 percent of all simulated heads were between ±5 m of measured values compared to51 percent and 85 percent for the prior model, respectively.

In general, all goodness of fit measures were significantly improved over the same measures for the prior model for the full calibration period from 1943-1996 which represents the entire period of Hanford operations. The most noteworthy improvement with the transient inverse calibrated model is its capability to fit historical trends of mound building in the 200 West Area. This improvement in the overall fit resulted in improved statistical performance in all categories (mean residual, range of residuals, sum of squared residuals) for the full calibration period.

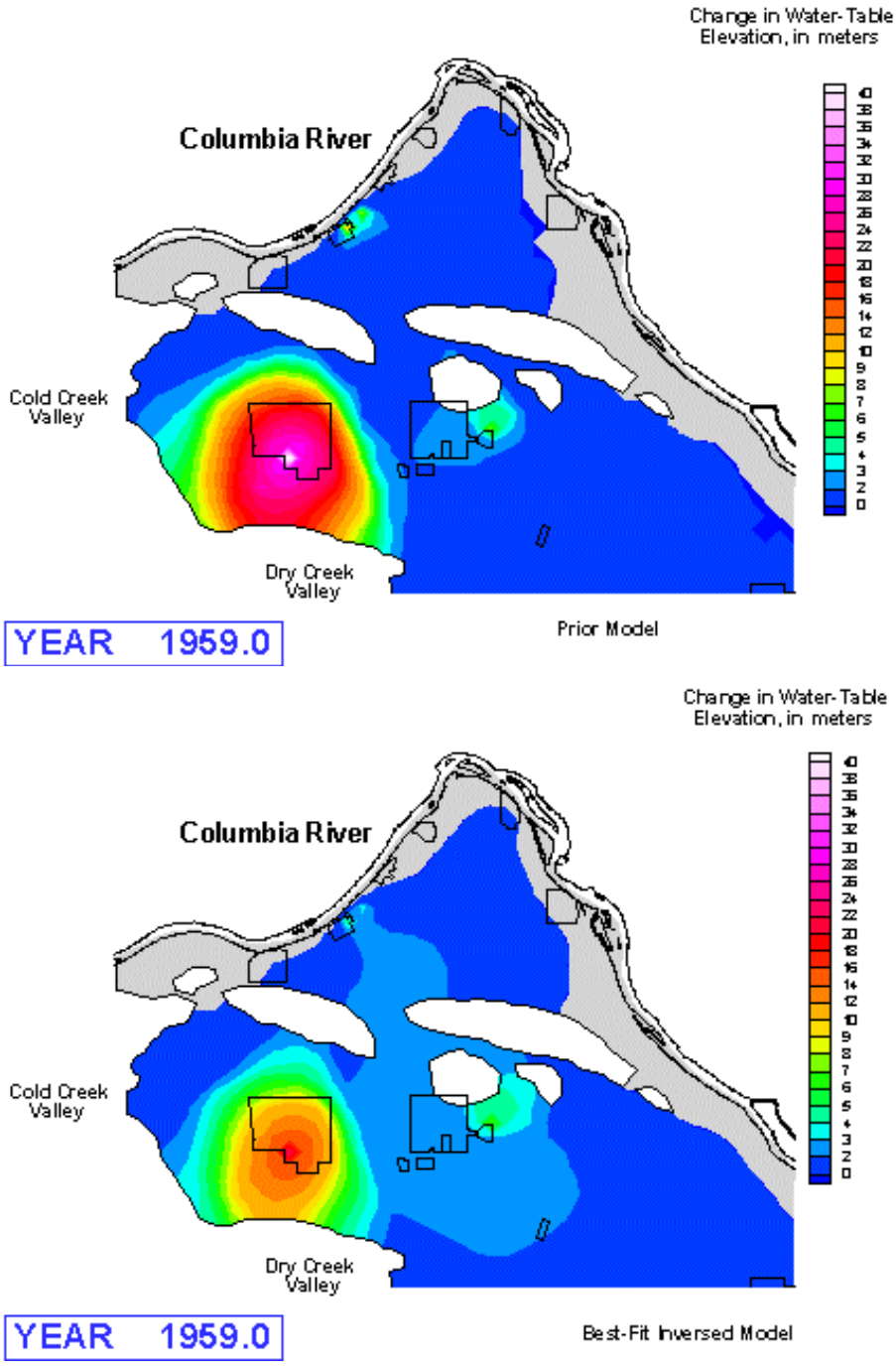

Figure 7.3. Water-Table Rise in Unconfined Aquifer Between 1943 and 1959 Using Prior and Best-Fit Inversed Models

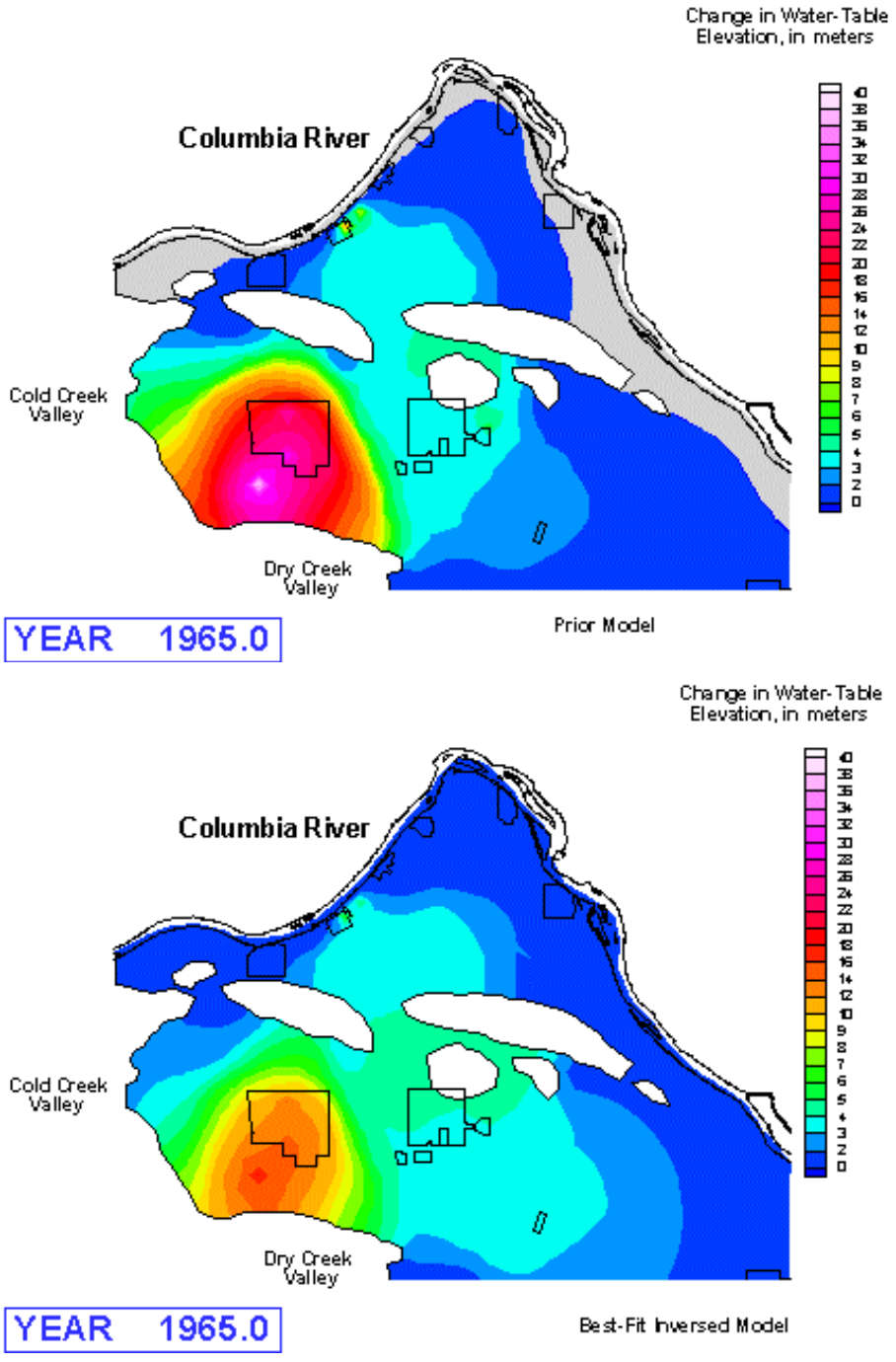

Figure 7.4. Water-Table Rise in Unconfined Aquifer Between 1943 and 1965 Using Prior and Best-Fit Inversed Models

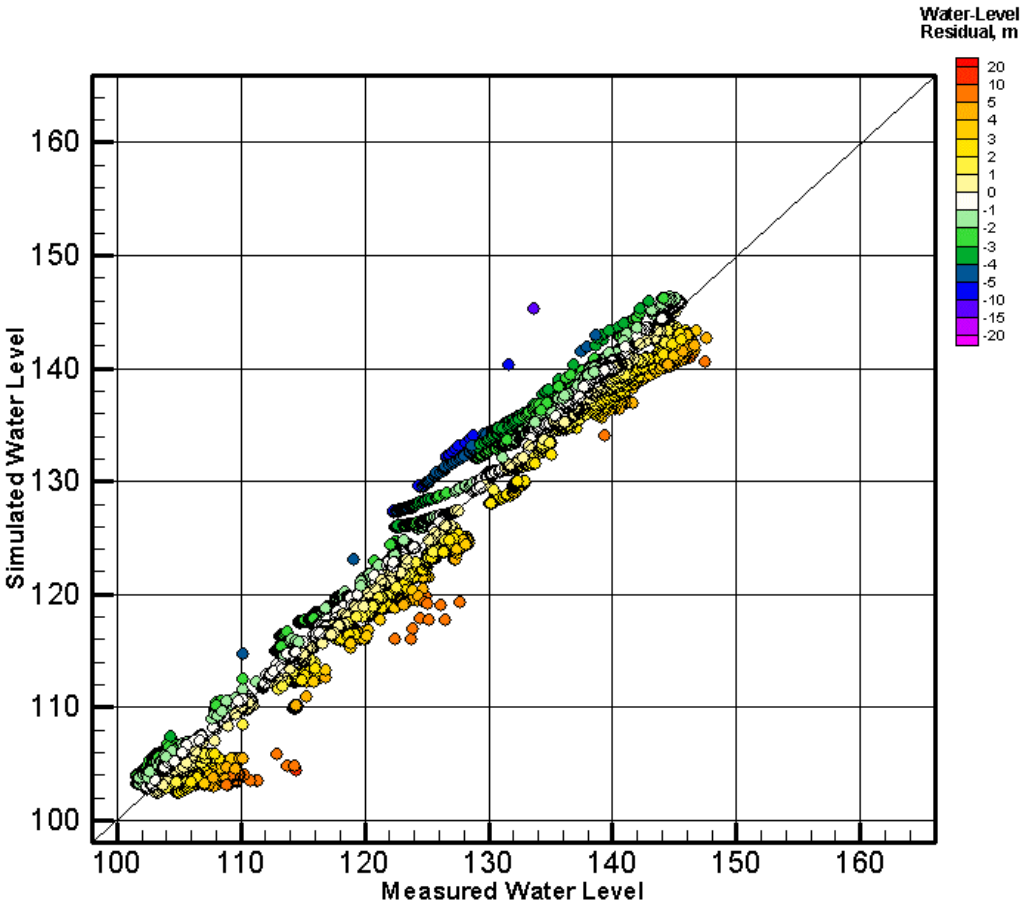

Figure 7.5. Measured Versus Predicted Heads (45-degree plot); All Observations for Best-Fit Inverse Model

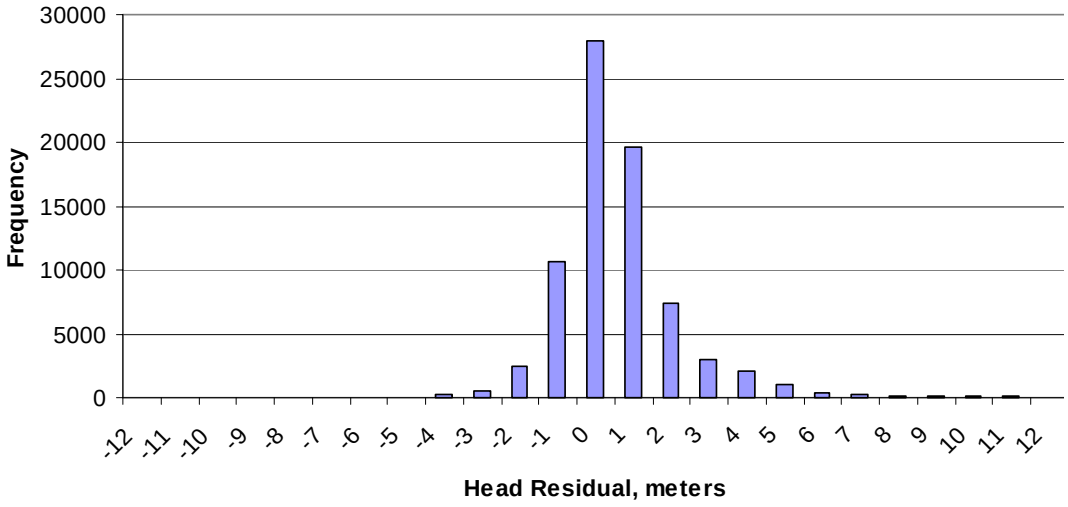

Figure 7.6. Histogram of Predicted Head Residuals; All Observations for Best-Fit Inverse Model

Table 7.1. Summary of Residual Error Statistics of Best-Fit Inverse Model

Residual Statistic		Number of Observations
Positive Residuals		
Mean (m)	0.91	41113
Standard deviation (m)	0.71	
Min (m)	0.10	
Max (m)	11.76	
Negative Residuals		
Mean (m)	-1.34	34922
Standard deviation (m)	1.55	
Max (m)	-11.07	
Min (m)	-9.3E-06	
Overall mean (m)	1.10	76035
Standard deviation (m)	1.62	
Sum of squared residuals (m^2)	2.01E+05	
Residual Range	*Percent of Total*	
Between 1 and -1 m	62	
Between 2 and -2 m	86	
Between 3 and -3 m	93	
Between 4 and -4 m	97	
Between 5 and -5 m	98	
Greater than 5 or less than -5 m	2	

7.2 Evaluation of Transient Inverse Model

This section provides an evaluation of the transient inverse calibration of the prior conceptual model using data for the historical period of Hanford operations (1943 to 1996). In this section, the regression results, model fit, and optimized parameter values are evaluated by examining various regression and statistical performance measures provided by UCODE (Poeter and Hill 1998). This evaluation, as was done for the discussions on the transient inverse implementation in Section 5, will follow the guidelines for effective model calibration (Hill 1998) and the checklist provided as part of the UCODE distribution documentation. The evaluation deals mainly with guidelines seven through fourteen presented in Table 7.2, adapted from Hill (1998, Table 2) (guidelines one through six, Table 5.1, were presented and discussed in Section 5 on implementation). In this section, most of the remaining guidelines (seven

Table 7.2. Guidelines 7–14 for Effective Model Calibration (adapted from Table 2 of Hill 1998)

Guideline	Description
7. Encourage convergence by making the model more accurate	Even when composite scaled sensitivities and correlation coefficients indicate that the data provide sufficient information to estimate the defined parameters, nonlinear regression may not converge. Working to make the model represent the system more accurately obviously is beneficial to model development, and generally results in convergence of the nonlinear regression. Use model fit and the sensitivities to determine what to change.
8. Evaluate model fit	Use the methods discussed in the sections of Hill (1998) "Statistical Measures of Model Fit" and "Graphical Analysis of Model Fit and Related Statistics."
9. Evaluate optimized parameter values	a) Unreasonable estimated parameter values could indicate model error. b) Identify parameter values that are mostly determined based on one or a few observations using dimensionless scaled sensitivities and influence statistics. c) Identify highly correlated parameters.
10. Test alternative models	Better models have three attributes: better fit, weighted residuals that are more randomly distributed, and more realistic optimal parameter values.
11. Evaluate potential new data	Use dimensionless scaled sensitivities, composite scaled sensitivities, parameter correlation coefficients, and one-percent scaled sensitivities. These statistics do not depend on model fit or, therefore, the possible new observed values.
12. Evaluate the potential for additional estimated parameters	Use composite scaled sensitivities and parameter correlation coefficients to identify system characteristics for which the observations contain substantial information. These system characteristics probably can be represented in more detail using additional estimated parameters.
13. Use confidence and prediction intervals to indicate parameter and prediction uncertainty	a) Calculated intervals generally indicate the minimum likely uncertainty. b) Include insensitive and correlated parameters, perhaps using prior information, or test the effect of excluding them. c) Start by using the linear confidence intervals, which can be calculated easily. d) Test model linearity to determine how accurate these intervals are likely to be. e) If needed and as possible, calculate nonlinear intervals (This is not supported in the present versions of UCODE and MODFLOWP). f) Calculate prediction intervals to compare measured values to simulated results. g) Calculate simultaneous intervals if multiple values are considered or the value is not completely specified before simulation.
14. Formally reconsider the model calibration from the perspective of the desired predictions	Evaluate all parameters and alternative models relative to the desired predictions using prediction scaled sensitivities (pss_j), confidence intervals, composite scaled sensitivities, and parameter correlation coefficients.

through fourteen), as applicable, are discussed. If a particular guideline was not followed or was not applicable, an explanation is given. This section includes discussions of

- regression measures
- evaluation of model fit
- evaluation of optimized parameter values
- summary of the evaluation and discussion of path forward

A shorthand is needed for the ten parameters identified in Table 5.3 for use in discussions, figures, and tables presented in the rest of this section. In the following discussions, the shorthand symbol for each of the ten parameters investigated are as follows:

- **K-U1** Hanford (Unit 1) Hydraulic Conductivity Multiplier
- **K-U5** Ringold (Unit 5) Hydraulic Conductivity Multiplier
- **K-U7** Ringold (Unit 7) Hydraulic Conductivity Multiplier
- **K-U9** Ringold (Unit 9) Hydraulic Conductivity Multiplier
- **SY-H** Hanford (Unit 1) Specific Yield Multiplier
- **SY-RU5** Ringold (Unit 5) Specific Yield Multiplier
- **F-CC** Cold Creek Valley Flux Multiplier
- **F-DC** Dry Creek Valley Flux Multiplier
- **F-NR** Natural Recharge Multiplier
- **F-RH** Rattlesnake Hills Flux Multiplier.

Additionally, a reference to a composite scaled sensitivity coefficient for a *parameter* will be written as *Css(parameter)* , e.g., *Css(F-NR)*.

7.2.1 Evaluation of Regression Measures

Besides the obvious failure-to-converge measure, UCODE provides additional information to evaluate whether the regression is well posed. This includes

- *Marquardt Parameter*, which as discussed in Section 4 is nonzero when the regression problem is ill-conditioned.
- *Parameter That Changed the Most*, because the parameter for which the maximum fractional change occurs is likely the to be at least a contributing problem when the regression does not converge.
- *The Amount of Change*, the magnitude of change.

Another measure is when the number of “parameter estimate iterations” need to converge exceeds the “on average” estimate given in Hill (1998) as approximately equal to twice the number of parameters (NP) being estimated (i.e., 2 NP).

The initial regression was run using all 10 parameters that, based on the phase 22 analysis, could potentially be estimated (Table 5.3), quickly indicated that the hydraulic conductivity distribution

multiplier for the K-U7 parameter could not be estimated because the composite scaled sensitivity coefficient ratio, *Css(maximum)/Css(K-U7)*, quickly increased from its initial value of 134 to values much greater than the value of ~100 discussed in the UCODE guidance document. As a result, this parameter was marked as "not to be estimated," and the regression was started over with the initial estimates for all nine remaining parameter multipliers to be estimated at their original estimate of 1.0, which corresponds to prior model values. This regression was allowed to proceed through iteration 9 with all 10 parameters bolded in Table 5.3 being estimated. By iteration nine, two additional composite scaled sensitivity coefficient ratios [*Css(maximum)/Css(K-U9)* and *Css(maximum)/Css(F-DC)]* were much greater than 100, as can be seen in Figure 7.7. Regression for these parameters was discontinued starting with iteration 10 when they were assigned values of 1.0 (i.e., their prior model values), and the regression was allowed to converge (Figure 7.8).

Once the insensitive parameter, K-U7, was removed from consideration, the convergence was generally well behaved, and all the regression measures for the transient inverse calibration of the prior conceptual model indicated that it was generally a well-posed problem. The 13 iterations for convergence was less than the "on average" estimate of 18 (i.e., 2*NP where NP=9), and the Marquardt parameter was zero for every iteration in the regression. An examination of the convergence history indicates that convergence could have probably been achieved in a few iterations if K-U7, K-U9, and F-DC had been excluded from the estimation process initially. Table 7.3 illustrates the numerical values for the parameter estimates and regression measures as a function of parameter estimate iteration (i). The composite scaled sensitivity coefficient ratio for the parameter with the smallest composite scaled sensitivity coefficient is termed (Max/Min Composite Scaled Sensitivity Ratio) in Table 7.3, and it is calculated as Css(Maximum)/Css(Minimum). This ratio should always less than ~100 for all the parameters to be estimated. Once K-U7, K-U9, and F-DC were excluded from the estimation process starting at iteration 10, it was well behaved and never exceeded 11.9. Similarly, the "Maximum Fractional Parameter Change," "Max. Frac. Change Limited To," and "Parameter with Maximum Fractional Change" regression measures were well behaved once K-U7, K-U9, and F-DC were excluded from the estimation process starting at iteration 10. Starting with iteration 10 the "Parameter With Maximum Fractional Change" changed with each iteration (i.e., F-CC, SY-RU5, F-NR, and SY-H).

7.2.2 Evaluation of Model Fit

Section 7.2.1 indicated that, in general, the regression was well posed and converged. This solution is thus the best fit for the problem posed. However, the quality of the model fit (e.g., how well it matches observations) must still be evaluated (guideline 8 in Table 7.2). As Hill (1998) points out, a powerful aspect of using nonlinear regression is the useful statistics generated that can be used diagnostically to measure the amount of information provided by the data and identify model error (bias), or to infer the uncertainty with which values are calculated. The difficulties common to nonlinear regression, according to Hill, make diagnostic statistics invaluable to successful regression. In the rest of this section, the various statistical measures for evaluating model fit and residual statistics provided as part of the UCODE distribution will be presented and discussed to evaluate the results of the regression.

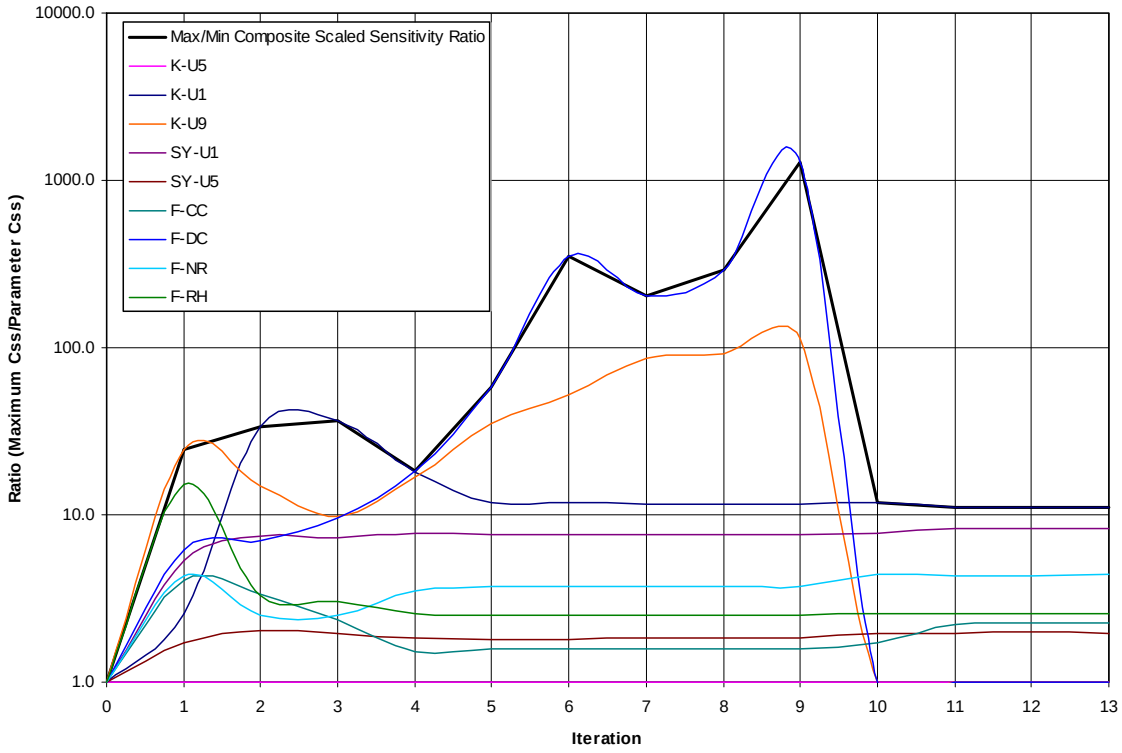

Figure 7.7. Composite Scaled Sensitivity Coefficient Ratios [Css(maximum)/Css(parameter)] as a Function of Parameter Estimate Iteration Until Convergence. Estimation of parameters K-U9 and F-DC ceased starting with iteration 10.

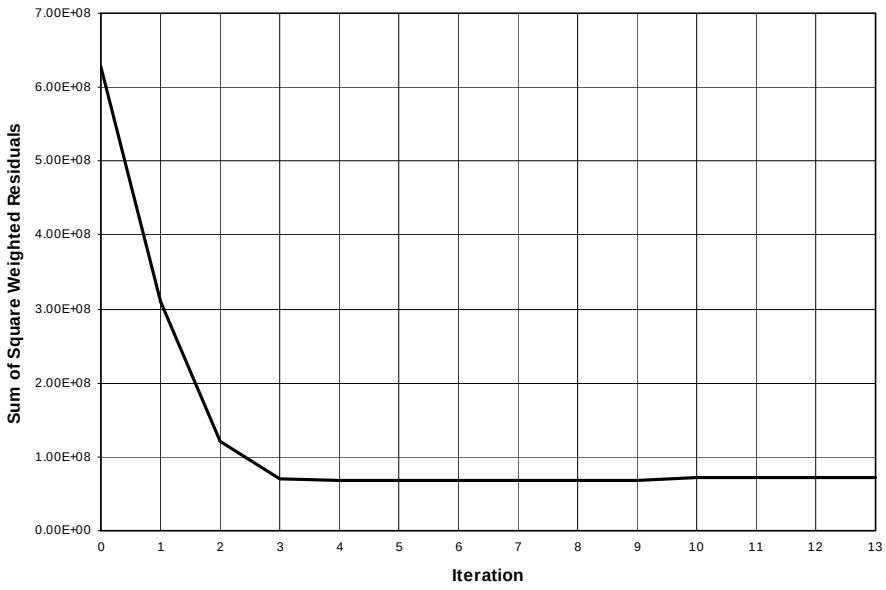

Figure 7.8. Sum of Square Weighted Residuals as a Function of Parameter Estimate Iteration Until Convergence

Table 7.3. Parameter Estimates and Regression Measures as a Function of Parameter Estimate Iteration (i)

i	K-U1	K-U5	K-U9	SY-U1	SY-U5	F-CC	F-DC	F-NR	F-RH	Max/Min Composite Scaled Sensitivity Ratio	Sum Sq. Wt. Residuals	Maximum Fractional Parameter Change	Max. Frac. Change Limited To:	Parameter With Maximum Fractional Change
0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.0	6.27750E+08			
1	0.985	1.269	0.419	0.829	1.224	1.256	0.748	1.372	1.900	24.6	3.10330E+08	4.674E+00	9.00E-01	F-RH
2	0.976	1.679	0.042	0.589	1.592	1.707	0.440	1.754	2.986	33.5	1.20860E+08	1.136E+00	5.71E-01	F-RH
3	0.943	2.089	0.008	0.320	2.051	2.362	0.149	1.788	4.119	36.6	7.03800E+07	-8.048E-01	none	K-U9
4	0.913	2.133	0.001	0.275	2.070	2.319	0.024	1.755	4.207	18.4	6.79910E+07	-8.368E-01	none	F-DC
5	0.913	2.135	0.001	0.277	2.062	2.316	0.002	1.755	4.214	58.1	6.79160E+07	-9.985E-01	-9.00E-01	F-DC
6	0.912	2.139	0.000	0.277	2.061	2.308	0.005	1.752	4.222	348.2	6.78440E+07	3.378E+01	9.00E-01	F-DC
7	0.912	2.139	0.000	0.277	2.061	2.308	0.003	1.752	4.222	204.5	6.78380E+07	-1.000E+00	-3.55E-01	F-DC
8	0.912	2.139	0.000	0.277	2.060	2.307	0.000	1.751	4.223	289.9	6.78270E+07	-1.000E+00	-9.00E-01	F-DC
9	0.912	2.139	0.000	0.277	2.060	2.307	0.000	1.751	4.223	1274.6	6.78259E+07	2.109E+17	3.19E-01	F-DC
10	0.901	2.243	1.000	0.277	2.129	2.097	1.000	1.723	4.349	11.9	7.13860E+07	-9.203E-02	none	F-CC
11	0.901	2.249	1.000	0.279	2.116	2.086	1.000	1.719	4.372	11.1	7.13812E+07	-5.769E-03	none	SY-RU5
12	0.901	2.248	1.000	0.278	2.117	2.086	1.000	1.715	4.374	11.2	7.13810E+07	-2.450E-03	none	F-NR
13	0.901	2.248	1.000	0.278	2.116	2.086	1.000	1.715	4.374	11.1	7.13810E+07	5.142E-04	none	SY-H

Removed From Regression

With more than 76,000 observations, it is impractical to present tables of observations, simulated values, residuals (calculated as the observations minus the simulated values), and weighted residuals. Inspections of residuals in the early stages of the regression process allowed bad data and weights to be identified, located, and removed. Graphical methods for the inspection of residuals were valuable. The spatial and temporal distribution of residual errors were presented and discussed in Section 7.1 and Appendix E. Figure 7.5 shows measured versus predicted heads (45-degree plot).

The following table contains statistics on the residuals from the converged regression printed by UCODE:

```
Statistics on residuals:
Sum of squared weighted residuals                    0.71381e+08
Sum of squared weighted residuals (with prior)       0.71381e+08
Maximum weighted residual:  0.689e+03                Obs#      12803 H12803
Minimum weighted residual: -0.406e+03                Obs#      45409 H45409
Average weighted residual:    0.440e+00
          # residuals ∃ 0:      34922
          # residuals < 0:      41113
Number of runs:     9668      in 76035 observations

Interpreting the calculated runs statistic value of   -205.
Note: the following applies only if
          # residuals ∃ 0 is > 10 and
          # residuals < 0 is > 10
The negative value may indicate too few runs:
  If the value is < -1.28, there is < 10% chance the values are random
  If the value is < -1.645, there is < 5% chance the values are random
  If the value is < -1.96, there is < 2.5% chance the values are random.
```

The maximum weighted residual indicates where the worst fit occurs relative to the expected fit and often reveals gross errors that can be identified and corrected. The minimum provides a little context by which to judge the maximum value. An average weighted residual near zero is needed for an unbiased model fit (usually satisfied if regression converges). No real comment can be made about the average weighted residual of 0.44 for this case, although it is relatively small given the nearly 70 m spread in the head values over all the Hanford Site. It will serve as a basis for comparison against future inverse efforts examining alternative conceptual models. However, Hill (1998) indicates that, if weights reflect the measurement errors, weighted residuals that are on average larger than 1.0 indicate that the model is worse than would be expected given anticipated measurement error, and values smaller than 1.0 indicate that the model fits better than expected given anticipated measurement error.

The number of positive and negative residuals indicates whether the model fit is consistently low or high. Preferably, the two values are about equal. The numbers, 34,922 negative and 41,113 positive, are fairly evenly balanced around the average of 0.44, as is better illustrated in the histogram in Figure 7.6. This figure provides more information about the distribution of residuals and indicates that the tails of the distribution are skewed on the positive side and thus not normally distributed. The number of test runs provides a statistic to help identify trends in spatially distributed weighted residuals since identifying trends (lack of non-randomness) by visual inspection is not always reliable. Too few or too many runs could indicate model bias. The number of runs, the runs' statistics value, and the means to interpret it are

printed by UCODE (see above). For this regression the runs statistic indicates too few runs, and the calculated statistic indicates that there is less than a 2.5% chance the values are random. Personal communication with Eileen Poeter indicates there may be a problem with this statistic related to the way the observations are ordered in our input file because, in both UCODE and MODFLOWP, the weighted residuals are analyzed using the sequence in which the observations are listed in the input file. Hill (1998) goes on to indicate that the runs test is included because it takes the order of the residuals into account, which is ignored in all the other summary statistics. If observations are grouped by location in transient simulations, too few runs commonly indicate positive serial correlation between residuals at individual locations. The effect of input observation ordering on the runs test will need to be investigated.

The following statistical information also provided by UCODE provides additional information about the model fit.

Least-Squares Objective Function	
(dependent variable only)	0.71381e+08
(W/parameters)	0.71381e+08
Calculated Error Variance	938.88
Standard Error Of The Regression	30.641
Correlation Coefficient	0.9999
Correlation Coefficient (W/parameters)	0.99990
Maximum Likelihood Objective Function	0.71534e+08
AIC Statistic	0.71534e+08
BIC Statistic	0.71534e+08
Hannan Statistic	0.71534e+08

Weighted least-squares objective function value of 0.71381e+08 is the same with and without parameters because no prior information was used. Given randomly distributed residuals and the same observations and weight matrix, a lower value of the least-squares objective function indicates a closer model fit to the data. This value for the transient inverse calibrated model will provide a basis for future comparisons. Smaller values of the calculated error variance for randomly distributed residuals are desirable. Values less than 1.0 (within user calculated confidence intervals: ns^2/X_U^2 ; ns^2/X_L^2) indicate that the model generally fits the data better than is consistent with the variances used to weight the observations and prior information, values greater than 1.0 indicate that the fit is worse. The standard error of the regression is the square root of the calculated error variance, and the same comments apply. The correlation coefficient with and without prior information is the correlation between weighted observed or prior information and simulated values. Correlation coefficient values below about 0.9 indicate poor model fit. This 0.9999 measure for the current regression would indicate a very good fit. The Maximum Likelihood objective function, the AIC, and BIC statistics for randomly distributed residuals can be used to compare one model with another where a lower absolute value indicate a better fitting model.

Another measure of model fit provided by UCODE uses ordered weighted residuals, which are the weighted residuals ordered smallest to largest. UCODE provides the correlation between ordered weighted residuals and normal order statistics. R_N^2 values above the critical value printed by UCODE

below indicate independent, normal weighted residuals. The value of 0.51 for this regression indicates the residuals are not independent normal weighted residuals.

The average weighted residual statistic of 0.44 and the correlation coefficient statistic of 0.9999 would indicate a good model fit. However, the other measures looking for patterns and trends in the residuals indicate the model fit is poor as indicated by the value of the calculated error variance, the 0.51 value for the correlation between ordered weighted residuals and normal order statistics which indicates the residuals are not independent, normal weighted residuals, and the runs test statistic which indicates trends are present in the spatially distributed weighted residuals. For a good model the residuals should be random. These measures, as indicated, will provide a baseline set of values for future reference as different conceptual models or conceptual model components are examined.

7.2.3 Evaluation of Optimized Parameter Values

Guideline 8 examines model error or fit by examining the optimized parameter values to determine if they are unrealistic and if the confidence intervals on the optimized values do not include reasonable values. The section evaluates the optimized parameter values through the discussions in the following subsections:

- Estimated Model Parameters
- Parameter-correlation coefficients
- Estimated confidence intervals in parameter estimates
- Comparison of model-parameter estimates with reasonable ranges.

7.2.3.1 Estimated Model Parameter Values

The optimal values of the parameters considered in the inverse calibration were calculated based on how well they reproduced the historical measurements of water levels during the period between 1943 and 1996. The objective function evaluated by the inverse calibration for the 1943 to 1996 period is the sum of the squared weighted residuals between simulated water levels and the approximately 76,000 water-level measurements made over the calibration period. The intent of the inverse-calibration procedure is to find the optimal combinations of parameters that minimize the objective function and to examine the various statistical measures that represent the quality of the regression.

Changes in the objective function during the calibration process (Figure 7.8) resulted in the optimal combination of parameters presented in Figures 7.9a-c. The final estimated parameters are summarized in Table 7.4. The inverse calibration of the selected model to the 1943–1996 period resulted in an overall reduction of the objective function (i.e., the sum of the squared residuals) by about an order of magnitude. The largest calculated changes to the base-parameter values associated with hydraulic properties (hydraulic conductivity and specific yield) were related to the hydraulic properties of the Ringold Formation (Unit 5). The largest changes to the parameters related to the boundary flux were associated with Rattlesnake Hills and Cold Creek model boundaries. The parameter value that changed the least was the multiplier for hydraulic conductivity of the Hanford Formation (Unit 1).

Changes in the ratio of the maximum composite scaled sensitivity coefficient divided by the composite scaled sensitivity coefficient for each parameter were monitored (Section 7.2.1) to identify the most sensitive and least sensitive parameters during the calibration process (Figure 7.7).

The combination of the low-sensitivity coefficients and the significant reduction in these parameter values during the calibration suggests that little confidence should be given to the final values for these parameters in the inverse. The significant reduction in the parameter associated with the Dry Creek boundary flux is an indication that changes in this relatively small boundary flux have little effect on simulated results. The reductions in the parameter associated with the hydraulic conductivity of the Ringold Formation (Unit 9) also suggest the same, but it is more likely that there is insufficient information in the observation dataset (water levels in wells completed in Unit 9) to allow the inverse procedure to effectively estimate the particular parameter value. The final inverse simulations made while fixing these values at their prior estimated values as can be seen in Table 7.3 had little effect on new estimates of all the other parameters considered in the model. All subsequent iterations used estimates made with the prior model.

Correlation Between Ordered Weighted Residuals
and Normal Order Statistics = 0.51
(Calculated using Eq. 38 of Hill 1992 or Eq. 23 of Hill 1998)

COMMENTS ON THE INTERPRETATION OF THE CORRELATION BETWEEN WEIGHTED RESIDUALS AND NORMAL ORDER STATISTICS:

Generally, IF the reported CORRELATION is GREATER than the critical value, at the selected significance level (usually 5 or 10%), the hypothesis that the weighted residuals are INDEPENDENT AND NORMALLY DISTRIBUTED would be ACCEPTED. HOWEVER, in this case, conditions are outside of the range of published critical values, as discussed below.

The sum of the number of observations and prior information items is 76035 which is greater than 200, the maximum value for which critical values are published. Therefore, the critical values for the 5 and 10% significance levels are greater than 0.987 and 0.989, respectively.

CORRELATIONS GREATER THAN these critical values suggest that probably the weighted residuals ARE INDEPENDENT AND NORMALLY DISTRIBUTED.

Correlations LESS THAN these critical values clearly indicate that we CAN REJECT the hypothesis.

The Kolmogorov-Smirnov test can be used to further evaluate the residuals.

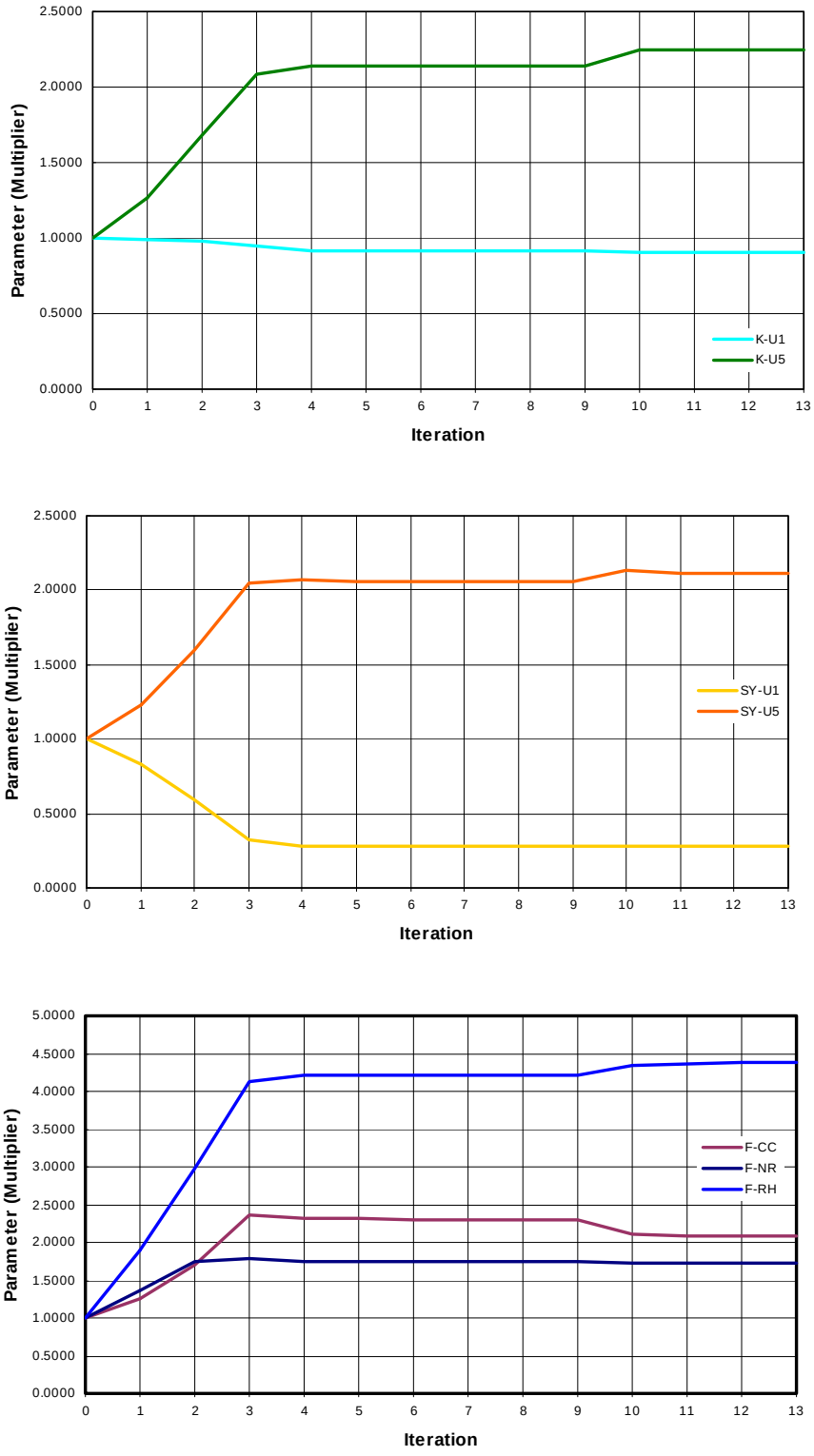

Figure 7.9. Parameter Changes During the Calibration Process

Table 7.4. Summary of Scaled-Parameter Estimates Derived for the Best-Fit Inverse Model

Model Parameter Factors	Scaled- Parameter Estimates
Hydraulic conductivity distributions within	
Hanford Formation (Unit 1)	0.901
Ringold Formation (Unit 5)	2.25
Ringold Formation (Unit 7)	Not estimated
Ringold Formation (Unit 9)	Not estimated
Specific yield of the	
Hanford Formation (Unit 1)	0.278
Ringold Formation (Unit 5)	2.12
Specific storage values of the	
Hanford Formation (Unit 1)	Not estimated
Ringold Formation (Unit 5)	Not estimated
Ringold Formation (Unit 7)	Not estimated
Ringold Formation (Unit 9)	Not estimated
Boundary Fluxes	
Cold Creek Valley	2.09
Dry Creek Valley	Not estimated
Rattlesnake Hills	4.37
Surface natural recharge	1.71
Anisotropy of all hydrogeologic units	Not estimated

7.2.3.2 Parameter-Correlation Coefficients

One output of the inversing process as implemented in UCODE is the correlation coefficients of all parameters estimated in the process. In UCODE, correlation coefficients are defined as the covariance between two parameters divided by the product of their standard deviations (Hill 1998). According to Hill (1998), correlation coefficients range from -1.0 to 1.0, with values very close to -1.0 and 1.0, indicative of parameter values that cannot be uniquely estimated with the observations used in the regression. As a diagnostic output for evaluating potential significant correlation coefficients, UCODE identifies parameter pairings with correlation coefficients between 0.85 and 0.9; 0.9 and 0.95; and greater than 0.95. Guidance in training materials for UCODE (Hill et al. 1999) and in previous work described in Hill et al. (1998) has suggested that parameter correlations in these ranges, particularly those in excess of 0.95, may be indicative of problematic strong correlation between parameter pairs. Strong correlation between two or more parameter values would bring into question the optimized estimates for the parameters in question. The parameter values, in question, would need some additional testing to the uniqueness of the optimized values estimated.

The correlation-coefficient matrix corresponding to the parameters estimated in the transient inverse calibration described in this section is summarized in Table 7.5. This summary indicates that no strong correlation was found among the parameter values estimated in the inversing process.

Table 7.5. Parameter Cross-Correlation Matrix Plot (best-fit inverse model)

Parameter	K-U1	K-U5	SY-U1	SY-U5	F-CC	F-NR	F-RH
K-U1	**1.000**	**0.309**	0.074	0.172	0.148	**0.623**	0.072
K-U5	**0.309**	**1.000**	**0.256**	-0.165	**0.288**	**0.456**	0.030
SY-U1	0.074	**0.256**	**1.000**	-0.083	-0.116	**0.329**	0.127
SY-U5	0.172	-0.165	-0.083	**1.000**	**0.605**	**0.270**	-0.158
F-CC	0.148	**0.288**	-0.116	**0.605**	**1.000**	-0.110	**0.353**
F-NR	**0.623**	**0.456**	**0.329**	**0.270**	-0.110	**1.000**	**-0.521**
F-RH	0.072	0.030	0.127	-0.158	**0.353**	**-0.521**	**1.000**

K-U1: Hanford (Unit 1) Hydraulic Conductivity Multiplier
K-U5: Ringold (Unit 5) Hydraulic Conductivity Multiplier
K-U9: Ringold (Unit 9) Hydraulic Conductivity Multiplier
SY-H: Hanford (Unit 1) Specific Yield Multiplier
SY-R(U5): Ringold (Unit 5) Specific Yield Multiplier
F-CC: Cold Creek Valley Flux Multiplier
F-NR: Natural Recharge Multiplier
F-RH: Rattlesnake Hills Flux Multiplier.

7.2.3.3 Confidence Intervals on Estimated Parameters

Another important output of the inversing process using UCODE is the calculation of confidence intervals around final parameter estimates. Technically, the confidence interval around a parameter estimate is a range that has a stated probability of containing the true value of the parameter. Narrow confidence intervals around a parameter estimate are a general indication or measure of the level of precision in the parameter estimate, given the types of observations used in the regression analysis. Narrow intervals imply greater precision; large intervals indicate less confidence in the optimized parameter value.

A summary of 95-percent confidence intervals and the estimated value for each parameter considered in the inverse analysis is summarized in Table 7.6. Calculated confidence intervals around all of the parameters exhibited a relatively narrow range, suggesting a good level of precision in the final parameter estimates.

Table 7.6. Scaled Parameter Estimates and Confidence Intervals (best-fit inverse model)

Model Parameters	95 % Confidence Interval	Best Fit Estimate	-95 % Confidence Interval
Hanford (Unit 1) K	0.91	0.90	0.89
Ringold (Unit 5) K	2.27	2.25	2.23
Hanford (Unit 1) SY	0.29	0.28	0.27
Ringold (Unit 5) SY	2.15	2.12	2.08
Cold Creek Flux	2.12	2.09	2.05
Natural Recharge	1.78	1.71	1.65
Rattlesnake Hills Flux	4.48	4.37	4.27

7.2.3.4 Comparison of Estimated Parameters with Reasonable Ranges

This section discusses the reasonableness of each of the estimated parameters.

Hydraulic Conductivities

Revised distributions of the hydraulic conductivities similar to those presented in Section 3.1.4 (see Figures 3.5, 3.6b, and 3.7b) but based on the scaling of the properties to best-fit multipliers given for Unit1 1 and 5 in Table 7.6 are presented in Figures 7.10 and 7.11(a,b) to illustrate the changes in properties of the Hanford and the uppermost Ringold sediments resulting from the inverse calibration. The most notable changes are seen in the distribution of hydraulic conductivities of the Ringold Formation (Unit 5), which outcrops at the water table over much of the western and southwestern part of the Site.

The best-fit estimates for the hydraulic conductivity of the Hanford Formation (Unit 1) were found to be well within the reasonable range of previous estimates for this parameter. Previous work summarized in Thorne and Newcomer (1992) and Wurstner et al. (1995) indicate that the hydraulic conductivity of Unit 1 generally ranges from about 1 to 1,000,000 m/d and is much higher than any of the other units that make up the unconfined aquifer system. Aquifer tests indicate that the minimum estimated hydraulic conductivity is about 1 m/d (Thorne et al. 1993), and the maximum estimated value is about 10,000 m/d (Thorne and Newcomer 1992; DOE 1988). However, the maximum hydraulic conductivity that can be estimated by an aquifer test is limited by the well efficiency and the flow rate that can be pumped with available equipment. Past calibration efforts by Wurstner et al. (1995) and Cole et al. (1997) have estimated that an upper limit of hydraulic conductivity for coarse-gravel flood deposits found in the central part of the Hanford Site is on the order of several tens of thousands of m/d. Estimates resulting from inverse calibration of the model were very similar to estimates in previous calibration efforts and are consistent with estimates developed largely from aquifer tests.

The best-fit estimates for the hydraulic conductivity of the Ringold Formation (primarily Unit 5) were found to be well within the reasonable range of previous estimates for this parameter. The Ringold Formation consists of sand to muddy sandy gravel with varying degrees of consolidation or cementation. Unit 5 is the most widespread unit within the unconfined aquifer and is found below the water table across most of the model region. According to Wurstner et al. (1995), hydraulic conductivities of Units 5, 7, and 9 determined from aquifer tests vary within the range of about 0.1 to 200 m/d. Because these units

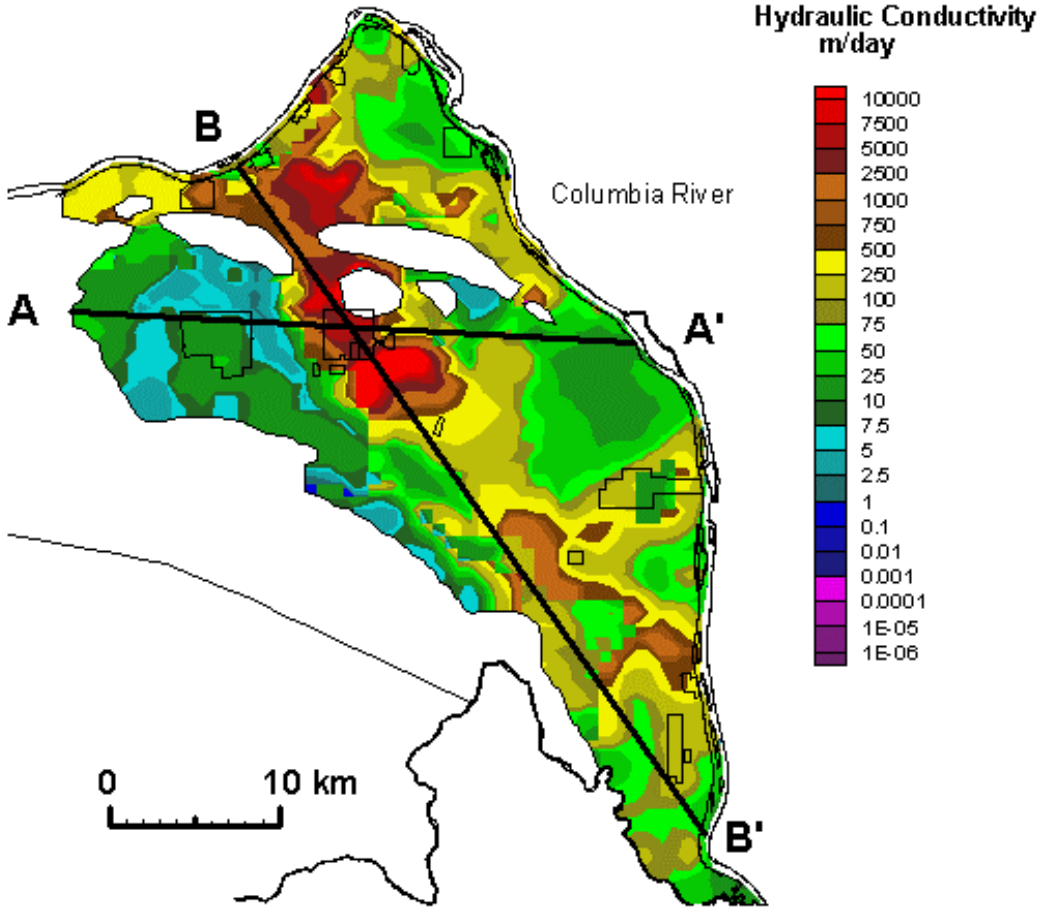

Figure 7.10. Hydraulic Conductivity Distribution from Best-Fit Inversed Model

are hydrologically similar, they were grouped together in areas where the intervening mud units do not exist. A few aquifer tests suggest that vertical anisotropy is in the range of 0.01 to 0.1. Therefore, the range of Kv is estimated at about 0.001 to 20 m/d. Estimates resulting from inverse calibration of the model were very similar to estimates in previous calibration efforts and are consistent with estimates developed largely from aquifer tests.

Storage Properties

According to Wurstner et al. (1995) and Thorne and Newcomer (1992), specific yield for the Hanford Formation (Unit 1) is estimated to range from about 0.1 to 0.3 and is expected to be higher for coarse, well-sorted gravel than for poorly sorted mixtures of sand and gravel. The best-fit estimate and associated confidence intervals for the specific yield of the Hanford Formation were below this range of reasonable values. From previous work (Wurstner et al. 1995; Thorne and Newcomer 1992), specific yields of the poorly sorted sediments of the Ringold formation are estimated to range from 0.05 to 0.2.

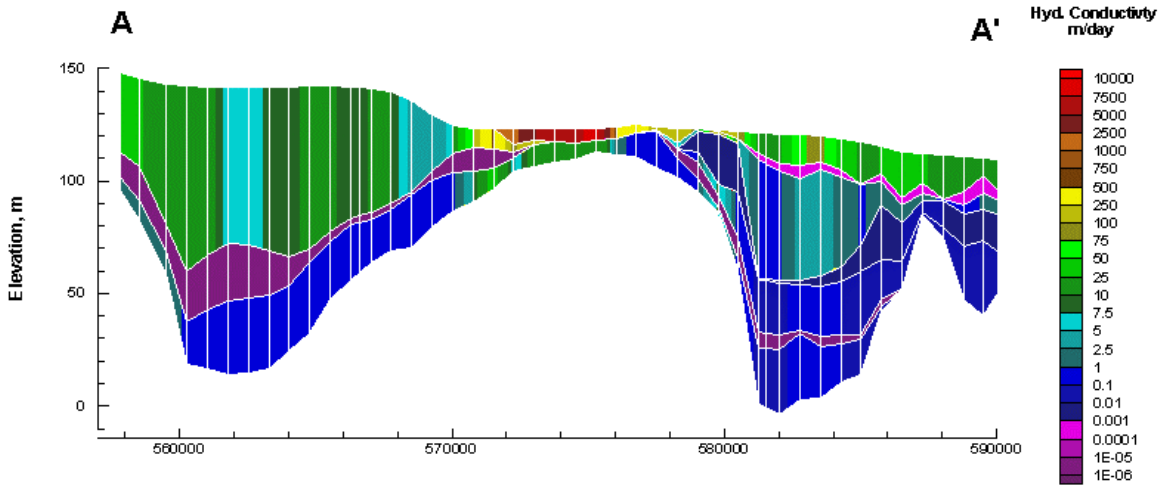

Figure 7.11a. Cross Section A-A' Showing Distribution of Hydraulic Conductivity

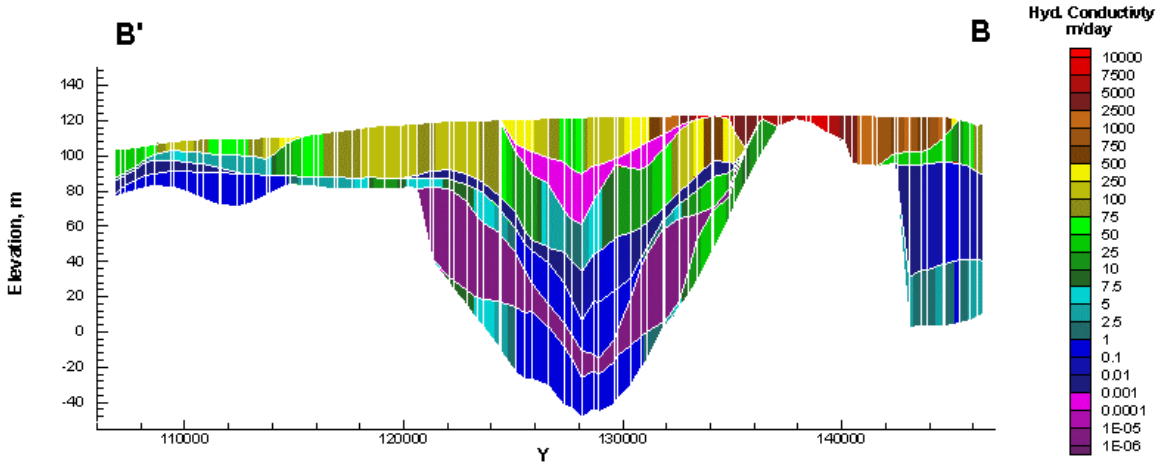

Figure 7.11b. Cross Section B-B' Showing Distribution of Hydraulic Conductivity

The best-fit estimate for the specific yield of the Hanford Formation (Unit 1) and the Ringold Formation (Unit 5) and associated confidence limits developed by the inverse calibration were not found to be consistent with reasonable ranges summarized in previous analyses of the hydraulic properties of these units. The best-fit estimate and associated confidence intervals for the specific yield of the Ringold Formation (Unit 5) were also found not to be consistent and slightly above the range of previously available estimates. These inconsistencies between the calibrated values and the prior information for both the Hanford Formation and the Ringold Formation suggest that processes other than those considered in the numerical implementation of the conceptual model may need to be evaluated in future calibration efforts.

One important component of the current understanding of the unconfined aquifer system that is not considered in the numerical model evaluated is the interaction of the unconfined aquifer system with the underlying uppermost confined aquifers of the Columbia River. Two specific aspects of this interaction may be very important to more accurately estimate model parameters from simulation of the unconfined aquifer system's behavior during the period of Hanford operations.

Downward leakage into the uppermost basalt-confined aquifer has likely occurred in the 200 Area plateau area during the historical period of Hanford operational waste-water discharges as groundwater mounds built up in the unconfined aquifer in response to the large volume discharges to surface facilities. Not considering this leakage down into the uppermost basalt-confined aquifers in the current model results in the over prediction of water transmitted laterally within the unconfined aquifer system. Not considering this interaction would likely have a large effect in estimating the position of the water table, the lateral hydraulic gradients, and the hydraulic properties (hydraulic conductivity and specific yield) in the vicinity of the artificial discharge areas. The current modeling efforts have shown that sensitivity to changes in the Ringold Formation properties is higher because of the significant mound building that has occurred within the Ringold Formation from past discharges in the 200 West Area.

Including estimates of regional discharge from the basalts into the unconfined aquifer system would also likely improve the ability of the current model to predict the behavior of the unconfined aquifer system between artificial discharge areas in the central plateau and the Columbia River. In these areas, flow from the basalt regional system is generally postulated to discharge upward into the unconfined aquifer system before it is discharged into the Columbia River. The current model does not account for this increased flow downgradient from the central plateau region and, as a result, likely underestimates the total amount of lateral flow moving through the unconfined aquifer in these regions. Evidence of these low predictions of the water table in areas between the central plateau and the Columbia River is quite evident in the series of residual head plots, provided in Appendix D, using the prior model estimates for hydraulic property parameters.

Our current interpretation of the estimates for the inverse model parameters is that the current model attempts to compensate for this lack of upward basalt leakage by reducing the storage properties (e.g., specific yield) of the Hanford Formations with the objective of moving additional water from upgradient areas to raise the predicted water levels between the central plateau and the Columbia River. The residual plots of head at observation well locations in these downgradient areas, as illustrated in the series of plots in Appendix E, certainly indicate that this change in the specific yield for the Hanford Formation has improved the overall fit of the model to the predicted water table in these downgradient areas. However, this estimated value of specific yield is not consistent with the general current understanding of hydraulic characteristics and the specific yields that would be expected for the Hanford Formation.

A separate inverse calibration was performed that held the specific yields of the Hanford and Ringold Formation at 0.25 and 0.1 to investigate the effect of putting these parameters in a more reasonable range on estimates of the hydraulic conductivities of the same units. Results of this calibration effort were unsuccessful in converging to a solution. This evidence is suggestive that the conceptual model being evaluated is incomplete and that other processes, such as intercommunication with the underlying basalt-confined aquifer, may be needed for the inverse calibration procedure to estimate these parameter values within the range of prior knowledge for these parameters.

Boundary Flux Estimates

The estimated fluxes developed from a constant head boundary during previous calibration efforts provide the basis for the initial estimate used in this inverse. While the various boundary fluxes at the Cold Creek and Dry Creek Valleys and at the base of Rattlesnake Hills cannot be independently verified, the estimated fluxes were found to be well within the reasonable range, given the uncertainty that could

be expected in hydraulic properties of the principal hydrogeologic units found close to these units. The best-fit flux at Cold Creek Valley was a little over a factor of 2 higher than the initial estimates. The best-fit flux estimated for Rattlesnake Hills was about a factor 4 higher than the initial estimate. Given the uncertainty in the overall hydraulic properties of the Ringold Formation found in the vicinity of Cold Creek Valley and the hydraulic properties of the Hanford and Ringold Formations found along the base of Rattlesnake Hills, the increase in the overall estimate is not considered unreasonable.

The range in confidence limits for the best-fit flux from Dry Creek Valley was very large, suggesting that either this flux is not a very sensitive parameter in the model or that insufficient information is available (or both) in the observational database to effectively estimate this flux using the inverse method.

Natural Recharge

The best-fit estimates for the natural recharge were about 71 percent greater than the estimates previously made by Fayer and Walters (1995). This increase in the overall estimate of recharge is not considered unreasonable for most areas of the site where recharge rates are on the order of 5 to 20 mm/yr. However, for some areas of the site where coarse soils exist and previous estimated recharge rates approach 50 to 60 mm/yr, this 71-percent increase in the overall recharge rate results in rates of 80 to 90 mm/yr, which are not considered reasonable. The best-fit estimate of recharge for this conceptual model may be higher to compensate for a variety of factors, including

- underestimating previous estimates of regional natural recharge
- underestimating artificial discharges at waste-water facilities in the operational areas due to
    - reporting errors
    - inadvertent losses of water within the Hanford Site infrastructure between points of withdrawal from the Columbia River to points of discharges within the operational areas
- not considering the interaction of the uppermost basalt confined aquifer in the current model leading to underestimates of flow into the unconfined aquifer system in areas along the Columbia River where the regional basalt aquifer is likely discharging upward into the unconfined system.

8.0 Summary and Conclusions

The inverse calibration methodology based on linking the Hanford Site Groundwater model with UCODE will provide a good working tool for evaluating alternative interpretations of components of the Site conceptual models. UCODE is a universal inverse modeling code developed jointly by the U.S. Geological Survey (USGS) and International Groundwater Modeling Center (IGWMC) of the Colorado School of Mines.

The transient inverse calibration uses over 76,000 water-level measurements taken in about 1200 wells at the Hanford Site since the mid 1940s. Because of the long simulation run times and large number of simulations required a serial computational approach for the coupled flow and transport inverse was estimated to require a year or more of computational effort. As a result, an innovative parallel computational approach that uses an isolated network of 23 computers was developed. The approach uses a recently developed parallel version of UCODE that communicates with a parallel task manager to propagate the multiple simulation tasks (i.e., the forward model runs) for simultaneous computation on an isolated network of dedicated computers. In addition, a customized version of the forward model code CFEST was developed to simplify the specification of inverse model parameters and the input of large numbers of observations.

The existing consolidated site-wide groundwater model (referred to as the prior model in this report) was calibrated using 1979 data and a steady state inverse approach in conjunction with additional trial and error transient model calibration runs using estimates of artificial discharges and a limited set of representative head observations between 1979 and 1996. The conceptual model for this prior model also provided the conceptual basis for the initial three-dimensional transient inverse modeling studies discussed in this report. The extended (1943–1996) calibration period dataset for the initial transient inverse modeling consisted of newly prepared estimates of the artificial discharges and river stage variations before 1979 to supplement the 1979–1996 data set prepared for the prior model and a complete set of head observations (1943–1996) as opposed to the limited set of observations (1979–1996) used in the prior model calibration efforts. As part of the process of evaluating and testing these newly developed artificial discharge estimates and data on head observations (1943–1996), the prior model was used to simulate this time period. This simulation using the prior model with the parameters determined from the previous calibration provided some good insights into the capability of the prior model to duplicate overall historical trends in water-table changes and, in particular, groundwater mound building and decline during the entire period of Hanford operations. Results of these preliminary simulations with the prior model indicate it was not a good predictor of the water-table configuration before 1979. While the prior model was generally capable of replicating overall trends over most of the site, using the prior model and its associated parameter estimates led to a significant over-prediction of the historical growth and decline of groundwater mounds in the 200 West Area. Simulated values of heads in this area were between 10 to 15 m higher than were observed in wells during the period of maximum discharges and mound building near discharge facilities during the 1950s, 1960s, and 1970s.

Recalibrating the prior model using the transient inverse calibration procedures in UCODE significantly improved the capability of the model to simulate historical trends in water-table changes over the entire site, but some parameters took on unrealistic values indicating that parameter zonation and/or

conceptual model improvements are needed. The improved inversing methodology provided in UCODE and the additional information provided in the longer period of calibration (i.e., 1943–1996) were both contributing factors that improved the ability of the recalibrated prior model to simulate historical trends in water-table changes. All goodness-of-fit measures were significantly improved over measures of fit for the previously calibrated prior model for the same simulation period. The most noteworthy improvement by the transient inverse calibrated prior model is its capability to fit historical trends of water-table changes and mound building observed near major discharge facilities in the 200 West Area. This improvement in the overall fit resulted in improved statistical performance in all categories (mean residual, range of residuals, and sum of squared residuals) for the 1943–1996 period of calibration.

The majority of changes in the parameter estimates derived from the transient inverse calibration of the prior conceptual model using the UCODE methodology produced new estimates and linear confidence intervals consistent with prior knowledge of these parameters. However, estimates for the specific yield of the Hanford Formation (between 0.06 to 0.07) and the Ringold Formation (between 0.20 and 0.21) from the three-dimensional transient inverse calibration were not consistent with current understanding of the characteristics of these sediments. The unrealistic estimates for these two parameters combined with the increased estimates from the inversing for natural recharge as well as boundary fluxes for Cold Creek Valley and Rattlesnake Hills Springs indicate that the conceptual basis of the prior model is incomplete. This initial three-dimensional transient inverse modeling study indicates that other conceptual model components not considered in the prior model are needed to approximate historical aquifer system behavior. The increased estimates of natural recharge resulting from this initial study, for example, are believed to indicate that recharge to the unconfined aquifer system of the prior model resulting from intercommunication of the unconfined aquifer system with the uppermost confined aquifer associated with the Columbia River Basalt is on the same order of magnitude as natural recharge. As a result, this is the first major alternative conceptual model that will be evaluated as studies of alternative conceptual-models are undertaken.

In summary, a three-dimensional transient inverse model approach for the estimation of site-wide scale flow and transport parameters, including their uncertainties, using data on the transient behavior of the unconfined aquifer system over the entire historical period of Hanford operations, has been developed and applied. The initial application of this newly developed methodology considered only the conceptual basis embodied in the prior site-wide model, and only the flow model and corresponding data set were used in this three-dimensional transient inverse calibration effort. Subsequent efforts will examine both flow and transport. Comparisons of the goodness of fit measures for the newly calibrated model with those for the prior model illustrate that the new three-dimensional transient inverse model approach will strengthen the technical defensibility of the final calibrated model(s) and provide the ability to incorporate uncertainty in model predictions. These initial results, however, indicate that improvements to the conceptual model framework of the prior model are required. Studies are under way to implement and test these improvements.

9.0 References

Ahlstrom SW, HP Foote, RC Arnett, CR Cole, and RJ Serne. 1977. *Multicomponent Mass Transport Model: Theory and Numerical Implementation* (Discrete-Parcel-Random-Walk Version). BNWL-2127, Battelle Northwest, Richland, Washington.

Anderman ER, MC Hill, and EP Poeter. 1996. "Two-dimensional advective transport in groundwater flow parameter estimation." *Ground Water*, 34(6):1001–1009.

Beck JV and KJ Arnold. 1977. *Parameter Estimation in Engineering and Science*. John Wiley & Sons, New York.

Bierschenk WH. 1959. *Aquifer Characteristics and Ground-Water Movement at Hanford*, HW-60601. General Electric Company, Hanford Atomic Products Operation, Richland, Washington.

Carrera J and SP Neuman. 1986. "Estimation of aquifer parameters under transient and steady-state conditions." *Water Resour. Res.*, 22(2):199–242.

Cearlock DB, KL Kipp, and DR Friedrichs. 1975. *The Transmissivity Iterative Calculation Routine - Theory and Numerical Implementation.* BNWL-1706, Battelle Northwest, Richland, Washington.

Chamness MA and JK Merz. 1993. *Hanford Wells*. PNL-8800, Pacific Northwest Laboratory, Richland, Washington.

Cole CR, FW Bond, SM Brown, and GW Dawson. 1984. *Demonstration/Application of Ground-Water Modeling Technology for Evaluation of Remedial Action Alternatives*. Contract 68-03-3116, Municipal Environmental Research Laboratory, U. S. Environmental Protection Agency, Cincinnati.

Cole CR, SB Yabusaki, and CT Kincaid. 1988. *CFEST-SC, Coupled Fluid, Energy, and Solute Transport Code, SuperComputer Version, Documentation and User's Manual*. Battelle, Pacific Northwest Laboratories, Richland, Washington.

Cole CR, SK Wurstner, MP Bergeron, MD Williams, and PD Thorne. 1997. *Three-Dimensional Analysis of Future Groundwater Flow Conditions and Contaminant Plume Transport in the Hanford Site Unconfined Aquifer System: FY 1996 and 1997 Status Report*. PNNL-11801, Pacific Northwest National Laboratory, Richland, Washington.

Connelly MP. 1994. *Capture Zone Analyses for the 200-ZP-1 and 200-UP-1 Pilot Scale Pump-and-Treat Tests*. WHC-SD-EN-TI-252, Westinghouse Hanford Company, Richland, Washington.

Connelly MP, BH Ford, and JV Borghese. 1992a. *Hydrogeologic Model for the 200 West Groundwater Aggregate Area*. WHC-SD-EN-TI-014, Westinghouse Hanford Company, Richland, Washington.

Connelly MP, BH Ford, and JW Lindberg. 1992b. *Hydrogeologic Model for the 200 East Groundwater Aggregate Area*. WHC-SD-EN-TI-019, Westinghouse Hanford Company, Richland, Washington.

Cooley RL. 1977. "A method of estimating parameters and assessing reliability for models of steady-state groundwater flow, 1, Theory and numerical properties." *Water Resources Research*, 13(2):318–324.

Cooley RL and RL Naff. 1990. *Regression modeling of groundwater flow*. U.S. Geological Survey Techniques in Water-Resources Investigations, Book 3, Ch. B4, p. 232.

Cooley RL. 1993. "Regression modeling of ground-water flow." Supplement 1 – *Modifications to the computer code for nonlinear regression solution of steady-state ground-water flow problems*. U.S Geological Survey Techniques of Water Resources Investigations, Book 3, Ch. B4, Supp. 1, p.8.

D' Agnese FA, CC Faunt, AK Turner, and MC Hill. 1998. *Hydrogeologic evaluation and numerical simulation of the Death Valley Regional Groundwater flow system, Nevada and California*. U.S. Geological Survey Water-Resource Investigation Report 96-4300, Water Resource Division, Denver.

Deju RA. 1974. *The Hanford Field Testing Program*. ARHC-00004-201, Atlantic Richfield Hanford Company, Richland, Washington.

Devary JL. 1987. *The CFEST-INV Stochastic Hydrology Code: Mathematical Formulation, Application, and User's Manual*. ICF Northwest, Richland, Washington.

DOE (see U.S. Department of Energy)

Doherty J. 1994. *PEST: Corinda, Australia*. Watermark Computing, p.122.

Dove FH, CR Cole, MG Foley, FW Bond, RE Brown, WJ Deutsch, MD Freshley, SK Gupta, PJ Gutknecht, WL Kuhn, JW Lindberg, WA Rice, R Schalla, JF Washburn, and JT Zellmer. 1982. *AEGIS Technology Demonstration for a Nuclear Waste Repository in Basalt*. PNL-3632, Pacific Northwest Laboratory, Richland, Washington.

Ebbert JC, SE Cox, BW Drost, and KM Shurr. 1993. *Distribution and Sources of Nitrates, and Presence of Fluoride and Pesticides, in Parts of the Pasco Basin, Washington, 1986-88*. Water Resources Investigations Report 93-4197, U.S. Geological Survey, Tacoma, Washington.

Ecology 1989. *Hanford Federal Facility Agreement and Consent Order Between the U.S. Environmental Protection Agency, the U.S. Department of Energy, and the State of Washington Department of Ecology*. May 15, 1989, as amended. Washington State Department of Ecology (Ecology), U.S. Environmental Protection Agency (EPA), and U.S. Department of Energy (DOE), Olympia, Seattle, and Richland, Washington.

Evans JC, DI Dennison, RW Bryce, PJ Mitchell, DR Sherwood, KM Krupka, NW Hinman, EA Jacobson, and MD Freshley. 1988. *Hanford Site Groundwater Monitoring for July Through December 1987*. PNL-6315-2, Pacific Northwest Laboratory, Richland, Washington.

Fayer MJ and TB Walters. 1995. *Estimated Recharge Rates at the Hanford Site*. PNL-10285, Pacific Northwest Laboratory, Richland, Washington.

Foley MG, DJ Bradley, CR Cole, JP Hanson, KA Hoover, WA Perkins, and MD Williams. 1995. *Hydrogeology of the West Siberian Basin and Tomsk Region*. PNL-10585, Pacific Northwest Laboratory, Richland, Washington.

Forsythe GE and EG Strauss. 1955. "On best conditioned matrices." *Proceedings of American Mathematical Society* 10(3):340–345.

Freshley MD and MJ Graham. 1988. *Estimation of Groundwater Travel Time at the Hanford Site: Description, Past Work, and Future Needs*. PNL-6328, Pacific Northwest Laboratory, Richland, Washington.

Friedrichs DR, CR Cole, and RC Arnett. 1977. *Hanford Pathline Calculational Program—Theory, Error Analysis and Applications*. ARH-ST-149, Atlantic Richfield Hanford Company, Richland, Washington.

Gale J, R Macleod, J Welhan, CR Cole, and LW Vail. 1987. *Hydrogeological Characterization of the Stripa Site*. Technical Report 87-15, SKB, Stockholm, Sweden.

Gee GW, MJ Fayer, ML Rockhold, and MD Campbell. 1992. "Variations in Recharge at the Hanford Site." *NW Sci.* Vol. 66, pp. 237–250.

Graham MJ, MD Hall, SR Strait, and WR Brown. 1981. *Hydrology of the Separations Area*. RHO-ST-42, Rockwell Hanford Operations, Richland, Washington.

Gupta SK. 1997. *Draft User's Manual, CFEST-96 Flow and Solute Transport, Constant/Variable Density, Computationally Efficient, and Low Disk PC/Unix Version*. Consultant for Environmental System Technologies, Irvine, California.

Gupta SK, CR Cole, CT Kincaid, and AM Monti. 1987. *Coupled Fluid, Energy, and Solute Transport (CFEST) Model: Formulation and User's Manual*. BMI/ONWI-660, Battelle Memorial Institute, Columbus, Ohio.

Hartman MJ and KA Lindsey. 1993. *Hydrogeology of the 100-N Area, Hanford Site, Washington.* WHC-SD-EN-EV-027, Westinghouse Hanford Company, Richland, Washington.

Hartman MJ and PE Dresel (eds.). 1997. *Hanford Site Groundwater Monitoring for Fiscal Year 1996.* PNNL-11470, Pacific Northwest National Laboratory, Richland, Washington.

Hill MC. 1992. *A computer program (MODFLOWP) for estimating parameters of a transient, three-dimensional, groundwater flow model using nonlinear regression.* U.S. Geological Survey Open-File Report 91-48, p. 358.

Hill MC. 1994. *Five Computer Programs For Testing Weighted Residuals And Calculating Linear Confidence and Prediction Intervals on Results from the Groundwater Computer Program MODFLOWP,* U.S. Geological Survey Open-File Report 93-481, p. 81.

Hill MC. 1998. "Methods and Guidelines for Effective Model Calibration." *Water Resource Investigations*. Report 98-4005, U. S. Geological Survey, Water Resources Division, Denver.

Hill MC, RL Cooley, and DW Pollack. 1998. "A Controlled Experiment in Ground-water Flow Model Calibration." *Groundwater* 36(3):520-535.

Hill MC, RL Cooley, E Poeter, and C Teideman. 1999. *Ground-Modeling Short Courses: Calibration and Uncertainty in Groundwater Models*. International Groundwater Modeling Center, Colorado School of Mines, Golden.

Isaaks EH and RM Srivastava. 1989. *An Introduction to Applied Geostatistics*. Oxford University Press, New York.

Jacobson EA. 1985. *A Statistical Parameter Estimation Method Using Singular Value Decomposition with Application to Avra Valley Aquifer in Southern Arizona*. Dissertation, Department of Hydrology and Water Resources, University of Arizona, Tucson.

Jacobson EA and MD Freshley. 1990. *An Initial Inverse Calibration of the Groundwater Model for the Hanford Unconfined Aquifer*. PNL-7144, Pacific Northwest Laboratory, Richland, Washington.

Kipp KL and RD Mudd. 1973. *Collection and Analysis of Pump Test Data for Transmissivity Values.* BNWL-1709, Battelle Northwest, Richland, Washington.

Kipp KL, AE Reisenauer, CR Cole, and LA Bryan. 1972. *Variable Thickness Transient Groundwater Flow Model: Theory and Numerical Implementation*. BNWL-1703, Battelle Northwest, Richland, Washington.

Levenberg K. 1944. "A Method for the Solution of Certain Non-Linear Problems in Least Squares." *Quart. Appl. Math* Vol. 2, pp 164-168.

Liikala TL. 1994. *Hydrogeology Along the Southern Boundary of the Hanford Site Between the Yakima and Columbia Rivers, Washington*. PNL-10094, Pacific Northwest Laboratory, Richland, Washington.

Liikala TL and RL Aaberg. 1988. *Geohydrologic Characterization of the Area Surrounding the 183-H Solar Evaporation Basins*. PNL-6728, Pacific Northwest Laboratory, Richland, Washington.

Lindberg JW. 1993a. *Geology of the 100-B/C Area, Hanford Site, South-Central Washington*. WHC-SD-EN-TI-133, Westinghouse Hanford Company, Richland, Washington.

Lindberg JW. 1993b. *Geology of the 100-K Area, Hanford Site, South-Central Washington*. WHC-SD-EN-TI-155, Westinghouse Hanford Company, Richland, Washington.

Lindberg JW and FW Bond. 1979. *Geohydrology and Ground-Water Quality Beneath the 300 Area, Hanford Site, Washington*. PNL-2949, Pacific Northwest Laboratory, Richland, Washington.

Lindsey KA. 1991. *Revised Stratigraphy for the Ringold Formation, Hanford Site, South-Central Washington*. WHC-SD-EN-EE-004 Rev. 0, Westinghouse Hanford Company, Richland, Washington.

Lindsey KA. 1992. *Geology of the Northern Part of the Hanford Site: An Outline of Data Sources and the Geologic Setting of the 100 Areas*. WHC-SD-EN-TI-011 Rev. 0, Westinghouse Hanford Company, Richland, Washington.

Lindsey KA and GK Jaeger. 1993. *Geologic Setting of the 100-HR-3 Operable Unit, Hanford Site, South-Central Washington*. WHC-SD-EN-TI-132, Westinghouse Hanford Company, Richland, Washington.

Lindsey KA, BN Bjornstad, and MP Connelly. 1991. *Geologic Setting of the 200 West Area: An Update*. WHC-SD-EN-TI- 008, Westinghouse Hanford Company, Richland, Washington.

Lindsey KA, BN Bjornstad, JW Lindberg, and KM Hoffman. 1992. *Geologic Setting of the 200 East Area: An Update*. WHC-SD-EN-TI-012, Westinghouse Hanford Company, Richland, Washington.

Loaiciga HA and MA Marino. 1986. "Estimation and inference in the inverse problem." *Proceedings of Water Forum '86, World Water Issues in Evolution*, pp. 973–980. ASCE, Long Beach, California.

Marquardt DW. 1963. "An algorithm for least-squares estimation of nonlinear parameters." *J. Soc. Indust. Appl. Math.* 11(2):431-441.

Neuman SP. 1999. "Methodology to Identify and Evaluate Conceptual Models and Uncertainty Related to Groundwater Transport at Nuclear Facilities and Sites." Draft report for U.S. Nuclear Regulatory Commission presented at Research Symposium: Hydrologic Conceptual Model and Parameter Uncertainty, July 25-26, 2000, Rockville, Maryland.

Neuman SP and S Yakowitz. 1979. "A Statistical Approach to the Inverse Problem of Aquifer Hydrology, 1, Theory." *Water Resour. Res.* 15(4):845–860.

Peck A, S Gorelick, G de Marsily, S Foster, and V Kovalevsky. 1988. *Consequences of Spatial Variability in Aquifer Properties and Data Limitations for Groundwater Modeling Practice.* IAHS Publication No. 175, International Association of Hydrological Sciences, IAHS Press. Institute of Hydrology, Wallingford, Oxfordshire, United Kingdom.

Peterson RE. 1992. *Hydrologic and Geologic Data Available for the Region North of Gable Mountain, Hanford Site, Washington*. WHC-SD-EN-TI-006, Westinghouse Hanford Company, Richland, Washington.

Poeter EP and MC Hill. 1996. "Unrealistic parameter estimates in inverse modeling: a problem or a benefit for model calibration." *Proceedings of the ModelCARE 96 Conference*. International Association of Hydrological Sciences Publication no. 237, Golden, Colorado, p. 277-285.

Poeter EP and MC Hill. 1998. *Documentation of UCODE, A Computer Code for Universal Inverse Modeling*. U.S. Geological Survey Water Resources Investigations Report 98-4080.

Puget Sound Power and Light (PSPL). 1982. *Skagit/Hanford Nuclear Project, Preliminary Safety Analysis Report*. Appendix 20, Amendment 23, Puget Sound Power and Light Company, Bellevue, Washington.

Richmond MC, WA Perkins, and Y Chien. 2000a. *Numerical Model Analysis of System-Wide Dissolved Gas Abatement Alternatives*. Final Report, prepared for U.S. Army Corps of Engineers, Walla Walla District under Contract DACW68-96-D-0002, Battelle Pacific Northwest Division, Richland, Washington.

Richmond MC, WA Perkins, and Y Chien. 2000b. *Numerical Model Analysis of System-Wide Dissolved Gas Abatement Alternatives*. Appendix B. Model Configuration and Optimization. Final report prepared for U.S. Army Corps of Engineers, Walla Walla District. Battelle Pacific Northwest Division, Richland, Washington.

Richmond MC, WA Perkins, and Y Chien. 2000c. *Numerical Model Analysis of System-Wide Dissolved Gas Abatement Alternatives*. Appendix F. Verification of MASS1 for Lower Columbia/Snake Temperature and Total Dissolved Gas Simulation. Final report prepared for U.S. Army Corps of Engineers, Walla Walla District. Battelle Pacific Northwest Division, Richland, Washington.

Seber GAF and CJ Wild. 1989. *Nonlinear Regression*. John Wiley & Sons, New York.

Simmons CS, CT Kincaid, and AE Reisenauer. 1986. *A Simplified Model for Radioactive Contaminant Transport: The TRANSS Code*. PNL-6029, Pacific Northwest Laboratory, Richland, Washington.

Sun NZ. 1994. Inverse problems in groundwater modeling. Kluwer Academic Publishers, Boston, p. 337.

Swanson LC. 1992. *Phase 1 Hydrogeologic Summary of the 300-FF-5 Operable Unit, 300 Area*. WHC-SD-EN-TI-052, Westinghouse Hanford Company, Richland, Washington.

Swanson LC. 1994. *1994 Characterization Report for the State Approved Land Disposal Site*. WHC-SD-CO18H-RPT-003, Westinghouse Hanford Company, Richland, Washington.

Theil H. 1963. "On the use of incomplete prior information in regression analysis." *Amer. Stat. Assoc. J.* 58(302):401-414.

Thorne PD and MA Chamness. 1992. *Status Report on the Development of a Three-Dimensional Conceptual Model for the Hanford Site Unconfined Aquifer System*. PNL-8332, Pacific Northwest Laboratory, Richland, Washington.

Thorne PD and DR Newcomer. 1992. *Summary and Evaluation of Available Hydraulic Property Data for the Hanford Site Unconfined Aquifer System*. PNL-8337, Pacific Northwest Laboratory, Richland, Washington.

Thorne PD, MA Chamness, FA Spane Jr, VR Vermeul, and WD Webber. 1993. *Three-Dimensional Conceptual Model for the Hanford Site Unconfined Aquifer System, FY 93 Status Report*. PNL-8971, Pacific Northwest Laboratory, Richland, Washington.

Thorne PD, MA Chamness, VR Vermeul, QC MacDonald, and SE Schubert. 1994. *Three-Dimensional Conceptual Model for the Hanford Site Unconfined Aquifer System, FY 1994 Status Report*. PNL-10195, Pacific Northwest Laboratory, Richland, Washington.

U.S. Department of Energy (DOE). 1988. *Consultation Draft, Site Characterization Plan, Reference Repository Location, Hanford Site, Washington*. DOE/RW-0164, Vol. 1 and 2, Richland, Washington.

U.S. Department of Energy (DOE). 1991. *Description of Codes and Models to be Used in Risk Assessment*. DOE/RW-9144, Richland, Washington.

U.S. Department of Energy-Richland Operations (DOE-RL). 2000. *Selection and Review of a Site-Wide Groundwater Model at the Hanford Site*. DOE/RL-2000-11, Richland, Washington,

Walters WH, MC Richmond, and BG Gilmore. 1994. *Reconstruction of Radionuclide Concentrations in the Columbia River from Hanford, Washington to Portland, Oregon, January 1950–January 1971*. BNWD-2225 HEDR, Battelle Pacific Northwest Division, Richland, Washington.

Wurstner SK and JL Devary. 1993. *Hanford Site Ground-Water Model: Geographic Information System Linkages and Model Enhancements FY 1993*. PNL-8991, Pacific Northwest National Laboratory, Richland, Washington.

Wurstner SK and MD Freshley. 1994. *Predicted Impacts of Future Water Level Decline on Monitoring Wells Using a Groundwater Model of the Hanford Site.* PNL-10196, Pacific Northwest Laboratory, Richland, Washington.

Wurstner SK, PD Thorne, MA Chamness, MD Freshley, and MD Williams. 1995. *Development of a Three-Dimensional Groundwater Model of the Hanford Site Unconfined Aquifer System: FY 1995 Status Report*. PNL-10886, Pacific Northwest Laboratory, Richland, Washington.

Appendix A

CFEST Utilities for UCODE Application

Appendix A

CFEST Utilities for UCODE Application

Contents

Tables

CFEST Utilities for UCODE Application

A.1 Introduction

The Coupled Fluid, Energy, and Solute Transport (CFEST) computer code is a three-dimensional finite element code that simulates coupled fluid, energy, and solute transport in groundwater systems. Documentation, including the theoretical basis and implementation, and guidance on its use, is provided in Gupta et al. (1987), Cole et al. (1988) and Gupta (1997). Recently, the most current version of CFEST was supplemented with additional utilities to facilitate the interface of CFEST with UCODE, a computer code for Universal Inverse Modeling (Poeter and Hill 1998). UCODE provides an efficient means for parameter optimization and quantitative uncertainty analysis. This appendix briefly describes CFEST-UCODE interface tools, a brief user's guide, and some example data files used in the preliminary application of UCODE to the Hanford Site aquifer system.

A.1.1 Overview of UCODE—A Computer Code for Universal Inverse Modeling

UCODE performs inverse modeling, posed as a parameter-estimation problem, by calculating parameter values that minimize a weighted least-squares objective function using nonlinear regression. UCODE minimizes objective functions using a modified Gauss-Newton method. UCODE provides options to include prior or direct information on estimated parameters. UCODE can be used in conjunction with any application models that use ASCII or text-only input, produce ASCII or text output, and can be executed in batch mode. The performance of UCODE has been tested in a variety of applications. Obtaining useful UCODE results depends on 1) defining a tractable inverse problem using simplified appropriate mathematical representation of the given system and 2) wise use of the statistics generated using calculated sensitivities and the match between observed and simulated values and appropriate graphical analyses of the statistical results. Guidelines and considerations in the application of UCODE for the construction and calibration to site-specific conditions are found in Hill (1998). The reader should also refer to several UCODE-related documents and guidelines that can be found at http://water.usgs.gov/software/ground_water.html/.

Each UCODE application requires the following key input files:

- A universal ("< >.**uni**") file which contains model parameters and observation data that will be examined in the application of UCODE.
- A prepare ("< >.**pre**") file, which includes a list of parameters, their start values, minimum and maximum ranges, perturbation size, and other associated data.
- A template (< >.**tpl**) file, which defines the specific format for UCODE to substitute updated parameters in the model input files.
- A substitute (< >.**sub**) file, which is a UCODE-generated revised input file with the perturbed parameters values.
- An extract (< >.**ext**) file, which includes the necessary information and details required by UCODE to extract simulation results for each of the observed data.

Following is a general overview of specialized CFEST utilities for generating the above files required for UCODE applications.

A.1.2 Overview of Specialized CFEST Utilities for UCODE Application

CFEST utilities have been developed to provide a set of streamlined procedures and tools to simplify the use of UCODE. The following two programs have been specifically developed for integrated execution of CFEST and UCODE.

CF_UCODE	Generates < >.uni, < >.pre, < >.sub, < >.tpl, < >.hed, and other UCODE interface files.
LP3UCODE	Reads UCODE-generated < >.sub file, updates parameters, performs simulations, and extracts result files (< >.cfo, < >.hed, < >.riv and others).

Before executing the above tools, the following three CFEST programs must be executed.

LPROG1	Reads material properties, nodal coordinates, boundary conditions, stream/river properties, and elemental details from LP1 files and generates binary files.
LBAND	Calculates structure of linear system of equations for solution of head and concentration.
LPROG3I	Reads options for steady-state or transient simulation of head and concentration, surface recharge, elemental source/sink terms, and nodal injection/extraction data.

CFUCODE reads UCODE parameters from a control file (cf_ucode.ctl) which has much of the same input data as a normal UCODE < >.uni file along with a specialized file to simplify parameterization of the 3-dimensional model for use with UCODE. CFUCODE also generates the UCODE interface files (< >.uni, < >.pre, < >.sub, < >.tpl, < >.hed, and others). The control file (cf_ucode.ctl) includes the names of the files containing the observation data and parameter codes that simplify the parameterization.

The CFUCODE utility also generates a "< >.bat file" for executing UCODE for the problem defined by the cf_ucode.ctl file. The script that launches the computational code, LP3UCODE, which is a "< >.bat" file on PC's or "< >.csh" on unix platforms is specified in the cf_ucode.ctl file and becomes part of the generated "< >.uni" file. This launch script is needed for UCODE to run LP3UCODE. It has to be prepared separately by the user as discussed in the Section 4.2.1.1 on parallel operation in the main text. This script must be generated by the user in order to properly interface with MasterTasker. A more detailed discussion of the command line arguments passed to the launch script by MasterTasker is found in Section B.4.2.3 of Appendix B.

CFUCODE and LP3UCODE program details are given in Sections A.2 and A.3, respectively.

A.2 CFUCODE Application

Key inputs to the CFUCODE are the control (cf_ucode.ctl), observation data, and model parameter files. CFUCODE generates "< >.uni," "< >.pre," "< >.tpl," " < >.ext," "< >.hed," and other UCODE interface files. Model parameter inputs used in UCODE are entered as model variables referred to as Paracodes. This chapter includes a discussion on model variables referred to as Paracodes and input and output files used in CFUCODE.

A.2.1 Model Parameter Variables

Sets of model parameter variables referred to as Paracodes have been designed for use with the CFEST-UCODE tools to simplify the parameterization process for inverse analysis (e.g. creating parameter zones). The Paracodes developed are grouped into the following broad categories: 1) hydraulic properties (i.e., hydraulic conductivities in the x, y, and z direction, specific yield, specific storage, porosity), 2) surface recharge, 3) stream and river properties, 4) boundary flux, and 5) well flow rates. These codes do not yet include transport-related parameters (e.g., longitudinal and lateral dispersivities, molecular dispersion, retardation, and other transport related parameters). Transport parameters will be added in the next update. There are two basic ways provided for parameterization of zones

A.2.1.1 Hydraulic Properties Paracodes

In CFEST, hydraulic properties can be defined by using material property numbers that have been assigned to the major hydrostratigraphic units or to create heterogeneity within a single aquifer unit by having different material property zones. Additionally the hydraulic parameters for hydraulic conductivity, storage coefficient, and specific yield can be assigned individually to each and every element. Hydraulic property paracodes include hydraulic conductivities in the x-, y-, and z-directions (Kx, Ky, Kz), specific yield (Sy), specific storage (Ss). Paracodes can be specified by material property number or by specifying parameter zones by identifying a list of elements within the zone or through a string of boundary points that forms a polygon that encloses the elements within the zone. Parameter can be specified as explicit parameter values, as a multiplying factor that is applied to the input distribution of parameter values specified in the normal CFEST input, or as an anisotropy factor that is applied to after the Kx and Ky values have been determined. To simplify the specification of the hydraulic conductivity, which can be isotropic, the user can specify that Kxyz change as a unit as specified in the normal CFEST input file and only one value is treated as a parameter (i.e., Kx) and the Ky and Kz keep the same Kx/Ky and Kx/Kz ratio specified in the original CFEST inputs. Similarly there are provisions for keeping Kxy as a single parameter. Anisotropy factor paracodes only apply to the hydraulic conductivity parameters. An anisotropy factor paracode allows the user to specify the anisotropy ratio by material number or as a zone through specification of a list of elements or through the use of polygonal zones. Use of explicit paracodes is straightforward as the value specified is directly used as CFEST input. However, when using factors the factor is applied to the parameters as specified in the original CFEST input. When using the material number factor, there is little difference from the explicit approach except in the value of the parameter specified (e.g., it would typically start at 1.0). When using zones specified as a list of elements or zone as specified by polygonal areas that enclose a group of elements the operation is different. The values within a zone may not all have the same parameter values. Using the factor approach to parameterization means that ratios between parameter values in the zone are maintained since all the original values within the zone as specified in the initial CFEST input are multiplied by the factor specified for this parameter zone. Contrast this to the explicit parameter approach where all the elements within the zone would have exactly the same parameter value.

The following is a list of paracodes associated with hydraulic properties. The last two characters (Columns 5 and 6) of each paracode corresponds to a specific material number (MN), user-defined list of elements (LS), or polygonal zone (ZN) (i.e., the elements within a polygon defined by a set of x and y coordinates).

Hydraulic Property Paracodes

Using Explicit Parameter Values

Mat. # Based	KX00MN	KY00MN	KZ00MN	KXY0MN	KXYZMN	SPECMN	SYLDMN
Elem. List Zone	KX00LS	KY00LS	KZ00LS	KXY0LS	KXYZLS	SPECLS	SYLDLS
Polygon Zone	KX00ZN	KY00ZN	KZ00ZN	KXY0ZN	KXYZZN	SPECZN	SYLDZN

Using Multiplier Factor

Mat. # Based	FCKXMN	FCKYMN	FCKZMN	FCKHMN	FCK3MN	FCSSMN	FCSYMN
Elem. List Zone	FCKXLS	FCKYLS	FCKZLS	FCKHLS	FCK3LS	FCSSLS	FCSYLS
Polygon Zone	FCKXZN	FCKYZN	FCKZZN	FCKHZN	FCK3ZN	FCSSZN	FCSYZN

Anisotropy Factors

Mat. # Based	FCXYMN	FCXZMN	FCYZMN
Elem. List Zone	FCXYLS	FCXZLS	FCYZLS
Polygon Zone	FCXYZN	FCXZZN	FCYZZN

Included below are some illustrative samples for uses of some of the hydraulic property paracodes taken from an example, < >.pra file. The TIED01 and "tied" concept found in 3-5 of the following example is discussed in a following section (A.2.2.5)

#PARACODE	ID	LY	START	VAL-MIN	VAL-MAX	PERTURB	LOG	IESTIM	PLn	TIED	FILE	Comment
'FCK3MN'	0	1	0.9	0.01	100	0.01	1	1	0	'NO'	'NONE'	Hanford
'FCK3MN'	0	5	2.1	0.01	100	0.01	1	1	0	'NO'	'NONE'	Ringold 5
'FCK3MN'	0	6	1	0.01	100	0.01	1	1	0	'TIED01'	'NONE'	Ringold 6
'FCK3MN'	0	7	1	0.01	100	0.01	1	1	0	'TIED01'	'NONE'	Ringold 7
'FCK3MN'	0	8	1	0.01	100	0.01	1	1	0	'TIED01'	'NONE'	Ringold 8
'FCK3MN'	0	9	0.52	0.01	100	0.01	1	1	0	'NO'	'NONE'	Ringold 9
'FCSYMN'	0	1	0.25	0.1	2	0.01	1	1	0	'NO'	'NONE'	Hanford
'FCSYMN'	0	5	2.06	0.1	2	0.01	1	1	0	'NO'	'NONE'	Ringold 5
'SPECMN'	0	1	1.00E-06	5.00E-07	5.00E-05	0.01	1	0	0	'NO'	'NONE'	Hanford
'SPECMN'	0	5	1.00E-06	5.00E-07	5.00E-05	0.01	1	0	0	'NO'	'NONE'	Ringold 5
'SPECMN'	0	7	1.00E-06	5.00E-07	5.00E-05	0.01	1	0	0	'NO'	'NONE'	Ringold 7
'SPECMN'	0	9	1.00E-06	5.00E-07	5.00E-05	0.01	1	0	0	'NO'	'NONE'	Ringold 9
'FCK3LS'	0	10	1	0.1	100	0.01	1	1	0	'NO'	'no-faults.txt'	EM Basalt
'FCK3LS'	0	10	1.76E+03	1.70E+02	1.70E+04	0.01	1	1	0	'NO'	'rmfault.txt'	RM Fault
'FCK3LS'	0	10	1.76E+03	1.70E+02	1.70E+04	0.01	1	1	0	'NO'	'gmfault.txt'	GM Fault
'FCK3LS'	0	10	5.30E+03	5.30E+02	5.30E+04	0.01	1	1	0	'NO'	'b-pond.txt'	B Pond
'FCXZMN'	0	1	1	0.0001	10	0.01	1	0	0	'NO'	'NONE'	Anisotropy

A.2.1.2 Surface Recharge Paracodes

Surface recharge can be defined in different ways using the variable, NSURFQ, with the following values:

NSURFQ	Recharge Definition
1	Constant in time and space
2	Variable in space and constant in time
3	Variable in space and time

The following is a list of paracodes associated with surface recharge that are used in CFUCODE:

Surface Recharge Paracodes

Using Explicit Recharge Rate or Multiplier Factor

Explicit Rate	RCQSUR	RCQSLS	RCQSZN
Multiplier Factor	FCQSUR	FCQSLS	FCQSZN

Note: RCQSUR can be used only with constant recharge rates (i.e. NSURFQ=1). RCQSLS and RCQSZN are used when recharge is variable in space and time (NSURFQ=2). FCQSLS and FCQSZN can only be used with recharge that varies in space and/or in time (NSURFQ=2 or 3).

Included below is an example of one of the surface recharge paracodes taken from an example, < >.pra file.

#PARACODE	ID	LY	START	VAL-MIN	VAL-MAX	PERTURB	LOG	IESTIM	PLn	TIED	FILE	Comment
'FCQSUR'	0	0	1.22	0.5	5	0.01	1	1	0	'NO'	'NONE'	Recharge

A.2.1.3 Head-Dependent Flux Paracodes

Head-dependent boundary conditions are typically associated with rivers/streams and can be defined by using an assumed stream-bed vertical hydraulic conductivity value ($SBKv_{streambed}$) or a streambed conductance variable, SB, which is defined as follows:

$SB = SB_{Length} * SB_{Width} * SBKv_{streambed} / SB_{Thick}$

Where: SB_{Length} is the stream bed length
SB_{Width} is the stream bed width
SB_{Thick} is the stream bed thickness

Paracodes used to manipulate these model variables include the following:

Stream Paracodes

Stream Coefficient

Explicit Value	STRALL	STRNOD	(STRALL=all values STRNOD=Node List)
Factor	FCSTRM	FCSTND	(FCSTRM=all values FCSTND=Node List)

Stream Bed Hydraulic

Explicit Values	STBKAL	STBKND	(STBKAL=all values STBKND=Node List)
Factors	FCSTKA	FCSTKN	(FCSTKA=all values FCSTKN=Node List)

Included below is an example of one of the streambed coefficient paracodes taken from a Hanford parameter file being used to represent the basalt erosional window (i.e., < >.pra file).

#PARACODE	ID	LY	START	VAL-MIN	VAL-MAX	PERTURB	LOG	IESTIM	PLn	TIED	FILE	Comment
'FCSTKA'	0	0	1	0.1	100	0.01	1	1	0	'NO'	'NONE'	Erosional Window

A.2.1.4 Boundary Flux Paracodes

Nodal or elemental boundary fluxes can be specified at a specific node or element. CFUCODE options are provided to specify a total flux along a line source using a specified unit length flux or a total flux specified through a vertical face of a given set of nodes. With these options, the CFUCODE internally distributes the specified flux in the following ways:

- Unit-length Line flux – the total line flux is distributed according to length of line represented by a given set of nodes
- Vertical-Face Area Flux- total area flux is distributed according to each nodal transmissivity (i.e hydraulic conductivity multiplied by the hydrogeologic unit thickness) represented by a given set of nodes.

These options can be used to reduce several nodal flux estimates to a single total flux- model parameter. Paracodes for these types of boundary fluxes (i.e., along a line or through a vertical face) are as follows:

Boundary Flux Paracodes

Total Line and Vertical Flux

```
Explicit Values  FLXQLN  FLXQAR    ("LN"=Line "AR"=Vertical Face)
Factor           FACQLN  FACQAR
```

Included below are two illustrative samples for uses of line-flux and area-flux boundary paracodes.

#PARACODE	ID	LY	START	VAL-MIN	VAL-MAX	PERTURB	LOG	IESTIM	PLn	TIED	FILE	Comment
'FACQAR'	1	0	2	0.8	5	0.01	1	1	0	'NO'	'coldcreek.afx'	Cold Creek
'FACQAR'	2	0	1	0.8	5	0.01	1	1	0	'NO'	'drycreek.afx'	Dry Creek
'FACQLN'	2	0	1	0.8	5	0.01	1	1	0	'NO'	'yakima.lfx'	Yakima Ridge Fault

The above lines are extracted from one of the early the hanford.pra files. Paracodes ending in “LN” and “AR” is used for line and area flux boundary paracodes. Files at the end of each line ending in “lfx” and “afx” are files containing more specific information about the line-flux and area-flux boundaries.

The file “yakima.lfx” is an example of a specified line-flux boundary file and is listed below.

21	1.00E-05	,total nodes involved with yakima fault, total flux									
-661	-662	-664	-697	-698	-700	-701	-747	-749	-751	-753	-755
-796	-841	-843	-891	-937	-939	-986	-1031	-1033			

In this example data input file, the line-flux boundary condition is used to represent vertical leakage from the underlying basalt confined aquifer at the approximate location of a thrust fault in the basalt. The total flux at this location is unknown and the inversing process is being used to estimate its approximate value. A nominal starting flux value is entered as an initial approximate and UCODE is being used to refine the estimate in the inversing process.

The first line of the file defines the total number of nodes and the total flux through the line-source boundary followed by comments. The second and other lines in the file specify the list of nodes that define the line-flux boundary location. In this case, the negative sign in front of each node indicates that the lowermost node below the specified node location will be used as the source node location in the calculation of flux distribution along the line. All nodes are entered sequentially from the start to the end of the line. Nodes are read in free format and can be comma separated, space separated, or entered one node on each line.

The file "coldcreek.afx" is an example of an area-flux boundary file and is listed below.

3	2881.32	,coldcreek flux area nodes, flux (m3/day)	
1078	1080	1126	surface nodes

In this example data input file, the area-flux boundary condition is being used to represent the influx of groundwater (in m^3/day) where Cold Creek Valley enters into the modeled region. The total flux at this location has been previously estimated and UCODE is being used to reevaluate this estimate in the inversing process.

The first line of the file defines the total number of nodes and the total flux entering into the model at this location followed by comments. The second line in the file specifies the list of nodes that define the surface nodes at this boundary location. This information is used to identify all nodes vertically below a given surface node in the calculation of flux distribution across the vertical plane at this location within the modeled region.

A.2.1.5 Well-Pumping/Injection Paracodes

The following paracodes are provided for changing flow rates assigned to a single or multiple nodes containing a flux into or out of the modeled region. These nodal sources and sinks are used to represent pumping or injection wells or any other specified flux to nodes. This option can be only used with time constant flow rates. Paracodes used for this option are as follows:

Well Flowrate Paracodes

Explicit Values

Specific node Q	QWELND (ND=Specific Node)
Constant Q all Nodes	QWELAL (AL=All Well Nodes)

Factor

Specific Node	FCWLND (ND=Specific Node)
Constant Factor all wells	FCWLAL (AL=All Well Nodes)

Examples of the single source/sink nodes where the explicit flow rate or a multiplier factor are used is given the following example < >.pra file where the specific node is identified these types of paracodes.

#PARACODE	ID	LY	START	VAL-MIN	VAL-MAX	PERTURB	LOG	IESTIM	PLn	TIED	FILE	Comment
'QWELND'	284	1	1.5	0.0001	10	0.01	1	1	0	'NO'	'NONE'	pumping well example
'FCWLND'	285	1	4.36	0.0001	10	0.01	1	1	0	'NO'	'NONE'	well factor example

A.2.1.6 "TIED" Paracodes

This is an option to "tie" together a group of paracodes so that a single input is applied to a series of similar parameters. The earlier example in subsection A.2.1.1 on hydraulic parameter codes uses the tied column (i.e., column 11) to specify the name of the TIED01 parameter to three different hydraulic conductivity specifications. Use of this TIED01 specification allowed the hydraulic conductivity parameters for Ringold 6, 7, and 8 to be tied together so that only one parameter was used to specify the multiplication factor for all three hydrostratigraphic units. The following example allows a single multiplier specification to be applied to the spatially variable distribution of Rattlesnake Hills flux so that only one parameter needs to be determined and the relationship between the various TIED02 components is maintained.

#PARACODE	ID	LY	START	VAL-MIN	VAL-MAX	PERTURB	LOG	IESTIM	PLn	TIED	FILE	Comment
'FCWLND'	285	1	4.36	0.0001	10	0.01	1	1	0	'TIED02'	'NONE'	Rattlesnake Springs
'FCWLND'	284	1	4.36	0.0001	10	0.01	1	1	0	'TIED02'	'NONE'	Rattlesnake Springs
'FCWLND'	287	1	4.36	0.0001	10	0.01	1	1	0	'TIED02'	'NONE'	Rattlesnake Springs
'FCWLND'	310	1	4.36	0.0001	10	0.01	1	1	0	'TIED02'	'NONE'	Rattlesnake Springs
'FCWLND'	312	1	4.36	0.0001	10	0.01	1	1	0	'TIED02'	'NONE'	Rattlesnake Springs
'FCWLND'	333	1	4.36	0.0001	10	0.01	1	1	0	'TIED02'	'NONE'	Rattlesnake Springs
'FCWLND'	335	1	4.36	0.0001	10	0.01	1	1	0	'TIED02'	'NONE'	Rattlesnake Springs
'FCWLND'	357	1	4.36	0.0001	10	0.01	1	1	0	'TIED02'	'NONE'	Rattlesnake Springs
'FCWLND'	359	1	4.36	0.0001	10	0.01	1	1	0	'TIED02'	'NONE'	Rattlesnake Springs
'FCWLND'	382	1	4.36	0.0001	10	0.01	1	1	0	'TIED02'	'NONE'	Rattlesnake Springs
'FCWLND'	384	1	4.36	0.0001	10	0.01	1	1	0	'TIED02'	'NONE'	Rattlesnake Springs
'FCWLND'	408	1	4.36	0.0001	10	0.01	1	1	0	'TIED02'	'NONE'	Rattlesnake Springs
'FCWLND'	410	1	4.36	0.0001	10	0.01	1	1	0	'TIED02'	'NONE'	Rattlesnake Springs
'FCWLND'	436	1	4.36	0.0001	10	0.01	1	1	0	'TIED02'	'NONE'	Rattlesnake Springs
'FCWLND'	435	1	4.36	0.0001	10	0.01	1	1	0	'TIED02'	'NONE'	Rattlesnake Springs
'FCWLND'	438	1	4.36	0.0001	10	0.01	1	1	0	'TIED02'	'NONE'	Rattlesnake Springs
'FCWLND'	465	1	4.36	0.0001	10	0.01	1	1	0	'TIED02'	'NONE'	Rattlesnake Springs

In the above example springs discharge into the aquifer system at the Rattlesnake Hills boundary. The uncertainty in the flux along this boundary, estimated in an earlier modeling effort, is being revaluated. In this example, node locations for each specified flux are identified by the number after the paracode "FCWLND" in column 2. The specific flux at the node is provided in the normal input to the CFEST module LPROG3I (i.e., <..>.l3i file). In this specific example the "TIED02" paracode is used to link the 17 nodal flux nodes (285, 284, 287, 310, 312, 333, 335, 357, 359, 382, 384, 408, 410, 436, 435, 438, 465) specified in the 17 rows above into a single model parameter multiplier factor (TIED02).

A.2.2 Input Files for CFUCODE

Brief guidelines for CFUCODE input file preparations.

A.2.2.1 Control File

Table A.1and A.2 include a copy of the detailed and short version of the control file, cf_ucode.ctl. New users may use a detailed copy of cf_code .ctl since it includes a set of comment lines explaining the content of each cf_ucode.ctl file input line.

The last three lines of the control file, cf_ucode.ctl, before the PARALLEL specification, which was discussed fully in Section 4.2.1.1, are used to specify the following data file names

- File containing observation well data
- File containing stream-flow observation data
- File containing paracodes that will be examined by UCODE.

Table A.1. Copy of Detailed cf_ucode.ctl File

```
#=============================================================================
#       Print control options nprint_ctl and idetail
#       nprint_ctl (0=default,1=echo ctl lines),
#       idetail (0=default,1=include comments)
#=============================================================================
0,0,                #nprint_ctl, idetail (0=default,1=include comments)
#=============================================================================
#       Filename “fn”. The program generates <  >.uni, <  >.ext, <  >.pre
#=============================================================================
'basecase,'         #filename “fn”
#=============================================================================
#       Phase
#=============================================================================
#       1= Para substitute and forward modeling-use starting parameters
#      11= Above + sum-of-squared, weighted residuals objective functions,
#          produces data for objective-function contour graph
#       2= Sensitivities at starting parameter values
#      22= Above+para variances, covariance and correlation at init.values
#       3= Perform regression
# To run phase 33,44&45 you must 1st Execute UCODE with Phase=3 and Graph=1
# in the same directory , When phase=33, 44, and 45 are run only the phase
# line is read
#      33= Calculate the modified Beale's measures of model linearity
#      44= Estimate predictions and their linear confidence and prediction
#          intervals
#      45= Use after Phase 44 for differences and their linear confidence
#          and prediction intervals.
#=============================================================================
1,                  #Phase (1,11,2,22,3,33,44,45)
#=============================================================================
#     Sensitivity and regression control
#     differencing {1=forward(recommended) 2=Central}.
#     Central option for sensitivities is used during parameter-estimation
#     iterations
#=============================================================================
1,                  #Differencing option
#=============================================================================
#     Tolerance, tol, (0.01 recommended) {convergence criteria-fractional
#     amount between regression iterations}
#=============================================================================
0.01,               #tol
#=============================================================================
#     Sum Of Squared weighted Residuals, sosr, changes parameter for the
#     convergence criteria. (0.01 or even 0.1 is recommended at early stages).
#     Convergence is considered complete when the sum of sum of squared
#     weighted residuals is less than the specified fractional amount “sosr”
#     over three-regression iterations -
#=============================================================================
0.0,                #sosr
#=============================================================================
#     nopt {0=No quassi-Newton update (recommended)
```

```
#           1= for highly non-linear problems}
#     This option is used used when sosr<0.01 over 3 regression iter.
#================================================================================
0,                  #nopt(0=recommended 1= use for highly non-linear)
#================================================================================
#     Maximum iterations for the regression, max-iter.
#     ~2-3 times number of parameters. Higher max-iter rarely helps for
#     non-conversion cases.
#================================================================================
8,                  #max-iterations
#================================================================================
#     max-change (fractional change of a parameter values in one regression
#     iteration
#================================================================================
2.0,                #max_change
#================================================================================
#     Command to execute inversion algorithm.
#     Edit in the correct path for the mrdrive executable supplied with UCODE
#================================================================================
C:\my-ucode\ucode1.09\bin\mrdrive
#================================================================================
#     Commands to run the application model (i.e., the forward model)
#     first line is the number of commands and subsequent are the path/names
#================================================================================
1,               #number of application models
'lp3ucode.bat'   #batch file for running application model
#================================================================================
#     Printing and plotting options
#       - scale-sensitivities applied to printed sensitivities- typically
#         1,2, or 3 are used. (0=no scaling, 1=dimensionless scaled,
#         2= 1% scaled 3= both dimensionless and 1% scaled are printed)
#       - Since graph data files are not large, 1 is recommended.
#================================================================================
3,              #Scale-sensitivities
0,              #print-intermediate (0=recommended,1=yes)
0,              #graph (Phase 3) 0=none 1=print post processing files.
0,              #number-residual-sets produced and written to file
#================================================================================
#     Note:<  >._rp, <  >._rd, <  >._rg for evaluation of non-randomness
#     of residuals.
#================================================================================
#     Base time data for converting observation time to model times
#     Option for handling observation times, iopt_time, and
#     Time of model time zero, ModelTime0, in time units specified by
#     iop_time.
#     iopt_time [0=(yyyy,mm,dd), 1=(hh:min:sec)],
#     ModelTime0 in iopt_time units (for steady state use 0000/00/00)
#================================================================================
0,'0000/00/00'
#================================================================================
#     Observation and parameter files
#================================================================================
'hcf_ex3.fld '     File_headObs
'scf_ex3.fld '     File_strmObs
'pcf_ex3.pra '     File_parameter
#================================================================================
#     Parallel options (no annotation)
#================================================================================
PARALLEL
C:\wrdapp\ucode3.0\master-tasker\shared\     #path to parallel job run directory
C:\wrdapp\ucode3.0\test-parallel2\test\      #path to variable files directory
C:\wrdapp\ucode3.0\test-parallel2\static\    #path to unchanging files directory
5                              #rpttime 1 sec
s                              #rptunit
```

```
1                               #default space 1M
m                               #space unit
600                             #default speed
m                               #speed unit MHz
5                               #default time 2 min
m                               #time unit
m                               #default launch for (model.bat)
1500                            #time out (secs)
leave                           #leave files on slaves in case of abort
#================================================================================
#     End of parallel processing specifications and start of optional feature
#     specifications
#================================================================================
RESTART_SAVE
ALTERNATE_PRINT
HEADER_PRINT
GROUP_STATS
END
```

Table A.2. Copy of Condensed cf_ucode.ctl File

```
#================================================================================
0,0,                #nprint_ctl, idetail (0=default,1=include comments)
#================================================================================
'basecase,'         #filename “fn”
#================================================================================
1,                  #Phase (1,11,2,22,3,33,44,45)
#================================================================================
1,                  #Differencing option
#================================================================================
0.01,               #tol
#================================================================================
0.0,                #sosr
#================================================================================
0,                  #nopt(0=recommended 1= use for highly non-linear)
#================================================================================
8,                  #max-iterations
#================================================================================
2.0,                #max_change
#================================================================================
C:\my-ucode\ucode1.09\bin\mrdrive
#================================================================================
1,               #number of application models
'lp3ucode.bat'   #batch file for running application model
#================================================================================
3,              #Scale-sensitivities
0,              #print-intermediate (0=recommended,1=yes)
0,              #graph (Phase 3) 0=none 1=print post processing files.
0,              #number-residual-sets produced and written to file
#================================================================================
0,'0000/00/00'
#================================================================================
'hcf_ex3.fld '     File_headObs
'scf_ex3.fld '     File_strmObs
'pcf_ex3.pra '     File_parameter
#================================================================================
PARALLEL
C:\wrdapp\ucode3.0\master-tasker\shared\     #path to parallel job run directory
C:\wrdapp\ucode3.0\test-parallel2\test\      #path to variable files directory
C:\wrdapp\ucode3.0\test-parallel2\static\    #path to unchanging files directory
5                               #rpttime 1 sec
s                               #rptunit
1                               #default space 1M
m                               #space unit
600                             #default speed
m                               #speed unit MHz
5                               #default time 2 min
m                               #time unit
m                               #default launch for (model.bat)
1500                            #time out (secs)
leave                           #leave files on slaves in case of abort
#================================================================================
RESTART_SAVE
ALTERNATE_PRINT
HEADER_PRINT
GROUP_STATS
END
```

Table A.3. Partial List of Hanford Head Observation Data Illustrating Data Format and Contents

```
0,0, Nprint(0=normal,1=echo each data line in dbg file),idetail(0=normal,1=echo comment lines also)
# User defines date format ** following format is for reading month, day, year, and time "11/17/1992 00:00:00"
(2(i2,1x),i4,3(1x,2i))
1, IOPT_STATPARA (0=use default statistical parameters, 1=data includes statistical parameters for each observation
1, well x-y conversion factor
#=============================================================================================================
# current version reads data up to isymbol, Other information will be used to internal element and time plane
# estimates with external estimates
#=============================================================================================================
#wellname,date,time,head,xcoord,ycoord,screen_top,screen_bot,stat,istflag,isymbol,timeplane,unitno,element,timewt,spacewt,head_err
'199-B2-12','11/17/1992 00:00:00',120.26,565368.62500,145363.93800,83.53,80.49,1,0,1,8,1,1367,0.848,1.00,0.999
'199-B2-12','12/21/1992 10:41:00',120.93,565368.62500,145363.93800,83.53,80.49,1,0,1,8,1,1367,0.152,1.00,0.999
'199-B2-12','01/18/1993 09:21:00',120.95,565368.62500,145363.93800,83.53,80.49,1,0,1,9,1,1367,0.098,1.00,0.999
...
'199-B2-12','10/27/1998 11:34:00',120.01,565368.62500,145363.93800,83.53,80.49,1,0,1,20,1,1367,0.151,1.00,0.999
'199-B2-12','11/19/1998 10:25:00',120.44,565368.62500,145363.93800,83.53,80.49,1,0,1,20,1,1367,0.299,1.00,0.999
'699-S32-E8','02/12/1990 13:47:00',110.25,592263.81200,113922.35900,93.35,90.29,1,0,1,3,1,138,0.505,1.00,1.000
'699-S32-E8','05/23/1990 09:13:00',110.32,592263.81200,113922.35900,93.35,90.29,1,0,1,3,1,138,0.369,1.00,1.000
...
'699-S32-E8','06/29/1998 13:53:00',111.61,592263.81200,113922.35900,93.35,90.29,1,0,1,19,1,138,1.000,1.00,1.000
'699-S32-E8','03/30/1999 13:47:00',111.59,592263.81200,113922.35900,93.35,90.29,1,0,1,21,1,138,1.000,1.00,1.000
END
```

Guidelines and key features for files listed in last these lines are as follows.

A.2.2.2 Observation Well Data File

Table A.3 includes a few lines from the Hanford head observation data file. Comments in Table A.3 identify the content and formats for preparing data files. A free format is used for the file. The well name and the date + time data (alphanumeric data) should have a single quote at the start and end of each well name and date data.

CFUCODE converts well names to a sequential head data identifier by using "H" in first left hand field and a five-digit sequential number of a given well as illustrated below.

The following lines are extracted from the Hanford.uni file to illustrate the translation of the well name in Table A.3.

```
###observations
H00001    120.26  1.000000    0    1
H00002    120.93  1.000000    0    1
H00003    120.95  1.000000    0    1
H00004    120.22  1.000000    0    1
H00005    121.20  1.000000    0    1
H00006    119.99  1.000000    0    1
H00007    119.87  1.000000    0    1
H00008    119.84  1.000000    0    1
H00009    119.94  1.000000    0    1
```

CFUCODE also generates < >.hed file. The following illustrates only part of the hanford.hed file.

```
#WellID SimResult ObsValue TimeFrom   Elem#  Inter  Local_X Local_Y  Local_Z DateObs    WellName
#                          Start           Elm#
H00001   118.5757 120.2600  1050.500   61367  6356  0.6004  0.8831  0.8267 11/17/1992  199-B2-12
H00002   118.5744 120.9300  1084.500   61367  6356  0.6004  0.8831  0.8267 12/21/1992  199-B2-12
H00003   118.7288 120.9500  1112.750   61367  6356  0.6004  0.8831  0.8267 01/18/1993  199-B2-12
H00004   118.7279 120.2200  1113.750   61367  6356  0.6004  0.8831  0.8267 01/19/1993  199-B2-12
H00005   118.7030 121.2000  1141.750   61367  6356  0.6004  0.8831  0.8267 02/16/1993  199-B2-12
H00006   118.6754 119.9900  1172.750   61367  6356  0.6004  0.8831  0.8267 03/19/1993  199-B2-12
H00007   118.6504 119.8700  1200.750   61367  6356  0.6004  0.8831  0.8267 04/16/1993  199-B2-12
H00008   118.6495 119.8400  1201.750   61367  6356  0.6004  0.8831  0.8267 04/17/1993  199-B2-12
H00009   118.6353 119.9400  1217.750   61367  6356  0.6004  0.8831  0.8267 05/03/1993  199-B2-12
H00010   118.6255 121.3000  1228.750   61367  6356  0.6004  0.8831  0.8267 05/14/1993  199-B2-12
```

The following is brief explanation of some of key features of the < >.hed file.

- A given observation well generally has multiple observation data taken at different times. The UCODE needs a unique ID for each observation. CFUCODE uses a sequential number for the given observation as identification for that observation. In the < >.hed file, the user-defined observation well name is included in the last column.

- Simulated and observed data are displayed in two adjacent columns for easy comparison of simulated results with the observed data.

- The internal element number, the local coordinate, and the time from model start time provide an easy means to interpolate simulated data for each observed value.

For transient data, CFUCODE includes only those observations that are within the simulation time range. Users should edit the standard CFEST control file, cfest.ctl, for starting and ending time steps (Line 9 – ITT [initial or restart time step], ITSTOP in cfest.ctl). If the cfest.ctl is changed after executing the CFUCODE, the user should re-execute the CFUCODE.

A.2.2.3 Stream/River Observation Data File Structure

Following is a copy of example transient stream-flow data file.

```
0,0, nprint
#StrmGage HHMMSSSSSSSS StrmFlow    Stat  ISFlag ISym NtNodStrm (Nod_strm(j),j=1,NtNodStrm)
(i2,i2,i8)
'flow','000000000000', -4.050001    0.10      2   2   0  #river gain must have prefix "-"
'flow','000000871620', -4.049938    0.10      2   2   0  #river gain must have prefix "-"
'flow','000024439065', -2.949642    0.10      2   2   0  #river gain must have prefix "-"
END
```

The stream-flow data file is on a similar line as the head-observation data file. The user must define the format for date. Observation and date/time data are entered with single quotes. River gain is entered as negative. The CFUCODE converts stream observation names to a sequential number with the leading character “S” in the first column. Following is a copy of stream flow lines in < >.uni and < >.riv file of the above data.

A copy of the stream-flow observation is in the < >.uni file. Note, the “flow” observation name is changed to “S” names.

```
S0001      -4.05  0.100000    2    2
S0002      -4.05  0.100000    2    2
S0003      -2.95  0.100000    2    2
END
```

A copy of the stream-flow observation is in the < >.riv file of flow data.

```
#StreamId   SimValue       ObsValue     TimeStep      DateObs      MonitoringName TotNodes  Nodes
(A,T10,2G15.7,f15.3,1x,a,1x,A,I5,200I8)
S0001      -4.057534      -4.050001     0.000         000000000000   flow          0
S0002      -4.057459      -4.049938     871620.000    000000871620   flow          0
S0003      -2.899951      -2.949642     24439065.000  000024439065   flow          0
```

A.2.2.4 Parameter Data File Structure

The last line in the < >.uni file before the parallel specification includes the name of the parameter (< >.pra) file. This file includes the names of all parameters that must be optimized. Following is a copy of a hanford.pra file.

hanford.pra file listing

```
1,1,  nprint_ctl (0=default,1=echo ctl lines),idetail(0=default,1=IncComments)
'F no',
'%15.7e',format in C++ for UCODE to write <  >.sub file
#=============================================================================================
#PCODE    ID LY  START    VAL-MIN VAL-MAX PERTURB  LOG IESTIM PLn TIED      FILE
'FCK3MN'   0  1  0.900000 .010000 100.000    0.01   1  1      0   'NO'     'NONE'
'FCK3MN'   0  5  2.100000 .010000 100.000    0.01   1  1      0   'NO'     'NONE'
'FCK3MN'   0  6  1.000000 .010000 100.000    0.01   1  1      0   'TIED01' 'NONE'
'FCK3MN'   0  7  1.000000 .010000 100.000    0.01   1  1      0   'TIED01' 'NONE'
'FCK3MN'   0  8  1.000000 .010000 100.000    0.01   1  1      0   'TIED01' 'NONE'
'FCK3MN'   0  9  0.520000 .010000 100.000    0.01   1  1      0   'NO'     'NONE'
'FCSYMN'   0  1  0.250000 .100000 2.00000    0.01   1  1      0   'NO'     'NONE'
'FCSYMN'   0  5  2.060000 .100000 2.00000    0.01   1  1      0   'NO'     'NONE'
'FACQAR'   1  0  2.000000 0.80000 5.00000    0.01   1  1      0   'NO'     'coldcreek.afx'
'FACQAR'   2  0  1.000000 0.80000 5.00000    0.01   1  1      0   'NO'     'drycreek.afx'
'FCQSUR'   0  0  1.220000 0.50000 5.00000    0.01   1  1      0   'NO'     'NONE'
'FCSTKA'   0  0  1.000000 0.10000 100.000    0.01   1  1      0   'NO'     'NONE'
'SPECMN'   0  1  1.00e-06 0.5e-06 5.0e-05    0.01   1  0      0   'NO'     'NONE'
'SPECMN'   0  5  1.00e-06 0.5e-06 5.0e-05    0.01   1  0      0   'NO'     'NONE'
'SPECMN'   0  7  1.00e-06 0.5e-06 5.0e-05    0.01   1  0      0   'NO'     'NONE'
'SPECMN'   0  9  1.00e-06 0.5e-06 5.0e-05    0.01   1  0      0   'NO'     'NONE'
'FCXZMN'   0  1  1.000000 0.00010 10.0000    0.01   1  0      0   'NO'     'NONE'
'FCWLND' 285  1  4.360000 0.00010 10.0000    0.01   1  1      0   'TIED02' 'NONE'
'FCWLND' 284  1  4.360000 0.00010 10.0000    0.01   1  1      0   'TIED02' 'NONE'
'FCWLND' 287  1  4.360000 0.00010 10.0000    0.01   1  1      0   'TIED02' 'NONE'
'FCWLND' 310  1  4.360000 0.00010 10.0000    0.01   1  1      0   'TIED02' 'NONE'
'FCWLND' 312  1  4.360000 0.00010 10.0000    0.01   1  1      0   'TIED02' 'NONE'
'FCWLND' 333  1  4.360000 0.00010 10.0000    0.01   1  1      0   'TIED02' 'NONE'
'FCWLND' 335  1  4.360000 0.00010 10.0000    0.01   1  1      0   'TIED02' 'NONE'
'FCWLND' 357  1  4.360000 0.00010 10.0000    0.01   1  1      0   'TIED02' 'NONE'
'FCWLND' 359  1  4.360000 0.00010 10.0000    0.01   1  1      0   'TIED02' 'NONE'
'FCWLND' 382  1  4.360000 0.00010 10.0000    0.01   1  1      0   'TIED02' 'NONE'
'FCWLND' 384  1  4.360000 0.00010 10.0000    0.01   1  1      0   'TIED02' 'NONE'
'FCWLND' 408  1  4.360000 0.00010 10.0000    0.01   1  1      0   'TIED02' 'NONE'
'FCWLND' 410  1  4.360000 0.00010 10.0000    0.01   1  1      0   'TIED02' 'NONE'
'FCWLND' 436  1  4.360000 0.00010 10.0000    0.01   1  1      0   'TIED02' 'NONE'
'FCWLND' 435  1  4.360000 0.00010 10.0000    0.01   1  1      0   'TIED02' 'NONE'
'FCWLND' 438  1  4.360000 0.00010 10.0000    0.01   1  1      0   'TIED02' 'NONE'
'FCWLND' 465  1  4.360000 0.00010 10.0000    0.01   1  1      0   'TIED02' 'NONE'
END
```

1st record: debug print options
2nd record: identifies that no function file is used (See UCODE guidelines for “pre” file).
3rd record: Format (using C+ convention) to write parameter values.
4th and other records: Separate line for each parameter change. A given paracode can be repeated for different material number or zones.

If a single multiplier is to be used for several parameters, user can use the “TIED01” paracode to tie together a group of paracodes associated with one common factor. The “TIEDxx” paracode designation is specified in column eleven.

A.3 LP3UCODE Application

The standard way of running simulations with CFEST involves the sequential use of the main simulation programs: LPROG1, LBAND, LPROG3I, and LPROG3. In the standard version of the

CFEST, parameter modifications are made by editing LP1 and L3I files and re-running LPROG1 and LPROG3I, the main input data processing programs prior to running LPROG3, the main simulation program.

LP3UCODE is an updated version of LPROG3 for UCODE applications which:

- Reads < >.hed, < >.riv, < >.cfo, and < >.sub file from lp3ucode.ctl (generated by CFUCODE and UCODE)
- Updates properties and flux rates using parameters listed in < >.sub (generated by CFUCODE and updated by UCODE).
- Performs simulations for time steps defined in cfest.ctl
- Extracts simulation results for the observation data points and updates < >.cfo, < >.hed, < >.riv and other output files.
- If variable IOPT_HEAD is not less than zero the CFEST binary result files are also written for processing input and output by TecPlot, EVS, and CFEST support programs.

Following is a copy of lp3ucode.ctl for Hanford test problem.

```
hanford.sub                          Parameter substitute UCODE output
hanford.cfo                          Lp3ucode results for UCODE output
hanford.hed                          CFUCODE file for head extraction
NONE                                 CFUCODE file for riv. extraction
NONE                                 CFUCODE file for conc extraction
```

Comments on right hand side of each record identify the file content. LP3UCODE first reads < >.sub file and updates parameter values. For each parameter change, LP3UCODE generates paracode.dbg (debug output). This file is overwritten with during each new run of LP3UCODE. Initially users are advised to edit the "dbg" files to check proper update of a given parameter. LP3UCODE updates < >.cfo, < >.hed, and < >.riv and if iopt_hsave is negative, LP3UCODE does not write normal binary result files.

Users are advised to rerun LP3UCODE after successful completion of UCODE. As can be illustrated by running test problems, UCODE final < >.sub file includes optimized values of parameters. Final results can be viewed in< >.hed and < >.riv files. The batch file,.lp3ucode.bat, executes the program. Section 4 discussed the interactive input to this batch file when running UCODE in parallel.

A.4 References

Cole CR, SB Yabusaki, and CT Kincaid. 1988. *CFEST-SC, Coupled Fluid, Energy, and Solute Transport Code, SuperComputer Version, Documentation and User's Manual*. Battelle, Pacific Northwest Laboratories, Richland, Washington.

Gupta SK. 1997. *Draft User's Manual, CFEST-96 Flow and Solute Transport, Constant/Variable Density, Computationally Efficient, and Low Disk PC/Unix Version*. Consultant for Environmental System Technologies, Irvine, California.

Gupta SK, CR Cole, CT Kincaid, and AM Monti. 1987. *Coupled Fluid, Energy, and Solute Transport (CFEST) Model: Formulation and User's Manual*. BMI/ONWI-660, Battelle Memorial Institute, Columbus, Ohio.

Hill MC. 1998. *Methods and Guidelines for Effective Model Calibration*. Water Resource Investigations Report 98-4005, U.S. Geological Survey, Water Resources Division, Denver.

Poeter EP and MC Hill. 1999. *Documentation of UCODE, A Computer Code for Universal Inverse Modeling*. U.S. Geological Survey Water Resources Investigations, Report 98-4080.

Appendix B

MasterTasker—A Perl Program for Management of Distributed Tasks

Contents

Figures

Tables

Appendix B
MasterTasker—A Perl Program for Management of Distributed Tasks

B.1 Introduction

B.1.1 What is MasterTasker?

MasterTasker is a task manager that allows a user to run multiple tasks simultaneously on multiple networked systems. The program distributes the tasks to the available systems and retrieves all results.

MasterTasker is a Perl[(a)] program that runs on all participating systems. The main server, which acts as the master, runs the code in the default mode. The same code, using the "slave" mode, is run on each of the slave systems. The specific software and connectivity prerequisites are discussed in Section B.2 of this appendix. The program is started from the command line and runs continuously, monitoring the specified directory for communication. All communication, between slaves and master, between master and MRGT, and between master and administrator, is done via simple ASCII files. The file naming conventions are very specific and are discussed in the Communication sections of this document. The master tracks current tasks, specific job progress, and slave availability and keeps all information in ASCII files. By using this simple form of management, the system can be easily restarted.

When a job is requested by the Multiple Realization Generator Task (MRGT), the process is started. This may be a complex program that will analyze the results and submit subsequent job requests or a simple batch file that follows certain criteria. This request starts a dialogue between the master and the MRGT that continues until the job is either completed or aborted. When MasterTasker and MRGT have established and confirmed their initial communication, MasterTasker begins its management phase. The master determines the systems currently available to act as slaves, assigns them jobs, and monitors their progress until the Slave system reports completion. After each job is completed, the master retrieves the results and, if necessary, assigns another job to the system. This process continues until all jobs have been successfully completed. At this point, MasterTasker informs the MRGT and awaits further instructions. During this dialogue, the MasterTasker remains open to other task requests.

B.1.2 Background

In modeling complex systems, it is necessary to first perform a model calibration. In this mode, the model/code is run many times with slightly different parameter values to determine the best value that produces the results that most closely match historical data. The model/code must be run with each of the test values—either sequentially on a single system or simultaneously on different systems. The results are then analyzed, and, based on the analysis, another set of runs may be started with different test values or a different parameter may be tested. Each of the runs is a separate task that must be performed.

[a] Practical Extraction and Report Language - -A Freeware Language Designed For Text Manipulation.

MasterTasker allows the runs to be made simultaneously using networked computer systems so the process can be completed quicker. Also, computer efficiency is increased because systems can be used during slack times (night and weekends).

B.1.3 Conventions Used in this Manual

Sample file listings are written in the `courier font`. The input parameters are discussed in the generic form; anything enclosed in angled brackets represents a value that the user will provide. For example, the shared directory—the location that all systems must have both read and write access to—is referred to as the <TASK_DIR>. The actual value might be D:\projects\MasterTasker. Examples are given in the discussion.

This document refers to jobs and tasks. In this context, a task is composed of multiple jobs. The MRGT will submit a task request. This task may consist of any number of realizations, each of which is considered a job. The MRGT is only concerned with task completion. The slave systems, where the individual modeling runs or realizations are processed, are only concerned with the specific job they have been assigned.

B.2 Prerequisites

B.2.1 Software

MasterTasker is written in Perl and you must have Perl installed on all systems. If you are running on a Windows system, version 5.6 of the Windows Operating system is required. This is the first version that provides a fork() emulation in the non-UNIX world. You can download the latest ActivePerl binaries from http://www.ActiveState.com/Products/ActivePerl/index.html.

B.2.2 Connectivity

The master system must be accessible to all available slaves. It must allow *simultaneous* connections to all of the current slaves. If you are running on Windows NT and plan on using more than ten slave systems, the master must be running NT Server operating system. NT Workstation will only allow 10 simultaneous connections. You will need to purchase enough licenses for the NT Server to support the desired number of slaves. If your master is a UNIX system, this is not an issue.

B.3 MasterTasker

B.3.1 Setup Overview

MasterTasker runs on each of the systems “listening” for assignments. On the system designated as the master, the program listens for communication from the MRGT and the slaves. On the systems running in the slave mode, the communication is from the master. The first step in the setup procedure is to determine which system is going to be the master. Then each system must be configured, which involves the creation of the initialization file taskr.ini. When the system has been configured, the code

runs from the command line. The master system is now ready and waiting for a task request from MRGT, and the slave systems are ready for a job assignment from the master. Each of these steps is discussed in this section.

B.3.1.1 Designate the Master System

The first step in setting up MasterTasker is deciding which system will be the master. The system is not necessarily going to do any modeling, so speed is not an issue. Connectivity and disk space are the priorities. The system must be capable of connecting to all of the currently available slaves at the same time. Although we have tried to make MasterTasker as robust as possible by storing all information in flat files that can be used to restart the system, the most reliable system should be the master. Space is the second issue—which system will have the most available room? All output files from the running jobs will be retrieved and stored on the master's system until MRGT has finished the analysis and signaled completion. This can add up to lots of space—depending on what the model/code is actually doing.

The final step to setting up the master system is to share the file system or folder that will act as the central shared space. All participating systems must have both read and write access to this directory. You must also create a sub-directory that will act as the holding area for the slave systems; the name of the directory is left up to you. After designating the master system, you need to install and configure the software on each of the participating systems.

B.3.1.2 Install the Software

Each of the systems, both master and slave, must be running a copy of the code. MasterTasker consists of the Perl program taskr.pl and the mtUtil.pm module. These programs must be copied to each of the systems that will participate in the process. Create a directory for MasterTasker and copy the program and module to this directory. The program will be launched from this directory. Each of the systems must then be configured.

B.3.1.3 Configuration

Configuration consists of designating the communication directory and the directories that will contain the management files. In addition, there are parameters that control the logging functions and describe the system, such as the name and the type of operating system. The parameters are used to create the initialization file.

After you have determined which system will act as the master, you must designate the sub-directory that will be monitored for communication files. This sub-directory must be shared to other systems with both read and write permissions. As an example, suppose the name of the master system is Wd23912. Figure B.1 is a snapshot of the setup on a slave system named computer02.

The arrow points to the communication directory, which was shared from the master (Wd23912) using the share name "testshare" and mapped to the local system as drive J:\. Below the arrow is a listing of the C:\mastertasker\slave-micki directory. This is the local directory that will contain all of the slave's management files, the <HOME_DIR>. To the right is the initialization file that reflects this setup. The TASK_DIR parameter is the full path to the directory; in our example, it would be J:\shared-micki. The sub-directory named "slaves" will serve as the holding area for job assignments.

Table B.1 lists all parameters and a brief description. If command line usage is supported, the associated switch is shown in the last column. Following the table, each of the parameters is discussed.

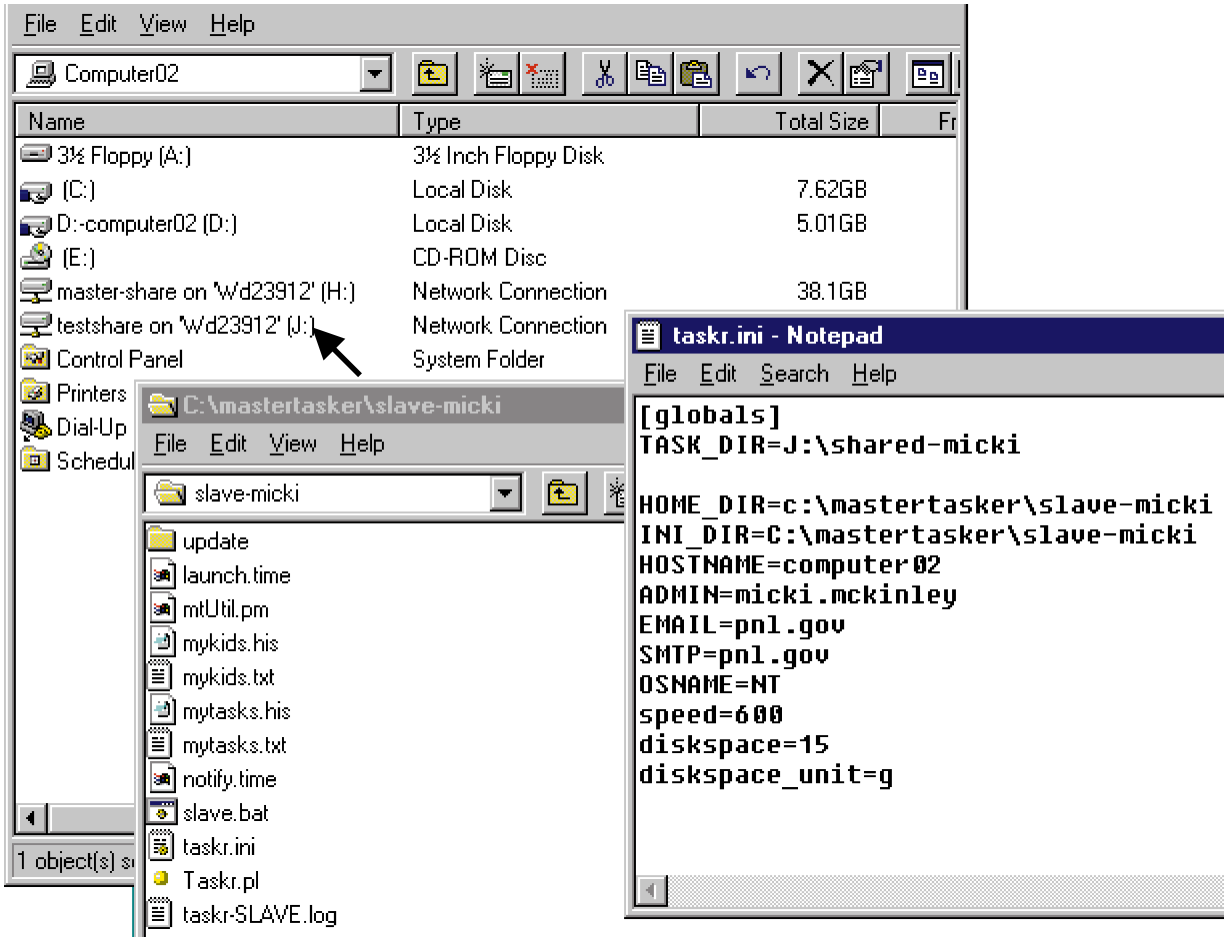

Figure B.1. Configuration of a Slave System

B.3.1.4 Parameter Definitions

Table B.1. Summary of Parameters

Parameter Name	Description	Switch
	prints usage message and exists	-? -u
	prints current version and exists	-v
	exit when messages are received out of order	-e
	interactive help	-h
	run as master	-m
	run as slave	-s
BATCH_FILE	a text file containing parameters	-B
DEBUG_LEVEL		-D
HOME_DIR	local home directory - not a shared area	-H
INI_DIR	directory containing the initialization file and log file	-I
TASK_DIR	shared directory accessible to master and all slaves	-T
TRACE_STRING	controls the amount of detail sent to screen and log file	-R
PROCESSOR	Processor number	-P
OSNAME	type of operating system [Windows2000|Windows98|NT|Unix]	-O
HOSTNAME	system name	-N
DOWNLOAD_SRC	location of shared source directory for software distributions	
ADMIN	user name to mail system messages to in the event of problems	
EMAIL	administrator address	
SMTP	outgoing mail server	
SPEED (s)	cpu speed of the slave system	
DISKSPACE (s)	available diskspace on local system	
DISKSPACE_UNIT(s)	unit for diskspace variable	

(s) = slave systems only

All parameters are optional unless otherwise noted.

HOME_DIR Full path to the local home directory
(example: D:\MasterTasker\master)
(command line example: -HD:\MasterTasker\master)

This is the location of all management files. On a slave system, this is the working directory where the jobs are actually run. The Management Files section of this document discusses the files themselves. This directory is not accessed by anyone but MasterTasker; it does not have to be shared. This variable is mandatory.

TASK_DIR Full path to shared directory on the master system
(example: H:\shared)
(command line example: -TH:\shared)

The shared directory is located on the master system and shared to all participating systems. The <TASK_DIR> variable is the local path to this shared resource. The slave system creates the local path by either mapping the drive (assign a drive letter to the network folder) or mounting the file system. The full path, including any drive specification, is entered. For example, assume the master is named primeur and the folder C:\MasterTasker\allshare has been shared with the share name of "taskr." On the local system, the user would map \\primeur\taskr to the local drive H:\. In this directory, there is a subdirectory named "share," which contains the subdirectory "slaves." The variable definition would be

TASK_DIR=H:\share

All communication takes place in this directory in the form of ASCII files that are written by the slave to the master or written by the master to the slaves. The ASCII files, or messages, are discussed in the Communication sections of this document. This variable is mandatory.

Enter the full path to the shared resource. Please see the TASK_DIR entry above for a discussion of mapping network drives. The "slaves" subdirectory must be an existing sub-directory in the <TASK_DIR> directory; MasterTasker will not create the sub-directory.

INI_DIR Full path to the local directory that contains the initialization file
(example: C:\projects\MasterTasker\local-slave)
(command line example: -IC:\projects\MasterTasker\local-slave)

The program uses the initialization file taskr.ini for setting parameters. The INI_DIR contains the initialization file and the log file. It can be the same directory as the HOME_DIR (see above).

If this path is not set on the command line, the program will search for the taskr.ini file as a means to establish the location. If the <HOME_DIR> variable has been set and a taskr.ini file is found, the <INI_DIR> will be set equal to the <HOME_DIR> value. As a last resort, the program will assign the <INI_DIR> equal to the directory where the program was launched (the current working directory). The log file is written to this directory.

OSNAME Operating System
(example: NT)
(command line example: -ONT)

This variable has four options: Windows98, Windows2000, NT, and Unix. The value is not case specific: Unix, UNIX, and unix will all be interpreted the same.

HOSTNAME Name of the computer system
(example: catch22)
(command line example: -Hcatch22)

The only special character allowed in the computer name is the underscore "_". If your computer supports the "hostname" command, MasterTasker can use this to determine the name of the computer. If your computer does not support the hostname command, you must set this variable.

TRACE_STRING Controls screen output and log file contents
(example: 00110011)
(command line example: -R00110011)

This is a eight-character string composed of zeros (0) and ones (1). It controls the amount of information that is sent to the screen and written to the log file: "0" toggles the trace message off, and "1" toggles the message on (see Table B.2).

Table B.2. Seven Positions and Corresponding Actions

Position	Action
1	Echoes subroutine names as they are called
2	Echoes the arguments passed into the subroutines (probably want 1 if 2)
3	Echoes incoming messages – format: message [filename]
4	Echoes outgoing messages – format: message [filename]
5	Echoes hash update information
6	Echoes synchronization messages
7	Retains a history of hash updates
8	Echoes timing information – for reassignment issues

The default setting is 00110000 and writes input and output messages to the log and screen.

This variable can be very helpful if you have to do any troubleshooting. For example, if you are building a MRGT program and things do not seem to be working correctly, you can see what messages are being sent and received; maybe an expected message is not received, or the messages are in the wrong order.

The last parameter is used during the reset process; see the Administrative Request section for a discussion of the reset process. If the user chooses to Update Software, all files located in the DOWNLOAD_SRC directory will be copied to the INI_DIR directory.

DOWNLOAD_SRC path to the shared directory that contains software for downloading
(example: J:\shared\bin\nt)

The remaining parameters are not currently used; they will be implemented in future versions and are included here for completeness.

The next three variables are used for email notification when an error has occurred. The administrator's email address is ADMIN@EMAIL. The SMTP parameter is the outgoing mail server. If you use Netscape, you can locate the value via Edit →Preferences → Mail & Newsgroups →Mail Servers. The value is listed as the Outgoing mail (SMTP) server, or you can contact your system administrator for this value.

ADMIN user name for administrator to notify
(example: jerri.jones)

EMAIL	administrators domain – used to construct email address (example: pnl.gov)
SMTP	outgoing mail server (example: smtp.mines.edu)

The next three variables are used as assignment criteria. All currently available slave systems will be rated according to their speed; the fastest will be assigned first. In the same manner, only slaves with sufficient diskspace will be assigned tasks. This is a planned future enhancement.

SPEED	the speed of the cpu (example: speed=600)
DISKSPACE	diskspace available on the slave system (example: diskspace=14)
DISKSPACE_UNIT	unit of measure for the diskspace variable (example: diskspaceunit=g)

B.3.1.5 Parameter Input

MasterTasker will either accept parameters from the command line or read them from a batch file or from the initialization file, or a combination of all three. The order of precedence, starting at the top, is the command line, the batch file, and the initialization file. For example, if you have completed your initialization file but wish to run with a different variable setting, you can use a temporary batch file. Any variables set in the batch file will override the setting in the initialization file. Alternatively, you can run the program using a switch to set a variable, and these settings will override any parameter settings in the initialization or batch file.

B.3.1.6 Initialization File—taskr.ini

MasterTasker uses an initialization file called taskr.ini. This ASCII file is located in the <INI_DIR> directory. Each of the variables should be written in this file using the format variable=value, with a single variable assignment per line.

The beginning of the variable section must begin with the word globals in square brackets: [globals]. Any comments (lines that begin with #) will be ignored. The taskr.ini file for an NT slave system is shown below.

```
# taskr.ini - initialization file
# example file for an NT slave system
#
# the following line is MANDATORY - it is case specific.
[globals]
#
# task_dir is a directory in shared space.
```

```
TASK_DIR=H:\mastertasker\share
#
# home_dir is a local home directory for system.
HOME_DIR=e:\projects\mastertasker\testing\local-slave
#
# ini_dir is a local directory that contains the initialization file taskr.ini.
INI_DIR=e:\projects\mastertasker\testing\local-slave
#
# name of the system - no special character allowed except the underscore "_"
HOSTNAME=slave1
#
# the options are:  NT | Windows98 | Windows2000 | Unix [not case specific]
OSNAME=NT
#
# admin, email, and smtp are used to send error notification to the administrator.
ADMIN=jane.user
EMAIL=pnl.gov
SMTP=smtp.mines.edu
#
# criteria for job assignments
SPEED=200
DISKSPACE=15
# DISKSPACE unit of measure: g {gigabyte} | m {megabyte} [not case specific]
DISKSPACE_UNIT=g
#
```

Tip – be consistent with the drive mappings, and this will reduce the necessary modifications to the initialization file; this can really save time if you are setting up a large number of slave systems.

B.3.1.7 Batch File

A batch file can be used to run MasterTasker. The -B switch is used on the command line to pass the name of the batch file to MasterTasker. The file must exist in the master's shared space (<TASK_DIR>). All parameters can be read from the batch file.

B.3.1.8 Command Line Switches

If you enter the parameters via the command line, you must use the one character switch shown in Table B.1. These switches can only be used on the command line. They are case sensitive, and there is NO SPACE between the switch and its associated parameter. Please see the parameter discussion for examples.

B.3.1.9 Launching MasterTasker

If you have been following the steps, all of the systems should be configured. You have prepared the initialization file and a batch file if you are going to use one. The last step is to actually start the code. On the master system, you use the following command:

```
perl taskr.pl
```

From the slave systems, you enter the following command:

```
perl taskr.pl –s
```

B.3.2 Management Files

MasterTasker uses ASCII files to keep track of all job information. The files are stored in the location specified by the parameter <HOME_DIR>. Each line represents a single record and describes a unique job or task. In addition, a log file is written named taskr-<mode>.log where the mode is either "MASTER" or "SLAVE." You can control the amount of detail written to the log file via the TRACE_STRING variable. A brief discussion of the files is presented in this section.

B.3.2.1 Running as Master

The master creates four files to keep track of all information. The log file is named taskr-MASTER.log.

phase1.db — This file contains all information regarding initial job requests – communication between MRGT and the master.

phase2.db — After a successful job request, the job is considered active. The current status and other information are kept in this file.

phase3.db — This file contains all information about the individual tasks that make up the job.

slave_reg.txt — When a new slave becomes available, it sends the master certain parameters regarding its available diskspace, etc. This information is kept in this "registration" file.

B.3.2.2 Running in Slave Mode

When MasterTasker runs in the "slave" mode, it creates two files to track information about its current jobs. The log file is called taskr-SLAVE.log.

mytasks.txt — This file contains information on assignments that have been accepted by the slave system.

mykids.txt — This file contains information on the child that was spawned to execute the launch script and actually do the processing.

B.3.2.3 Administrative Requests

MasterTasker currently supports two administrative tasks. Both requests are made by placing a file, with a very specific name, in the master's shared directory <TASK_DIR>. The administrator can request a status report and reset the system in this manner.

If a file with the name *taskr-status* appears in the <TASK_DIR>, the program will generate a status report on tasks that are currently active. An example is shown in Figure B.2. In this example, there are

six jobs, numbered 0–5. Four of the six jobs have been completed by the two active slave systems: computer02 and computer03. The 5th job has been assigned to computer02, and the 6th job is still waiting in the queue.

If the file named *taskr-reset* appears in the directory, the program will remove all management files, both on the master and all active slave systems. This cleanup occurs even if the file *taskr-reset* is empty.

Additional options are determined by the contents of the file. The heading [reset] must precede any variable settings; any comments (lines that begin with #) are ignored. The options and allowed values are outlined in the Table B.3. All settings are optional. All variables and values are case specific (i.e., logfile is not the same as Logfile). The default values are shown in bold.

Table B.3. Options and Allowed Values of Variables

Variable	**Allowed Values**	**Action**
LogFile	DELETE, RETIRE	Delete or rename the existing log files
KillSlave	YES, NO	The program running on the slaves will terminate
SoftwareUpdate	YES, NO	Used to distribute new or updated version of software to slave systems
InitSlave	YES, NO	Instruct the slave system to re-initialize by reading the ini file
UpdateMaster	YES, NO	Used to distribute new or updated version of software to the master
InitMaster	YES, NO	Instruct the master to re-initialize by reading the ini file

```
[reset]
KillSlave=YES
LogFile=RETIRE
SoftwareUpdate=NO
```

The sample file listing shown on the left instructs the slaves to retire the existing logfile: the original logfile will be renamed taskr-SLAVE.old, and a new logfile will be created. Finally, the MasterTasker program running on the slave will exit.

```
REPORT FOR TASK 8953857

DETAILS

4 job_completion_acknowledged computer02
5 job_completion_acknowledged computer03
0 job_completion_acknowledged computer02
1 job_completion_acknowledged computer03
2 assigned computer02
3 queue no comment

SUMMARY

queue: 1
assigned: 1
confirmed: 0
awaiting_completion: 0
job_completion_acknowledged: 4
```

Figure B.2. Report for Task 8953857

B.4 Multiple Realization Generator Task (MRGT)

MasterTasker provides a service, but it is the MRGT that makes the job request and starts the whole ball rolling. MRGT provides all the information necessary to run the job, including all input files, any programs that will actually be executed (or its location if the program already resides on the slave systems), and a launch script that starts the process. The launch script, a batch file (extension: csh or sh) or perl program (extension: pl), will be executed by MasterTasker on the slave system.

B.4.1 MRGT Communication

MRGT must be able to communicate with the master. It must know the location of the master's shared directory and be capable of reading from and writing to this directory. It must also know the "language," the dialogue necessary to communicate with the master. A very specific exchange of messages must take place between the MRGT and the master system. A missing, or out of order, message can cause the entire job to be canceled. The dialogue has three phases: 1) the request phase, 2) the management phase, and 3) a cleanup phase.

All Phase 1 communication is written to the <TASK_DIR> directory, and all filenames use a timestamp in the filename. The timestamp is set by MRGT and is the time that the initial request is generated. The job request from the MRGT is the first step in the process.

timestamp format is: MMDDYYHHMMSS - using a 24-hour clock (GMT)
where:
MM = month range: 01-12
DD = day range: 00-31
YY = year range: 00-99
HH=hour range: 00-24
MM=minute range: 00-60
SS=second range: 00-60

example: 032800092033.req

B.4.1.1 Phase 1—Initial Job Request

The dialogue goes like this:

Step 1. The MRGT sends a request by writing a file in the shared directory <TASK_DIR> named <timestamp>.**req**.
Step 1 starts the entire process. This initial file contains no information; it is the timestamp and the extension that are important.

Step 2. MasterTasker sends a reply in the form of a file named <timestamp>.**rpl**.
When MasterTasker sees the <timestamp>.req file, it assigns a unique task_id to the request, and the <timestamp> is associated with this unique job_id. It sends this task_id back to the MRGT in Step 2 as the only contents of the file <timestamp>.rpl.

Step 3. The MRGT acknowledges receipt of the request by writing a file named <timestamp>.**rcv**.
In Step 3, the MRGT reads the job_id contained in the <timestamp>.rpl file. It then writes a file, called <timestamp>.rcv, containing the same task_id, back to the MasterTasker.

Step 4. MasterTasker either confirms the request by writing a file named <timestamp>.**go** or aborts the process by writing a file named <timestamp>.end.

If the task_id, returned by the MRGT in Step 3, is the same as the task_id assigned by MasterTasker, then the file <timestamp>.go is sent to the MRGT as an acknowledgment. The MasterTasker creates the directory with the name <task_id>. This directory is assigned to the specific job for the duration of its existence. It is always located in the MasterTasker shared directory (TASK_DIR).

The job has been assigned and confirmed. Phase 1 has been completed. If the job_id is not the same, then the process fails, the file <timestamp>.end is written, and the entire job is canceled.

B.4.1.2 Phase 2—Management

Assuming the successful completion of Phase 1, the dialogue continues with the next phase – management of the actual job. All subsequent communication between MRGT and the MASTER use the <task_id> and the name of the sender to construct the filename. Message files from MRGT to the master are named <task_id>-mrgt and from the Master to MRGT are named <task_id>-master. These messages are written to the directory created specifically for this task. For example, if the unique task_id assigned by the master is 78865644, and the <TASK_DIR> is H:\share, the messages will all be written to the H:\share\78865644 directory.

Step 1. MRGT writes the file <task_id>-mrgt that contains the message "**creating_pf**."
MRGT creates a parameters file named <task_id>.parameters and puts it in the directory that has been assigned to this job (Step 4 of Phase1). The parameter file contains the variables necessary for setting up and managing the tasks that make up the requested job. The parameters are discussed in the following section of this document.

Step 2. MasterTasker writes the file named <task_id>-master containing the term "**waiting_pf**."
This message is more of tracking device than a necessary step in the dialogue process. MasterTasker is really waiting for the instruction to "check_pf."

Step 3. MRGT writes the file <task_id>-mrgt that contains the term "**check_pf**."
If MRGT encounters problems creating the parameters file, the contents of this message may read "pf_creation_error_clean" or "pf_creation_error_leave."

Step 4. MasterTasker writes the file named <task_id>-master with the contents "**checking_pf**."
The MasterTasker must find the parameters file in the directory assigned to this specific job and titled <task_id>. The file is checked for any missing variables. (*Note: MRGT should not consider this message mandatory – the process of checking the parameters may be so fast that before MRGT reads the file, the message will have been replaced*.)

Step 5. MasterTasker writes the file named <task_id>-master with the contents "**accept_pf**."
If the parameters file is OK, the master informs the MRGT that all necessary parameters have been received by writing the response "accept_pf." An example parameters file is shown in Figure B.3 with a snapshot of the directory, as populated by MRGT. In this example, the job has been assigned the task_id 13601684, and the TASK_DIR is E:\testing\share.

If the parameters file is not acceptable, the response is "reject_pf," the name of the parameter file is changed to <task_id>-param-rejected, and it is placed in the <TASK_DIR> directory.

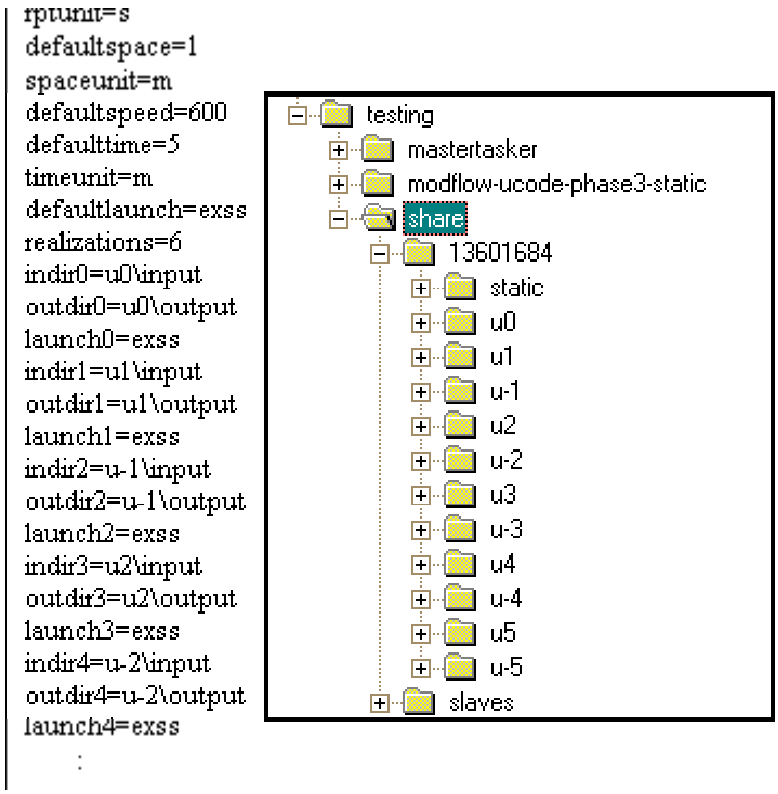

Figure B.3. Example Parameters File

Assuming the parameters file is accepted, the process continues with the following steps.

Step 6. MRGT writes the file <task_id>-mrgt that contains the message "**waiting_results_mr**." MRGT must wait for all of the tasks to complete. It continues to monitor the file named <task_id>-master for incoming messages.

Step 7. MasterTasker sends status reports to MRGT in the file <task_id>-master.
The first line of the file contains the term "**processing_mr**." The second line contains information about the job progress: the number of slaves currently working on the job, the number of tasks running, the number of tasks in the queue, and the number of tasks actually completed. In the following example, there are 12 tasks and 3 slaves available. Five of the jobs have been completed, three are running, and four are waiting in line.

example: processing_mr
3 5 4 5

B.4.1.3 Phase 3—Analyze Results

When all tasks have been completed, the master reports to MRGT. MRGT then does any analysis of the results. Depending on the outcome of the analysis, MRGT may want to generate another set of parameters and run another set of tasks, or it may decide the job is complete.

MasterTasker writes the file <task_id>-master with the contents "processing_complete."

MRGT writes the file <task_id>-mrgt containing the term "analyzing_results_mr."

MasterTasker writes the file <task_id>-master with the term "waiting_analysis_results_mr."

MRGT writes the <task_id>-mrgt "creating_pf" if another suite of tasks is to be done. If the job is complete, the file will contain the term "processing_done_clean" or processsing_done_leave." If MRGT determines that another set of tasks is necessary, the process starts over with Step 1 of Phase 2.

B.4.2 Task Parameters

The parameters file contains the parameters associated with the tasks to be performed in parallel. This section describes the nature and contents of this "parameters file." After each of the parameters has been defined, some example files are presented for illustration. The parameters and a brief description are shown in Table B.4.

B.4.2.1 Parameter Definitions

Table B.4. Parameter Names and Descriptions

Variable Name	Description
realizations (m)	number of realizations (individual jobs) that make up the task
defaultrpttime (m)	
defaulttime (m)	
defaultspace (m)	
defaultspeed (m)	
defaultindir	
defaultoutdir	
defaultlaunch	
indir	input directory for a specific realization
outdir	input directory for a specific realization
launch	name of the script that will start the process

The "parameters file" uses the format convention of: <name>=<value> to allow for maximum expandability and flexibility. Each line must contain a single variable, value pair. All parameter names are lower case.

The parameters file contains default information for those parameters that do not vary with each realization such as:

defaultruntime	the default run time requirement to run a realization task (example: defaultruntime=1440)
runtimeunit	units that describe the defaultruntime variable (example: runtimeunit=h)

The options are “m” for minute (the default) or “h” for hour.

defaultindir	the default input directory for a realization task (example: defaultindir=hanford)

This might be applicable for setup tasks where the only function of MRGT is to distribute software to a specific directory.

defaultoutdir	the default output directory for a realization task (example: defaultoutdir=results)
defaultlaunch	the default launch code for running a realization task (example: defaultaunch=hanford_setup).

Do not specify the extension unless the script is a perl program; in that case, the extension must be “.pl..” MasterTasker adds the appropriate extension based on the platform: csh for UNIX and “.bat” for Windows systems. This allows you to use the same script name for different platforms and goes towards platform independence.

defaultrpttime	the default monitoring time for checking the status of a realization on a slave computer (example: defaultrpttime=10)
defaultspeed	the default speed estimate upon which the run time is based for a realization task (example: defaultspeed=500)
speedunit	units that describe the defaultspeed variable (example: speedunit=m)

The only option currently supported is m for mhz.

The parameters file contains specific information that varies with each realization. Obvious examples are the directory names for each realization. In the following, <num> is used to represent a realization number.

rpttime<num>	the monitoring time for checking the status of a specific realization on a slave computer

rpttimeunit<num> unit used to describe the rpttime<num> variable

launch<num> the default launch code for running a realization task
(example: launch4=runit)

This script must place all the realization output in a sub-directory at the same level as “input” named “output.” At the completion of the realization task on the slave machine, the “output” sub-directory will be moved into the realization directory outdir<num>.

To be platform independent, the input directory for each realization should have both a “.bat” file and a “.csh” launch file. That way, the slave machine will know how to launch the code properly based on the type of platform of the slave system. Typically, the launch script would be set using the default method above. However, in the case of just plain multiple runs, each may be different.

NOTE: Even if all realizations use the same script and it is the designated defaultlaunch, a copy of the script must be included in each of the realization directories.

The following example might be relevant for a composite analysis set of runs.

launch1=tritium
launch2=sr90
launch3=uranium
…
launch6=chromium

indir<num> the directory name for a specific realization task
(example: indir8=ucode)

This directory contains all the input files and directory structure needed to make a specific realization run. The launch script must be located in this directory.

outdir<num> the output directory for a specific realization
(example: outdir7=p3out)

After the task has been completed, the output will be retrieved from the slave system and placed in this directory. MRGT will use it for analysis.

realizations the number of realization tasks to be run in parallel
(example: realizations=6)

Each realization must have a defined input directory (either indir<num> or defaultindir), an output directory (either outdir<num> or defaultoutdir), and a launch script (launch<num> or defaultlaunch).

The following parameters are not currently used. They are included for future reference.

defaultspace	the default space requirement on the local disk to run a realization task (example: defaultruntime=1440)
spaceunit	units that describe the defaultspace parameter (example: spaceunit=g)

The options are m for megabyte (the default) and g for gigabyte.

B.4.2.2 Example Parameter Files

The first example uses all the default values. None of the units are specified because the defaults are acceptable.

```
realizations=4
defaultrpttime=10
defaulttime=55
defaultspace=10
defaultspeed=400
defaultindir=hanford
defaultoutdir=results
defaultlaunch=runcode
```

In the second example, the unit defaults are accepted, but an input and output directory is assigned for each of the realizations.

```
realizations=3
defaultrpttime=20
defaulttime=55
defaultspace=2
defaultspeed=400
indir1=u0\input
indir2=u1\input
indir3=u2\input
indir4=u3\input
outdir1=u0\output
outdir2=u1\output
outdir3=u2\output
outdir4=u3\output
```

The third example demonstrates the most flexibility. A combination of default and realization-specific parameter values is used.

```
realizations=4
defaultrpttime=20
defaulttime==45
```

defaultspace=2
defaultspeed=500
defaultindir=hanford
defaultoutdir=results
defaultlaunch=runcode
indir1=tritium\input
indir4=special\input
outdir1=tritium\output
outdir4=special\output

In this example, the input and output directories for Realizations 2 and 3 are the defaultindir and defaultoutdir. The remaining input and output directories are realization specific. All realizations use the default launch script.

B.4.2.3 Launch Script

The launch script will be launched by a child of the slave and passed to the following input parameters:

output file	Full path of the output file for status reporting (format: <task_id>-child.txt)

When the process has been completed, the launch script must inform the slave; it does so by writing "process_complete" to this output file. This is the only way the slave system will ever realize that the process has been completed by the child process. When this file is read, the slave system reports the status to the master.

static directory	name of the static directory (example: 79580688_static)

The static directory is always located at the same level as the directory that contains the input and output sub-directories for the specific task. Using this directory saves both time and space on the slave system. The static directory contains all input files that are consistent for the extent of the task. This way, when a slave finishes a specific job and is assigned another, the master does not have to download another set of input files, wasting time and space. Only the dynamic input files, which will be written to the input directory, have to be moved to the slave.

job_id	unique number assigned to this job request (example: 8953857_1)

The job_id consists of the task_id and the realization number.

B.5 What Does a Slave System Do?

When the code is launched from a Slave system, the first thing it does is notify the master of its availability. It then monitors the shared space (<TASK_DIR>) for incoming messages from the master. All communication files from the master contain the name of the slave (<slave_id>) in the file name. It

will only recognize mail that is specifically addressed to it. The slave also monitors its home directory (<HOME_DIR>) for messages from its children. There are three phases of communication: a request phase, a processing phase, and a cleanup phase.

B.5.1 Slave Communication

B.5.1.1 Phase 1—Initial Request

1. MasterTasker writes the file <task_id>_<slave_id>-assign in the shared space.

When the master receives a job request from MRGT, it determines which slaves are available for assignment. It sends a request to each of the slaves. The first line of the file contains the <job_id>, which is composed of the <task_id> and the realization number. If there are multiple jobs running, a slave may be requested to do multiple tasks. However, a slave can only run a single task for each job.

2. The Slave system writes the file <task_id>_<slave_id>-accept with the contents "assignment_received."

The Slave system reads the assignment file and responds with a message of acceptance. This message acts as a confirmation of the <task_id>, which is used in building all subsequent communication filenames.

3. MasterTasker writes the file <task_id>_<slave_id>-confirm.

The master confirms the task_id and that the job assignment is successful. This completes Phase 1 – the job has been assigned, accepted and confirmed. The Slave system now begins the actual processing.

B.5.1.2 Phase 2—Processing

After completion of Phase 1, the file-naming convention changes. All files from the Slave system to the master are named <task_id>_<slave_id>-slave, and all files from the master to the Slave system are named <task_id>_<slave_id>-master. The files are copied from the directory created for the specific slave in the slave holding area to the local working directory on the Slave system (<HOME_DIR>), where all processing occurs. If a static directory is being used that does not already exist on the slave, a copy is made.

1. The Slave system writes the file <task_id>_<slave_id>-slave with the contents "processing." If the Slave system has some problem, the contents may be "processing_failed."

2. MasterTasker responds with the file <task_id>_<slave_id>-master which contains the term "waiting_results."

3. The Slave system writes the <task_id>_<slave_id>-slave file with the contents "process_complete."

When the Slave system receives notification of completion from the child process that was spawned to execute the launch script, it informs the master. When the process is complete, the output files are returned to the shared space; the input files are not returned. The working directory for this task is removed, and the static directory remains in anticipation of further assignments.

B.5.1.3 Phase 3—Cleanup

When the MRGT informs the master that the entire process is done, all slaves who worked on the job are informed. At this point, the static directory will be deleted.

B.5.2 Processing the Job

As each job is assigned, MasterTasker will copy the input sub-directory from the specific realization directory to the directory created for the specific slave in the slaves' holding area (<SLAVE_DIR>) [Step 1 in Figure B.4]. When the slave has accepted the task, it will download the input files from its holding area to its local disk space [Step 2], and the launch script will be executed. Upon completion of the task, the slave will copy the "output" directory back to the slaves' shared space. The master will then move the files to the appropriate MRGT output directory.

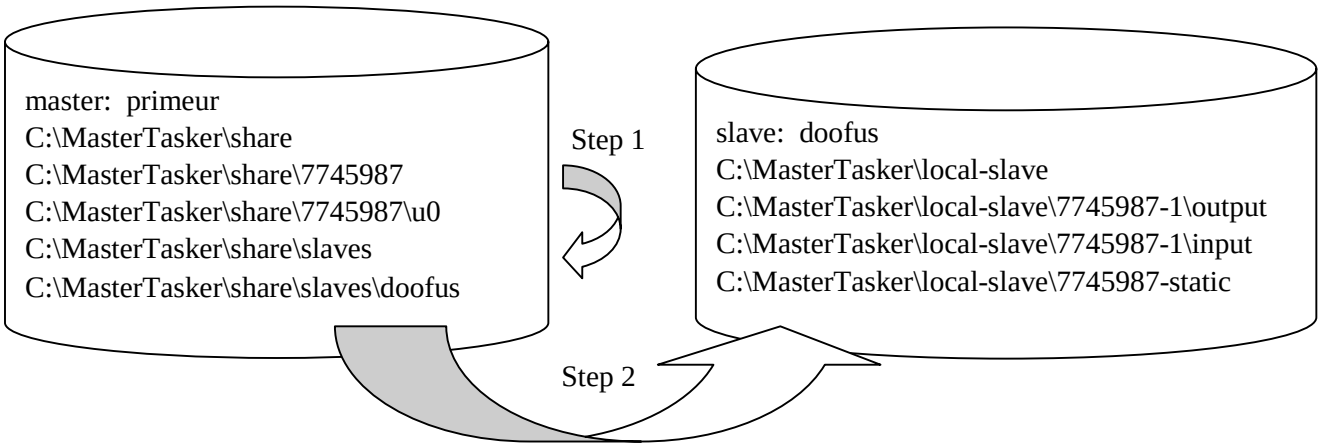

Figure B.4. Step 1 and Step 2 in Processing the Job

B.6 Putting It All Together

After MasterTasker is launched, it keeps watch over its assigned directory (TASK_DIR) for incoming messages. All communication takes the form of files or messages that are written to the shared directory by the master, the MRGT, the slave systems, and the administrator. There are three types of messages MasterTasker recognizes: communications from a MRGT about a new or existing job, messages from slave systems regarding an existing task or a change in availability, and requests from the administrator. In this section, the dialog will be discussed from the angle of the actions that occur. This section will describe the full dialogue beginning with a slave system, announcing its availability and ending with the final instruction from the master to the slave systems to clean up and remove all remnants of the complete job. The dialog is written as if every step were successful. Files written by the master are shown in green, messages from MRGT are shown in blue, and communication from the slave is shown in violet. Message files from the child to the slave are shown in pink. In this example (Table B.4), we have a single slave named doofus.

Table B.5. Example Case

File Name	Contents	Discussion
H:\share\082700205859.req	none	Initial request from MRGT.
H:\share\082700205859.rpl	91677856	Reply from the master – contains the unique id for this job.
H:\share\082700205859.rcv	91677856	Receipt by MRGT – returns unique id for confirmation.
H:\share\082700205859.go	91677856 confirmed	Initial handshake process was successful. The master creates the job directory in the shared space: H:\share\91677856
H:\share\91677856\91677856-mrgt	creating_pf	MRGT begins creation of the parameters file.
H:\share\91677856\91677856-master	waiting_pf	Master acknowledges receipt from MRGT.
H:\share\91677856\91677856-mrgt	check_pf	MRGT informs the master that the parameters file has been written: H:\share\91677856\91677856-parameters
H:\share\91677856\91677856-master	accept_pf	Master informs the MRGT that parameters are accepted.
H:\share\91677856\91677856-master	processing_mr	The master creates the MRGT working directories as assigned by the parameters file in the slave share space, including: H:\share\91677856\u0\input H:\share\91677856\u0\output A copy of the static directory is put in the slave shared space: H:\share\slaves\91677856_static The master creates a directory for each of the slaves currently available: H:\share\slaves\doofus
H:\share\91677856_doofus-assign	91677856_1!doofus	Master assigns realization number 1 to doofus. It passes in the task_id, the identification number that has been assigned to the task, and the name of the directory that contains the files for this realization.
H:\share\91677856_doofus-accept	91677856_1!assignment_received	Doofus acknowledges assignment and returns the task_id for confirmation.

Table B.5. Example Case

File Name	Contents	Discussion
H:\share\91677856_doofus-confirm	91677856_1!confirmed	The master confirms the task_id. The slave has all the information necessary to locate and download the file necessary to run the realization. In our example (SLAVE_DIR=H:\share\slaves): H:\share\slaves\doofus\91677856_1\input
H:\share\91677856_doofus-slave	91677856_1!processing	End of Phase 1 – processing has actually started on the slave system.
H:\share\91677856_doofus-master	91677856_1!waiting_processing	The master confirms receipt of status report.
H:\share\91677856\91677856-mrgt	waiting_results_mr	MRGT checks in – still waiting for results…
H:\share\91677856\91677856-master	processing_mr 1 1 3 3	The master replies – still processing… Only have one slave, one job is currently submitted, three are in the queue, and three are complete.
C:\tasker\local\91677856_1-child	process_complete	Child process informs the slave that the process is complete.
H:\share\91677856_doofus-slave	91677856_1!process_complete	The slave informs the master that the task is complete. The output files are returned to the output directory reserved for doofus in the slaves shared area; the name is built in the same manner as the input directory: H:\share\slaves\doofus\91677856_1\output All files specific to this task are removed from the local disk, with the exception of the static directory.
Time passes… the remaining tasks are assigned and completed.		
H:\share\91677856\91677856-master	processing_complete	The master notifies the MRGT that all tasks have been completed for this job.
H:\share\91677856\91677856-mrgt	analyzing_results_mr	MRGT acknowledges receipt and proceeds with its analysis.
H:\share\91677856\91677856-master	waiting_results_mr	The master acknowledges MRGT and waits for further instructions.
H:\share\91677856\91677856-mrgt	processing_done_clean	The job is complete. MRGT could have replied with “creating_pf” in which case the dialogue continues with creation of a new parameters file.

Appendix C

Columbia River Boundary Condition Generation

Appendix C

Columbia River Boundary Condition Generation

C.1 Introduction

This appendix describes the procedure followed to gather the necessary data and run the Modular Aquatic Simulation System 1D (MASS1) for the Hanford Reach of the Columbia River to generate water-surface elevation boundary conditions for the transient inverse calibration of the Hanford site-wide groundwater model to conditions between 1943 to 1996.

The MASS1 model, developed at the Pacific Northwest National Laboratory (PNNL), is a one-dimensional, unsteady, hydrodynamic and water quality model for branched river systems (Richmond et al. 2000a, 2000b, and 2000c). This appendix also provides plots of input data, output data, and validation comparisons, as well as the sources (web sites, agencies, etc.) from which data were gathered. The following data were required for a time period from 1944 to the present: fore bay elevations at McNary dam and flows for the Columbia (below Priest Rapids Dam), Yakima, Snake, and Walla Walla Rivers.

C.1.1 Data Sources

The system modeled included the main stem of the Columbia River from the tail water of Priest Rapids Dam to the fore bay of McNary Dam, as well as the tributary inputs from the Yakima, Snake, and Walla Walla Rivers (Figure C.1). The fore bay elevations for McNary were downloaded from the U.S. Corps of Engineers (COE) dam operations data found at http:/net/gehenna/usr/files0/usace-data/pub/datarequest/columbia/mcnary/daily.

Fore bay elevation data were not available for the period between dam completion (1953) and 1961. Consequently, a fore bay elevation of 103.63 m (340 ft) for this period was used based on the documented full-pool elevation. For the time period before McNary's existence, a water-surface elevation of 82.3 m (270 ft) was used, which is a conservative (high) estimate based on the amount that McNary elevates the pre-dam water-surface elevation (about 24.38 m [80 ft]) and the water-surface elevation of John Day Reservoir immediately downstream.

The flow data for the Columbia, Yakima, and Walla Walla Rivers was downloaded from the U.S. Geological Survey (USGS) Washington National Water Information System -West (NWIS-W) data retrieval page: http://waterdata.usgs.gov/nwis-w/WA/

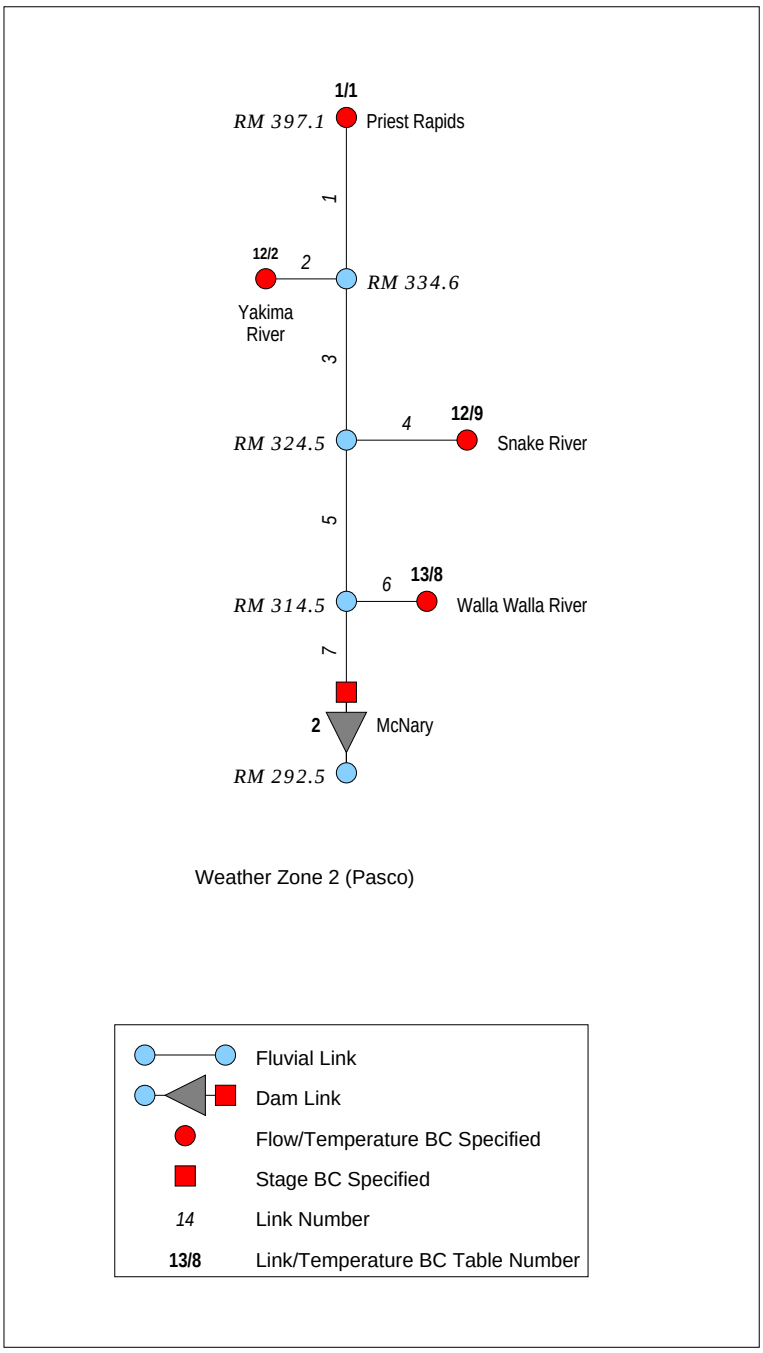

Figure C.1. Model Schematic for the Hanford Reach to McNary Pool

The flow data for the Snake River were created by combining operations data for Ice Harbor Dam (downloaded from the same site from McNary fore bay elevations were derived) with the pre-dam data extracted from Earth-Info databases.

Each of these data sets was transformed into model input format and plotted (Figures C.2 and C.3).

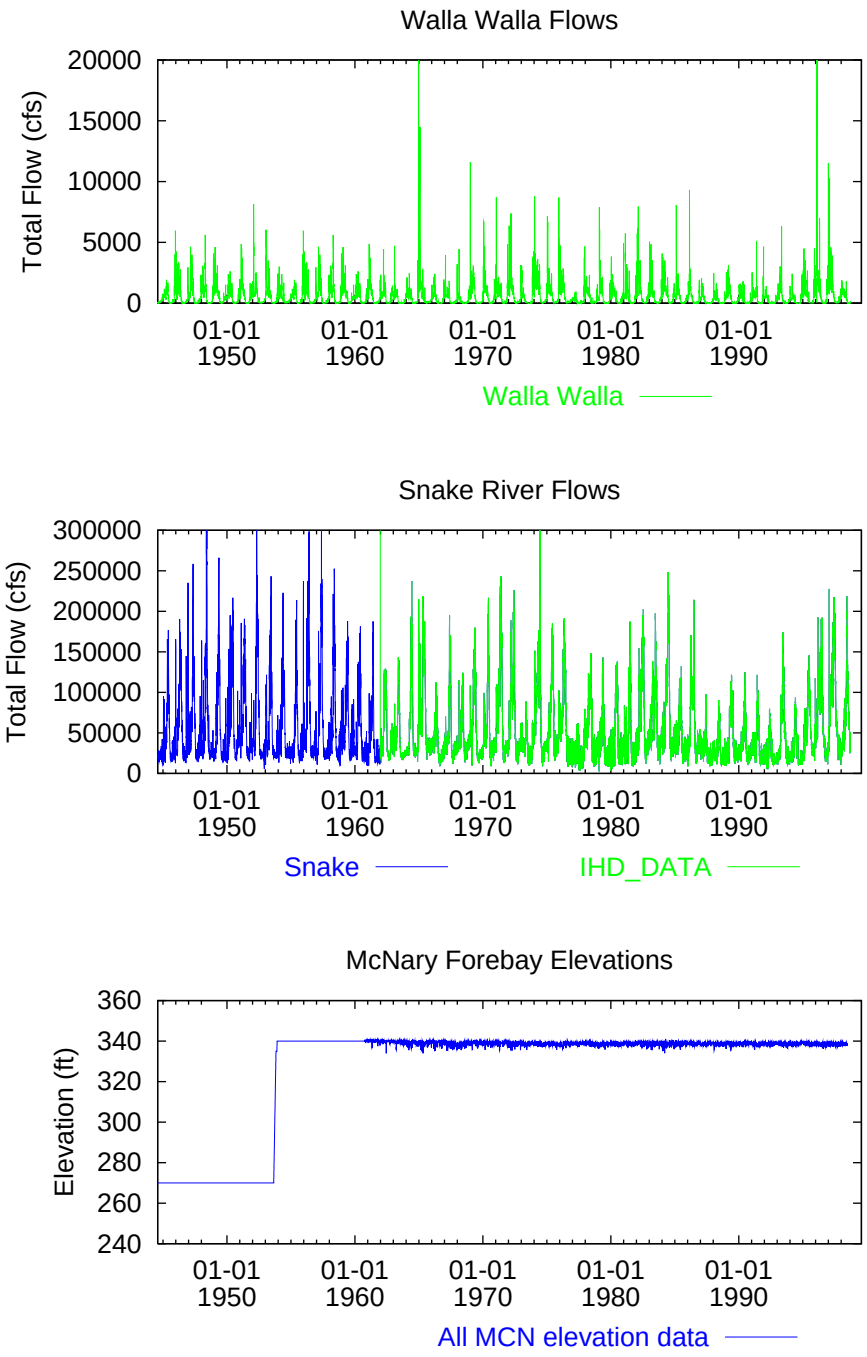

Figure C.2. Data for Snake and Walla Walla River Flows and the McNary Forebay Elevations

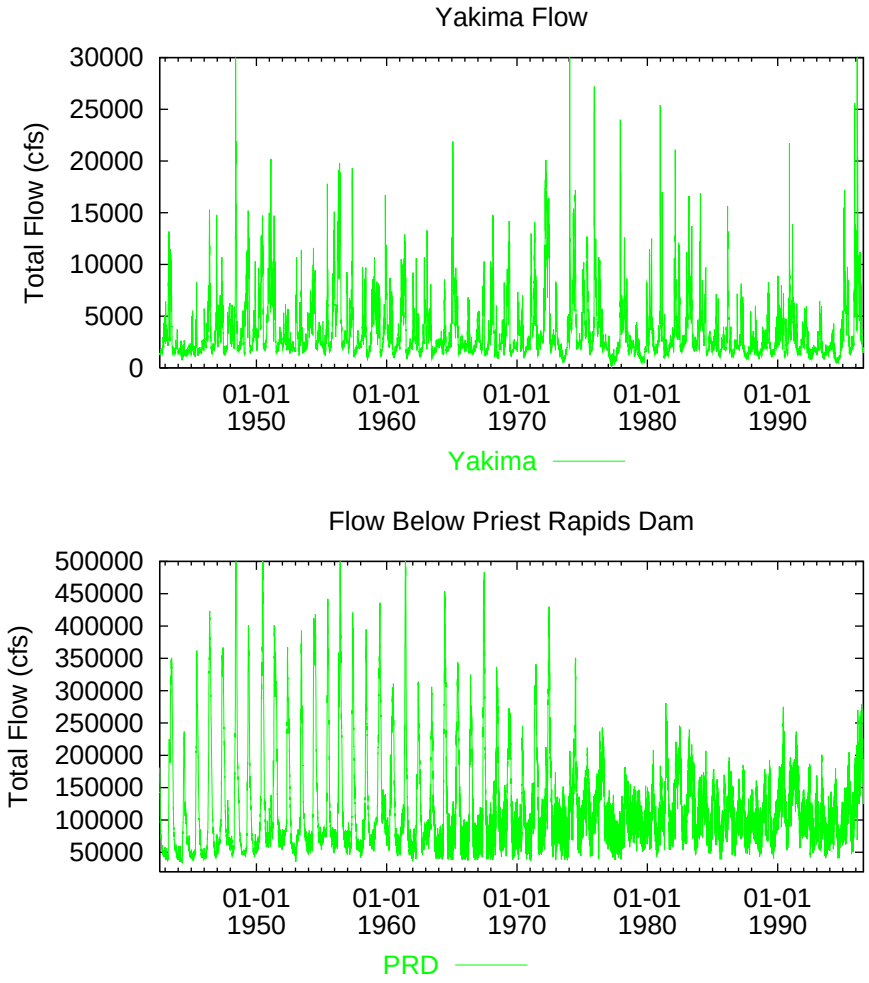

Figure C.3. Data for Columbia and Yakima River Flows

C.1.2 Validation of Results

The model results were checked by verifying that the simulated flows and water-surface elevations matched the respective input data (Figure C.4) and by comparing the simulated daily water-surface elevations with real hourly data for locations within the Hanford Reach (Figure C.5). We also verified that the bi-monthly averages corresponded with the daily data from which they were calculated (Figure C.6). Lastly, we included pre- and post-dam longitudinal profiles (Figure C.7) as well as numerous plots of hourly model output superimposed on hourly data (Figures C.8, C.9, C.10).

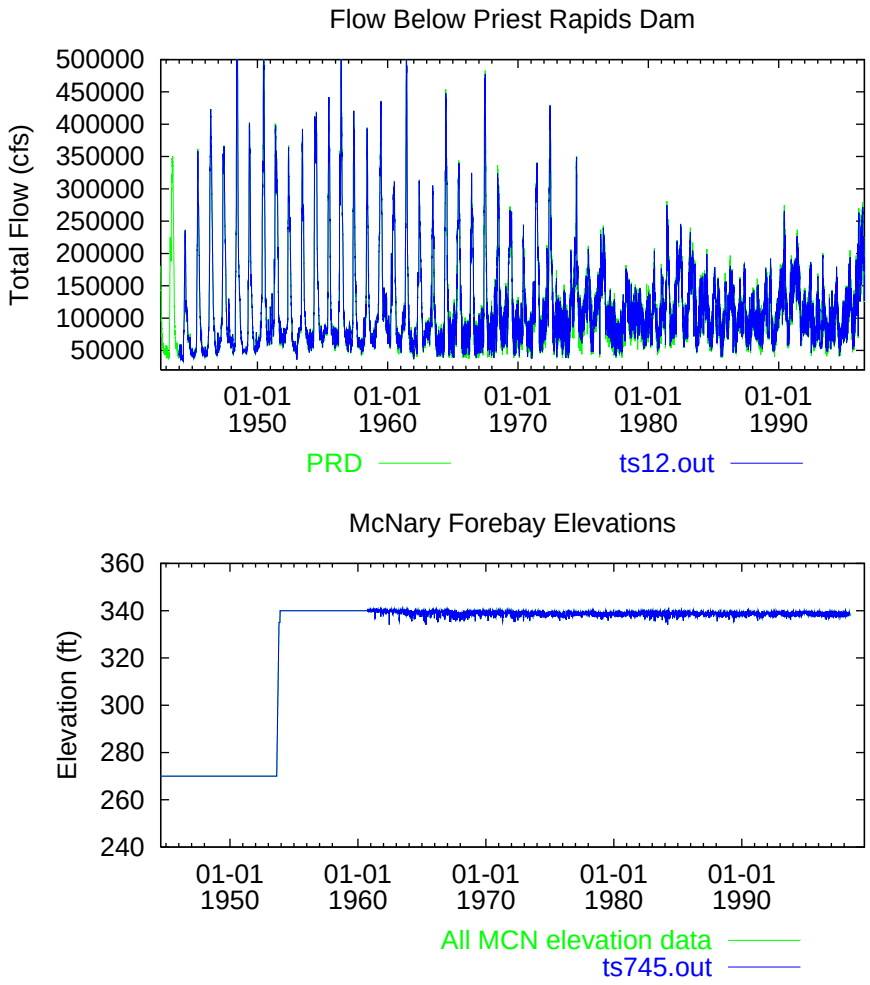

Figure C.4. Data for Flow Below Priest Rapids Dam and McNary Fore Bay Elevations Overlain with Model Output

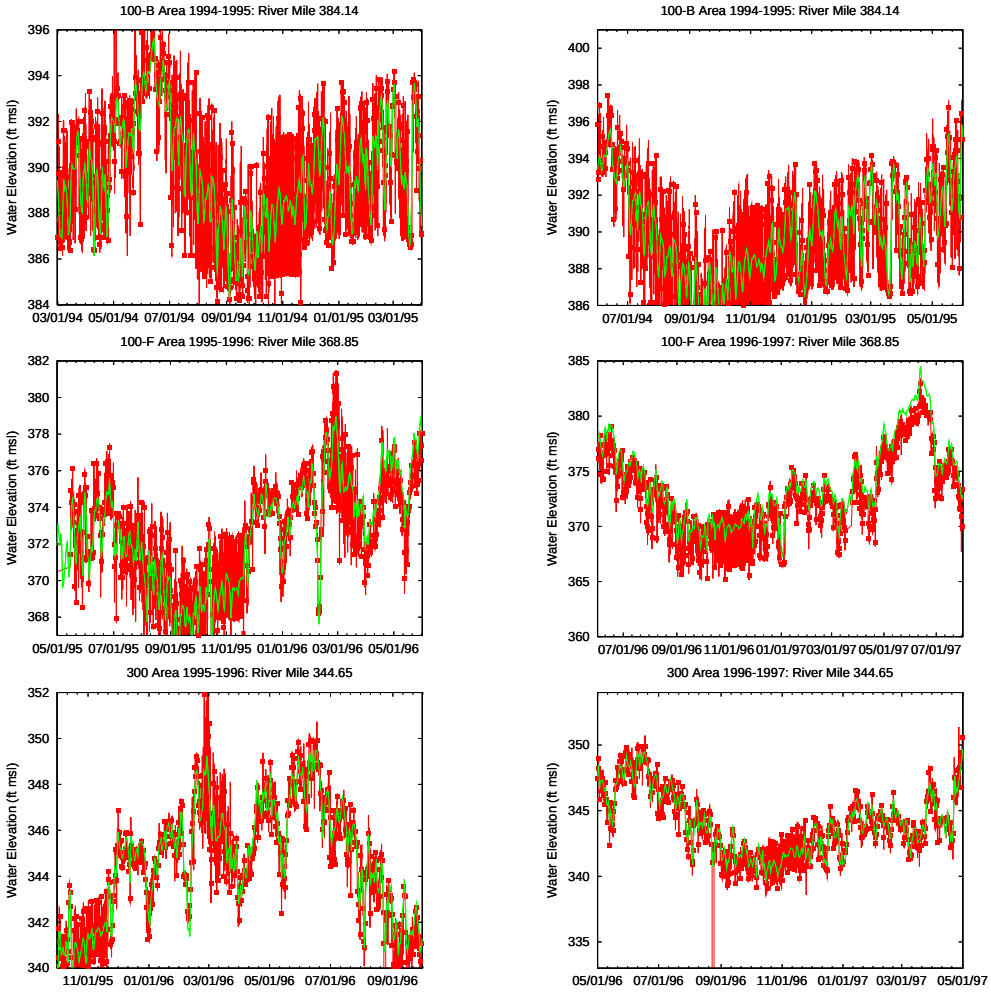

Figure C.5. Comparison of Simulated (Green) and Measured (Red) Water-Surface Elevations (ft) for the Hanford Reach

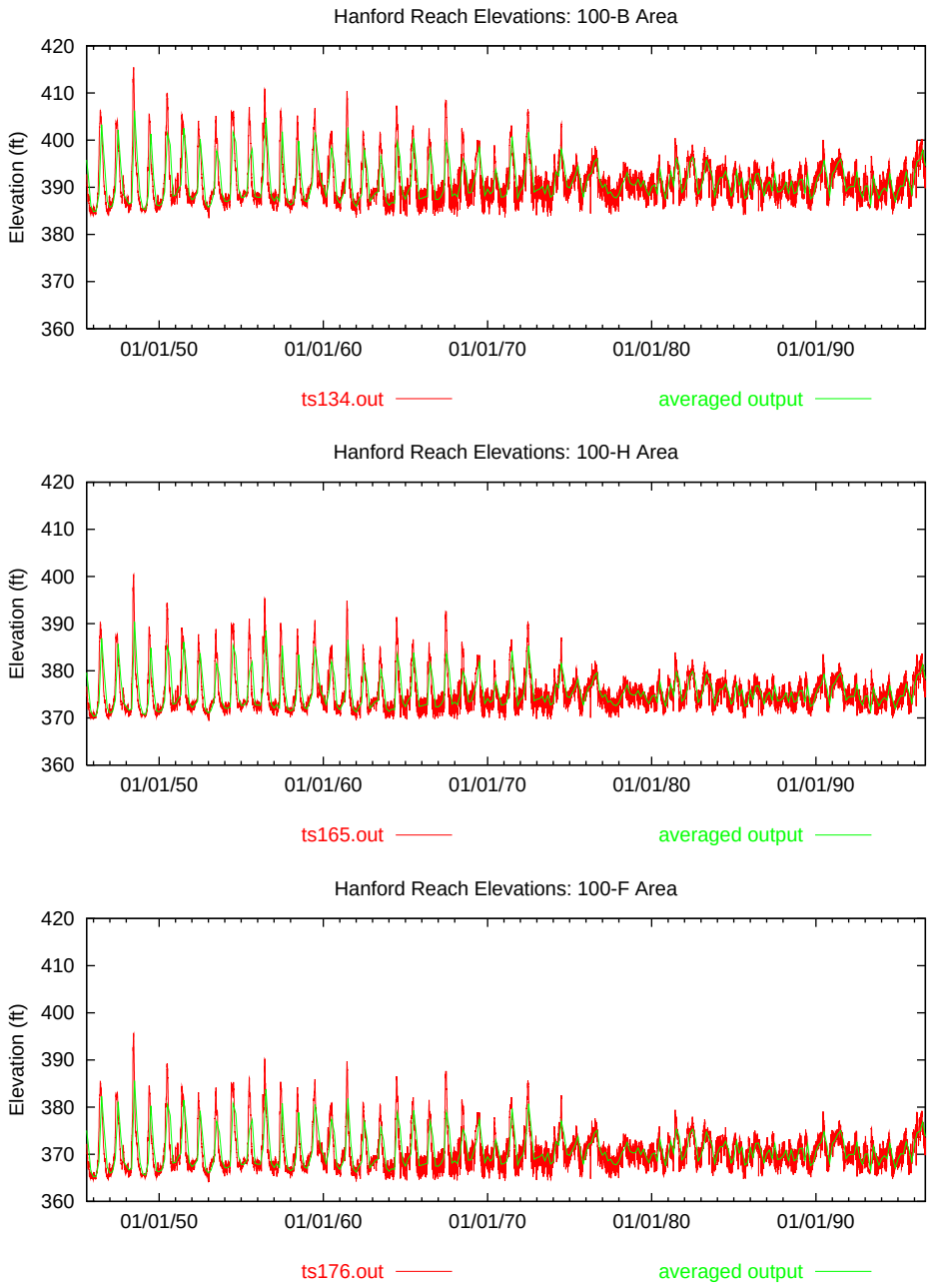

Figure C.6. Plots of Daily and Bi-Monthly Average Model Output for Selected Sites

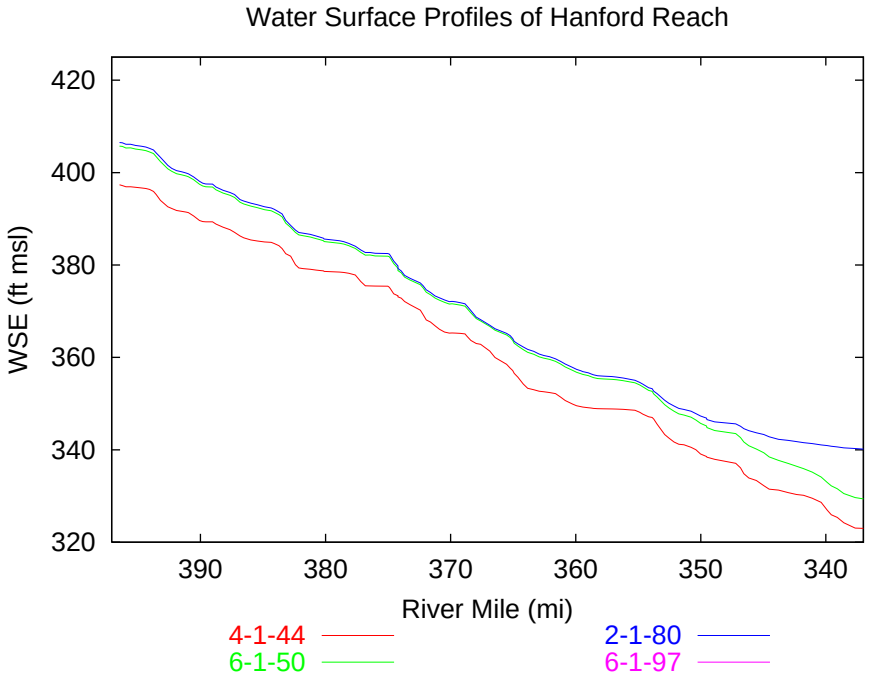

Figure C.7. Longitudinal Profiles of Averaged Data for Selected Dates

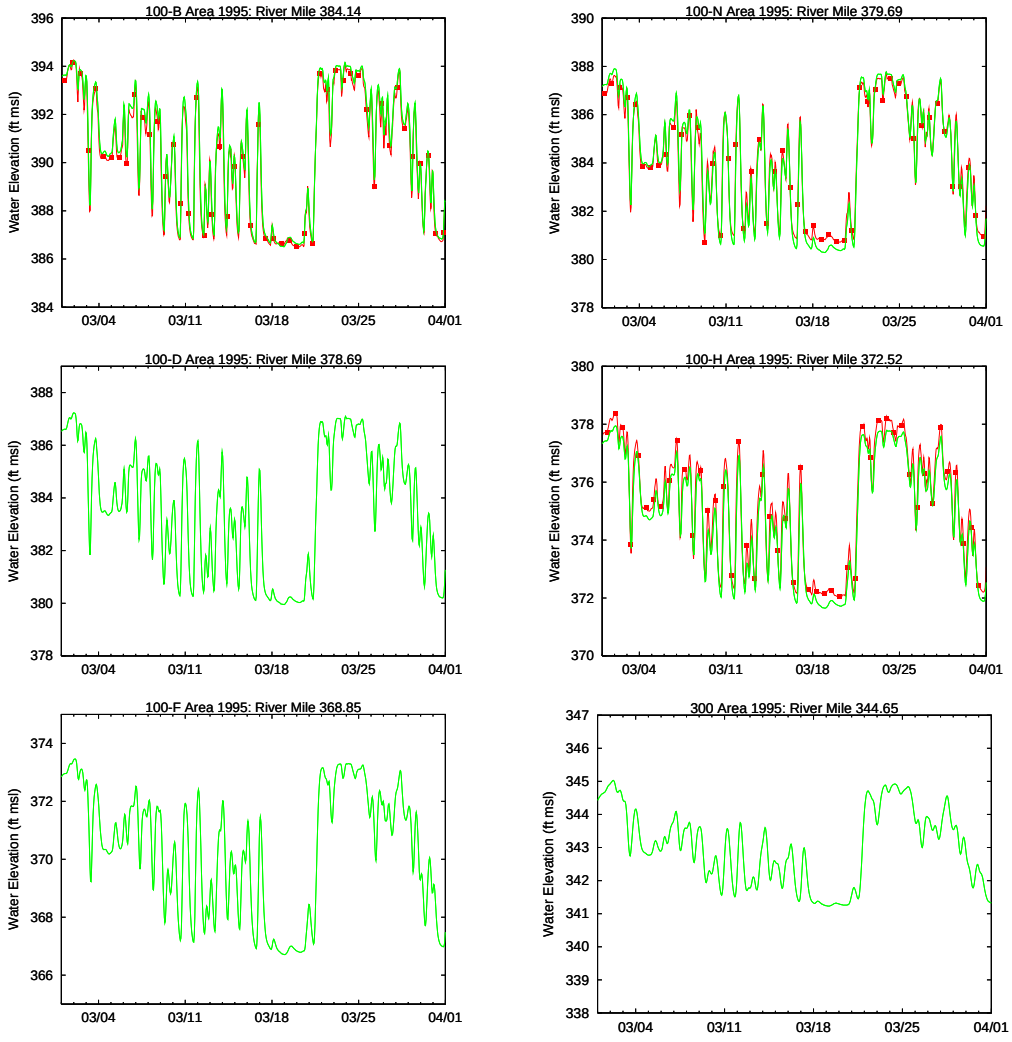

Figure C.8. Comparison of Simulated and Measured (symbols) Water-Surface Elevations for March 1995

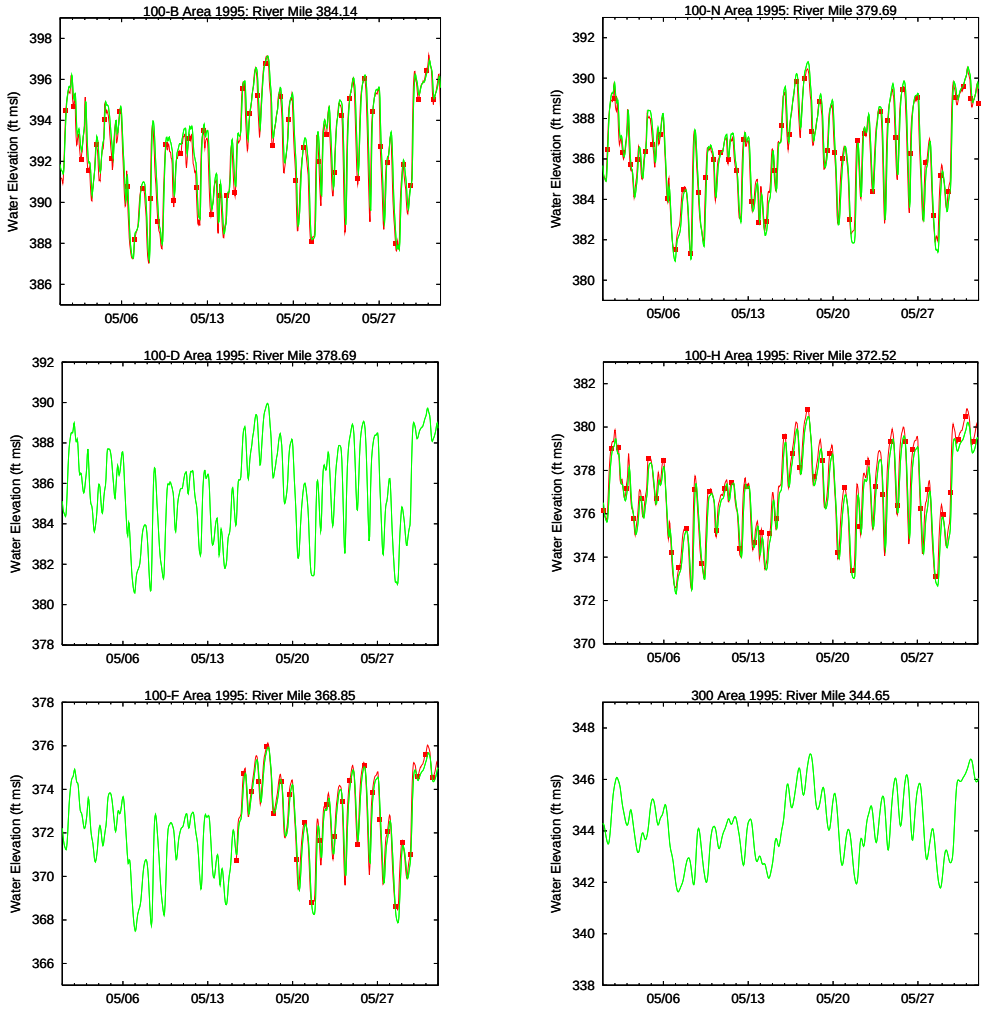

Figure C.9. Comparison of Simulated and Measured (symbols) Water-Surface Elevations for May 1995

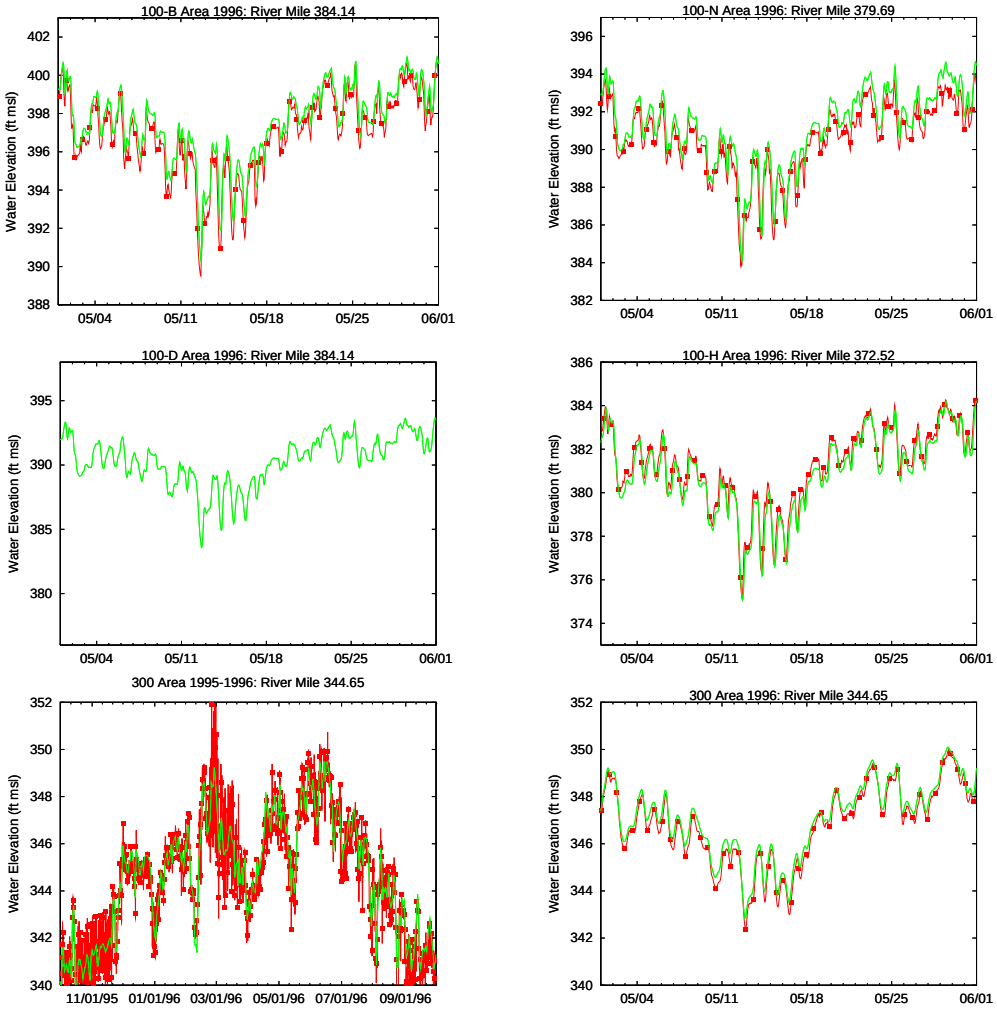

Figure C.10. Comparison of Simulated and Measured (symbols) Water-Surface Elevations for May 1996

C.2 References

Richmond MC, WA Perkins, and Y Chien. 2000a. *Numerical Model Analysis of System-Wide Dissolved Gas Abatement Alternatives*. Final report prepared for U.S. Army Corps of Engineers, Walla Walla District, by Battelle Pacific Northwest Division, Richland, Washington.

Richmond MC, WA Perkins, and Y Chien. 2000b. *Numerical Model Analysis of System-Wide Dissolved Gas Abatement Alternatives, Appendix B, Model Configuration and Optimization*. Final report prepared for U.S. Army Corps of Engineers, Walla Walla District, by Battelle Pacific Northwest Division, Richland, Washington.

Richmond MC, WA Perkins, and Y Chien. 2000c. *Numerical Model Analysis of System-Wide Dissolved Gas Abatement Alternatives, Appendix F, Verification of MASS1 for Lower Columbia/Snake Temperature and Total Dissolved Gas Simulation*. Final report prepared for U.S. Army Corps of Engineers, Walla Walla District, by Battelle Pacific Northwest Division, Richland, Washington.

Appendix D

Selected Plots of Water Table Elevations and Head Residuals from Simulation of Hanford Wastewater Discharges (1943–1996) Using the Prior Model

Appendix D

Selected Plots of Water Table Elevations and Head Residuals from Simulation of Hanford Wastewater Discharges (1943–1996) Using the Prior Model

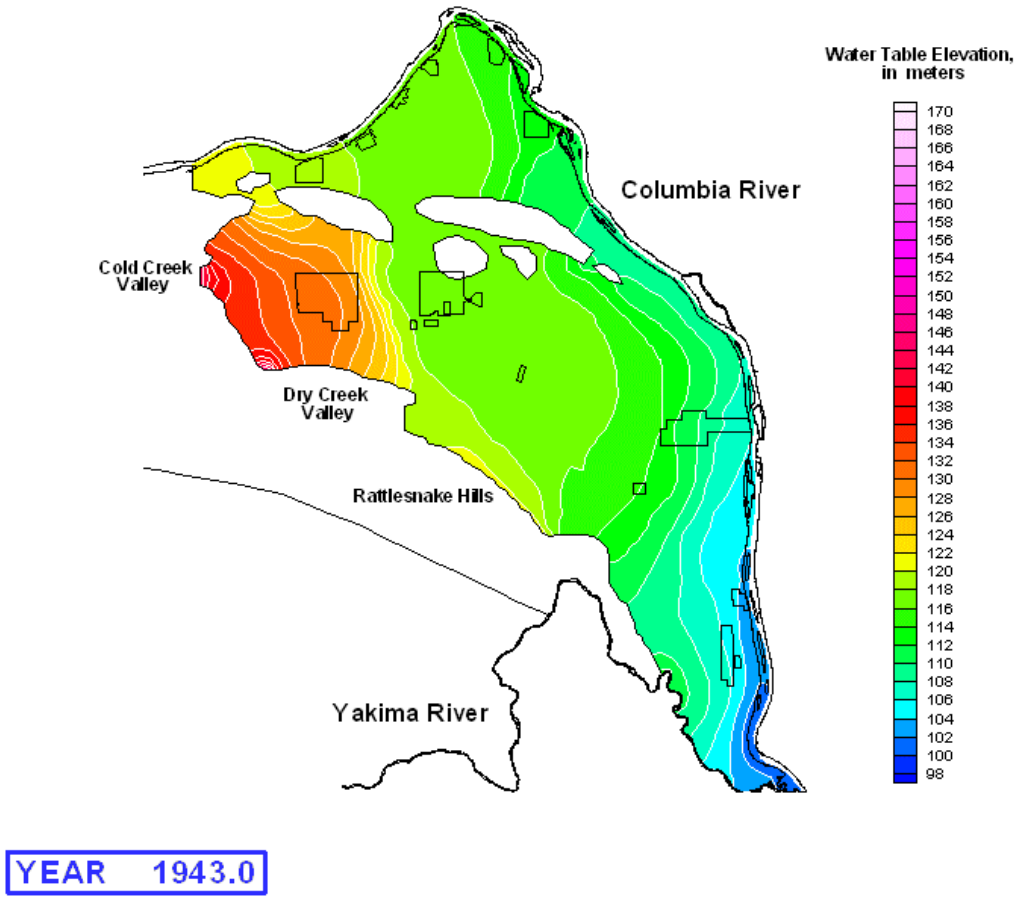

Figure D.1a. Simulated Water-Table Elevations for 1943

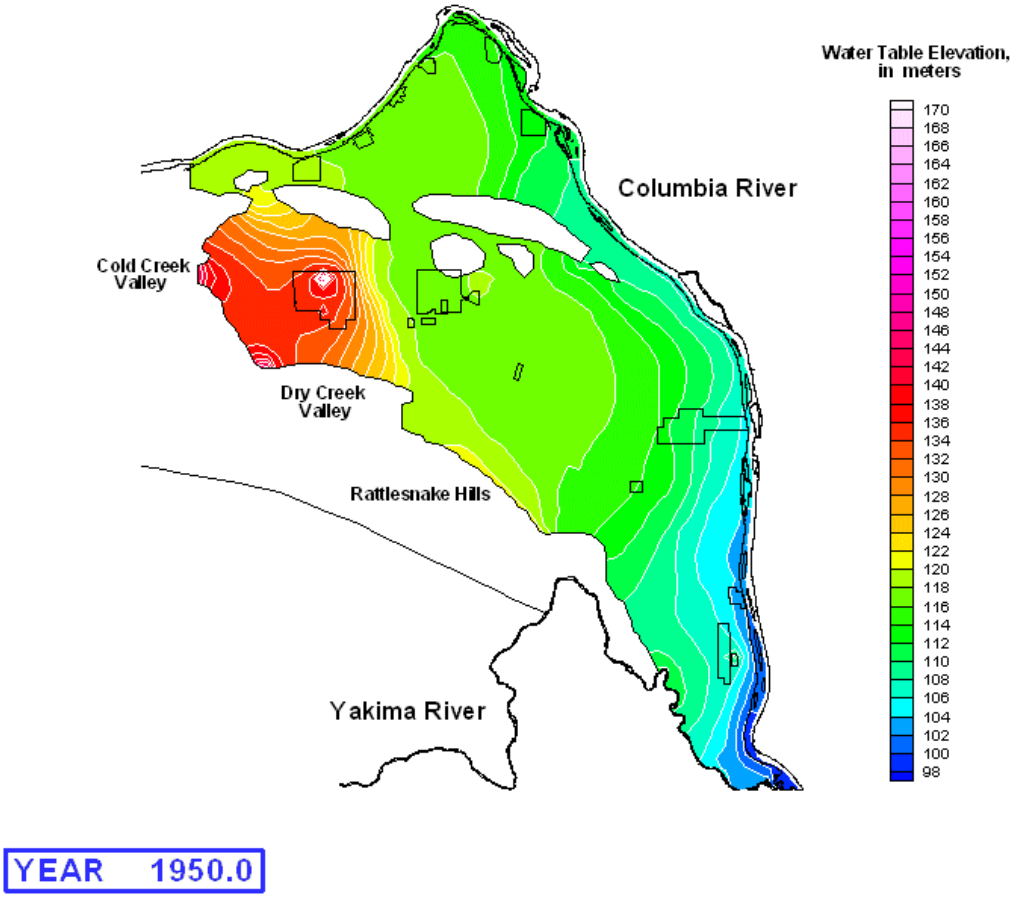

Figure D.1b. Simulated Water-Table Elevations for 1950

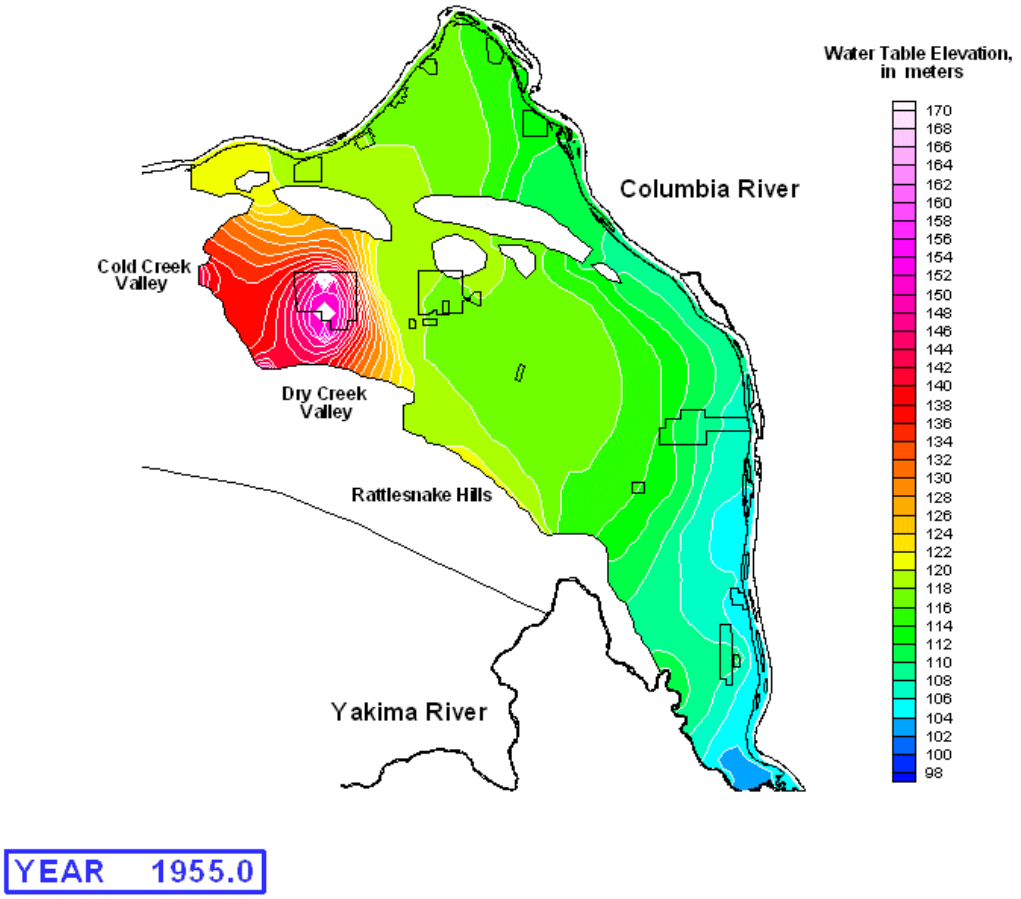

Figure D.1c. Simulated Water-Table Elevations for 1955

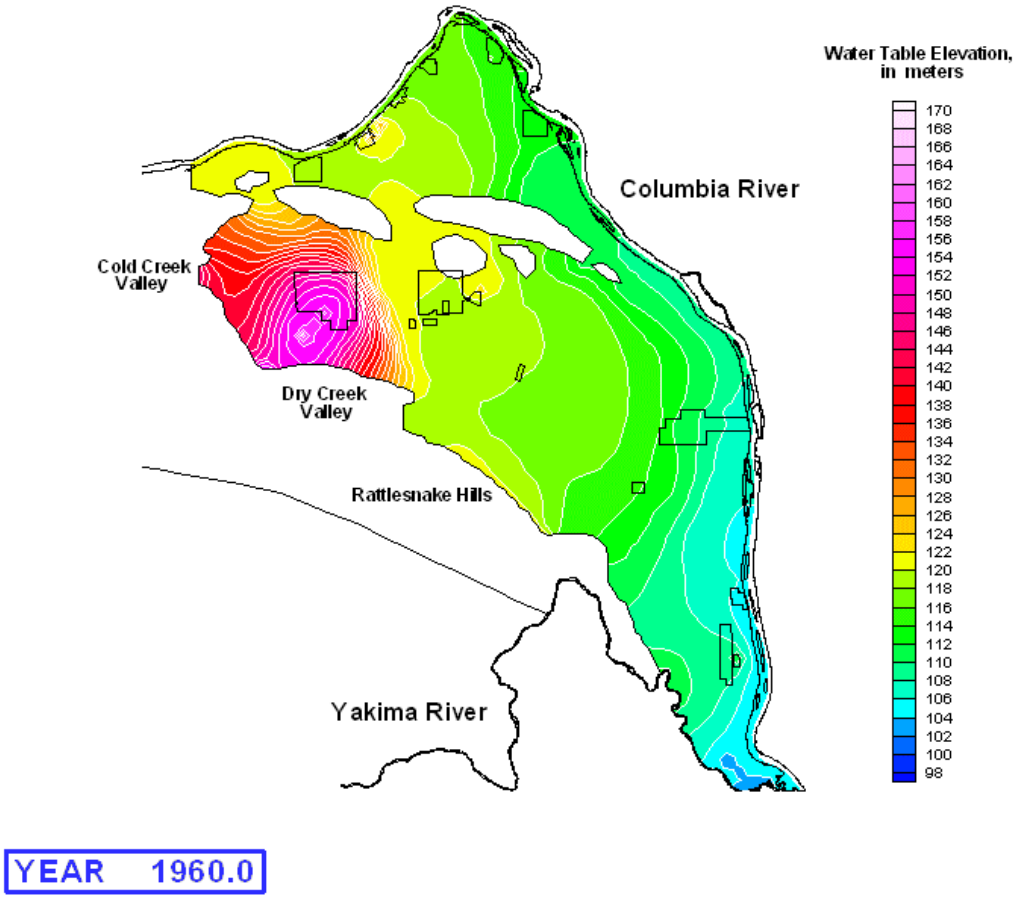

Figure D.1d. Simulated Water-Table Elevations for 1960

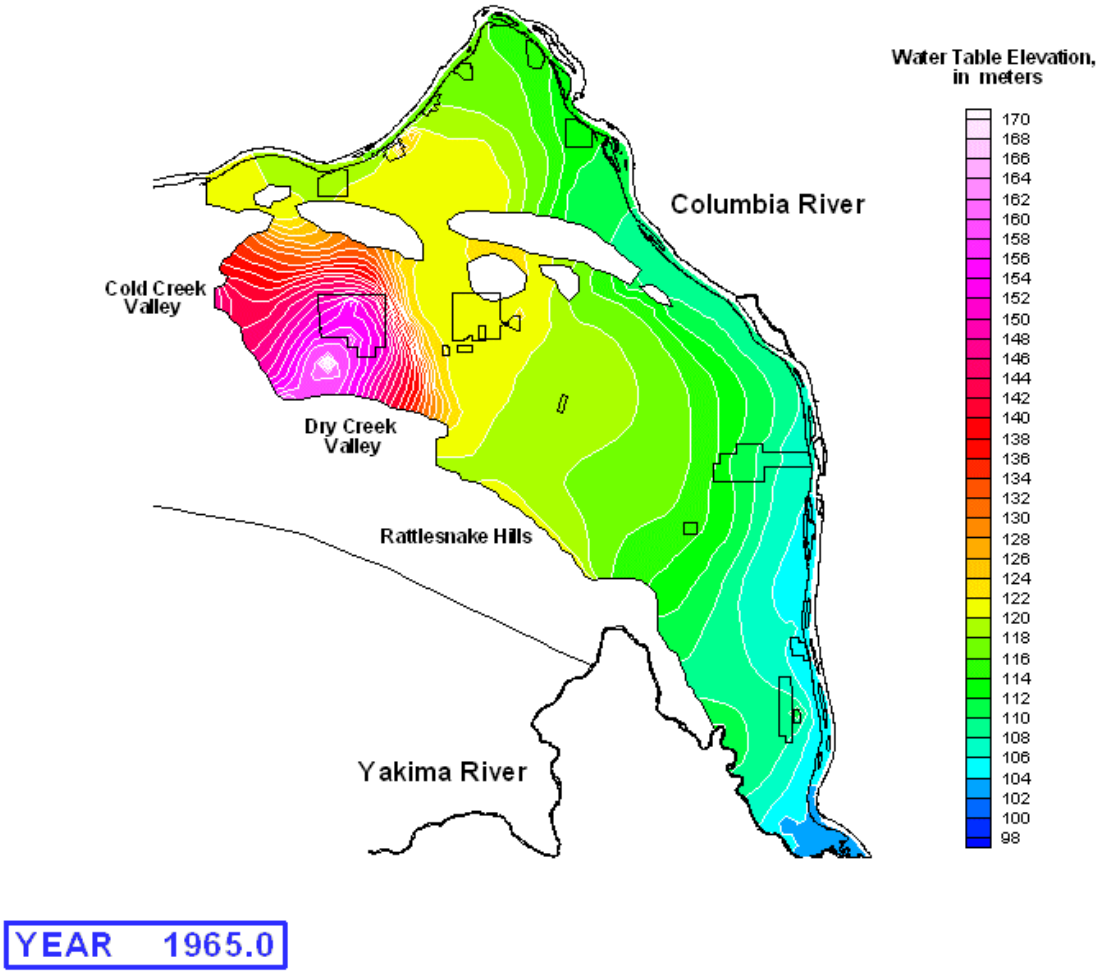

Figure D.1e. Simulated Water-Table Elevations for 1965

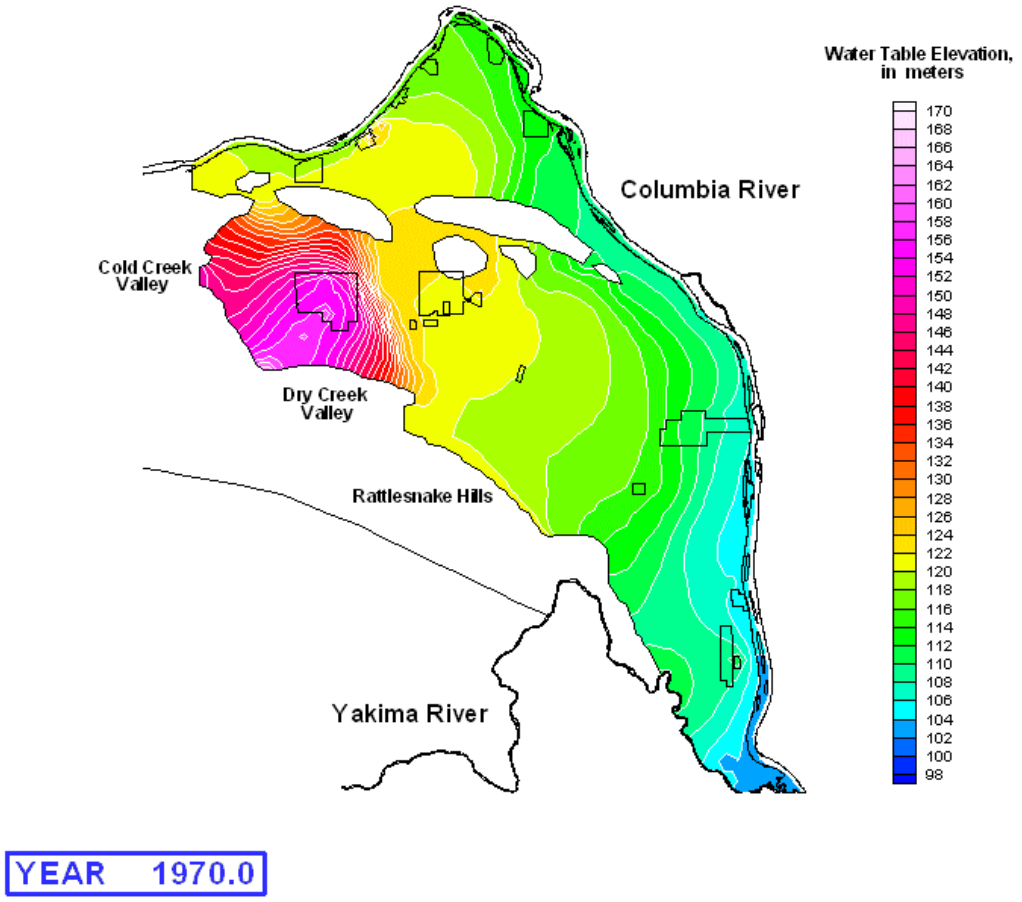

Figure D.1f. Simulated Water-Table Elevations for 1970

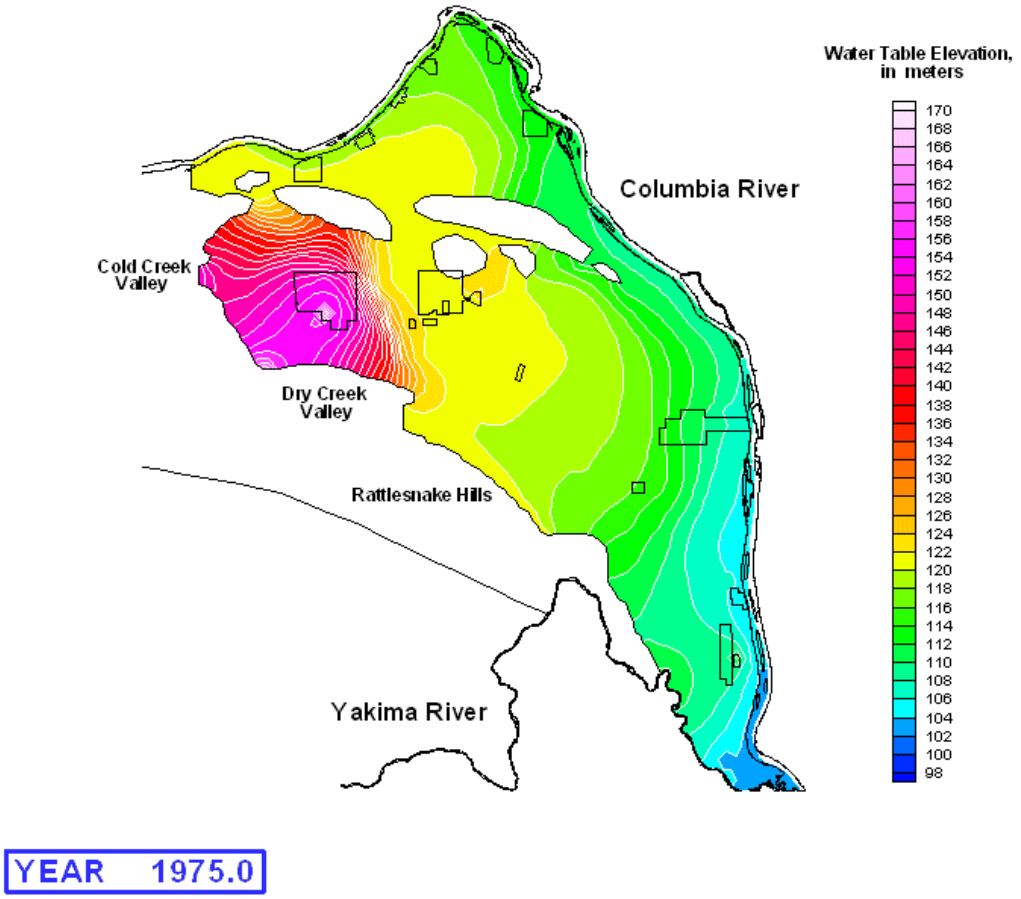

Figure D.1g. Simulated Water-Table Elevations for 1975

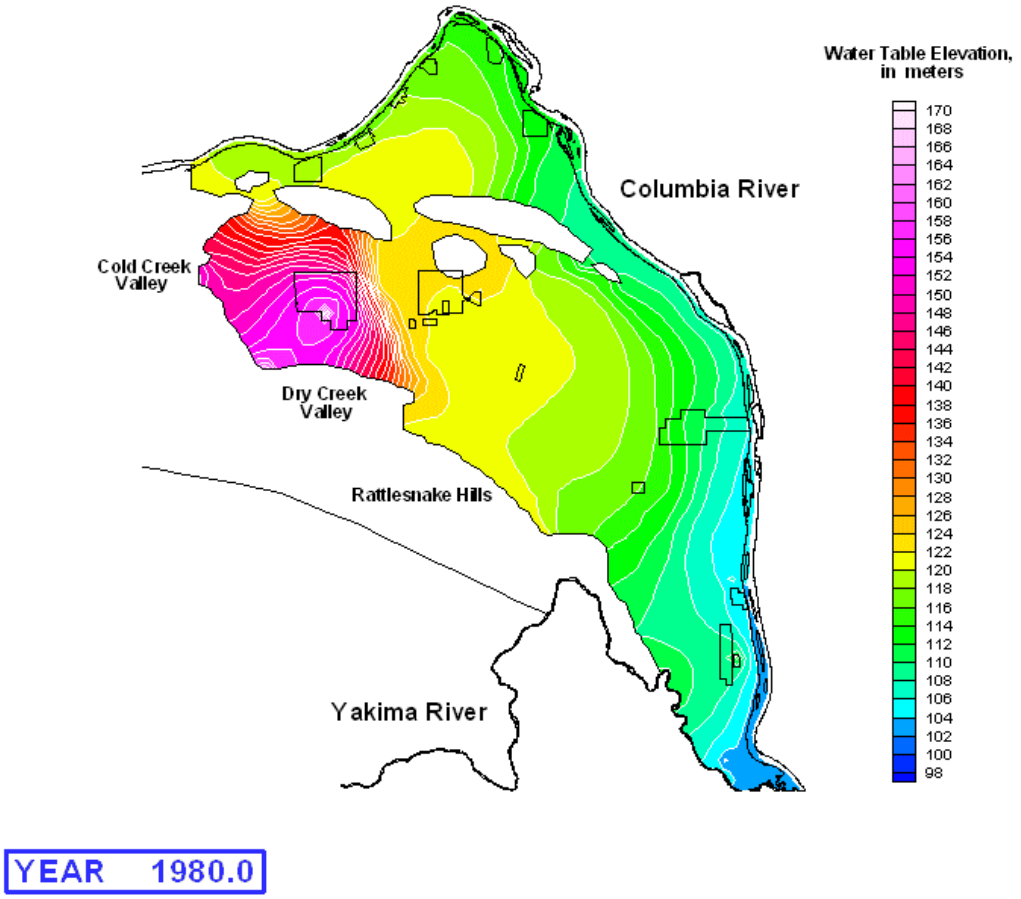

Figure D.1h. Simulated Water-Table Elevations for 1980

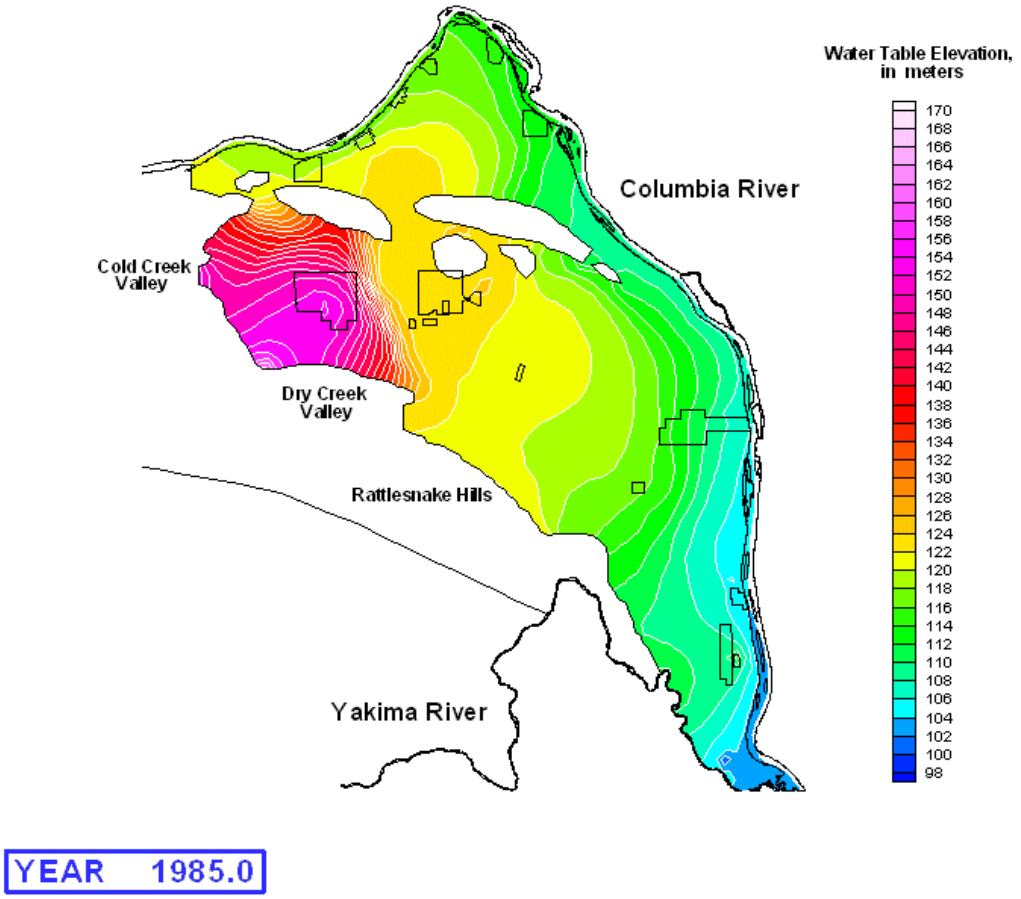

Figure D.1i. Simulated Water-Table Elevations for 1985

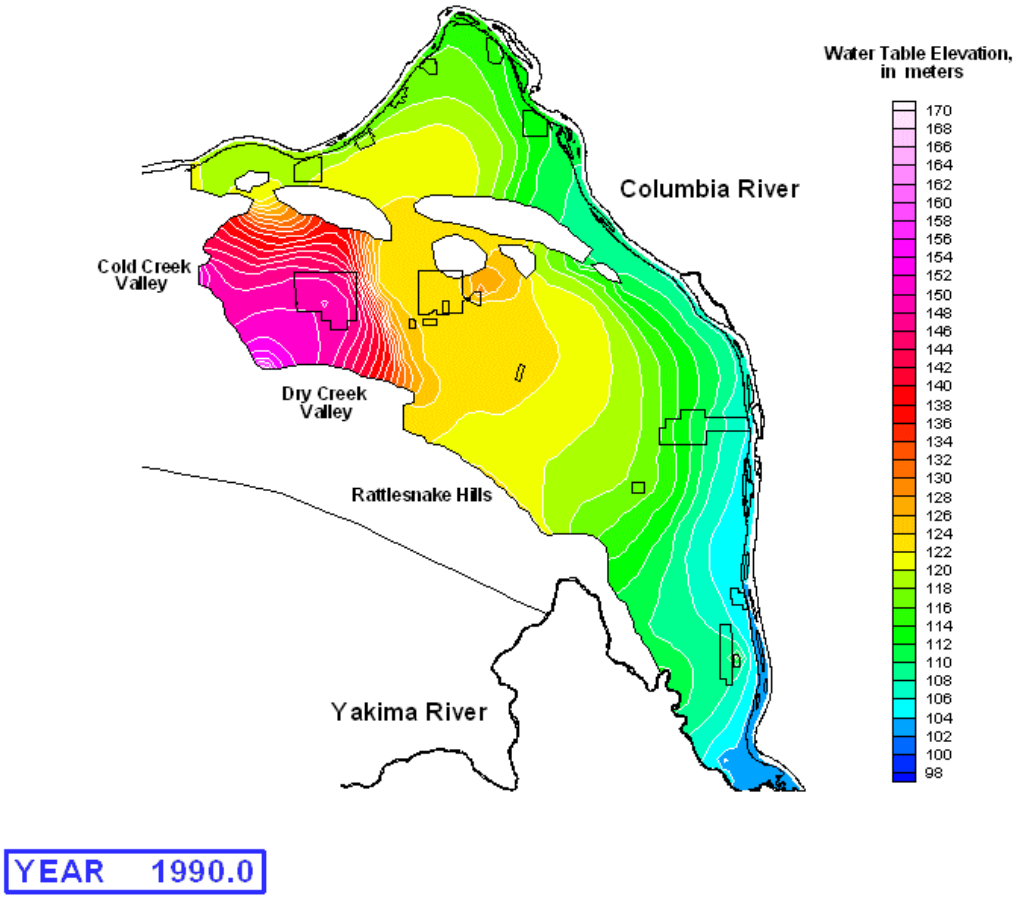

Figure D.1j. Simulated Water-Table Elevations for 1990

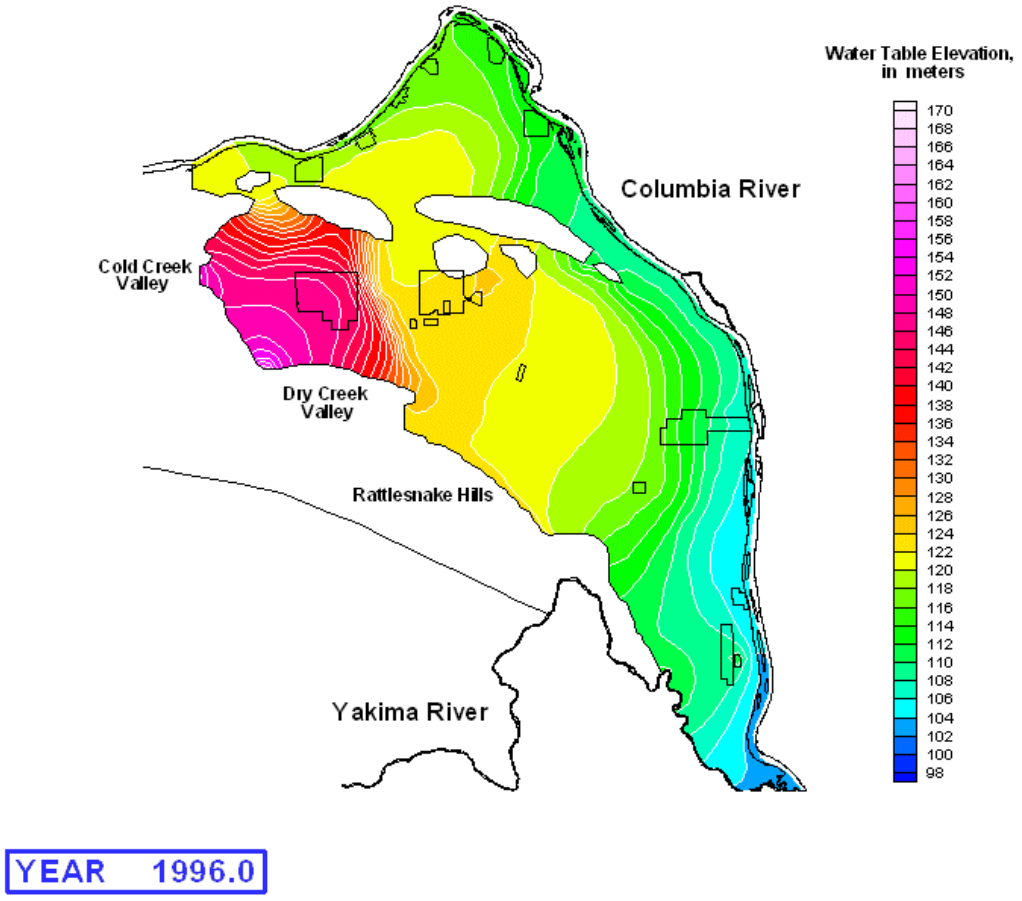

Figure D.1k. Simulated Water-Table Elevations for 1996

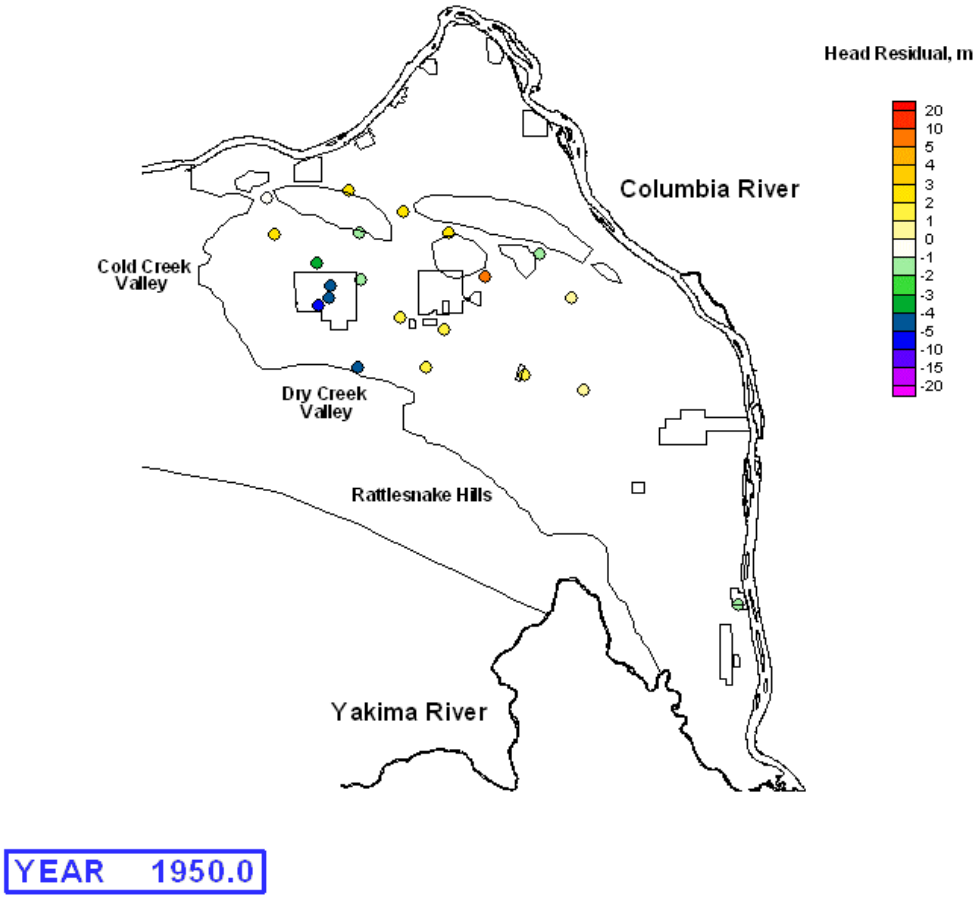

Figure D.2a. Hydraulic Head Residuals (simulated - measured water levels) for 1950

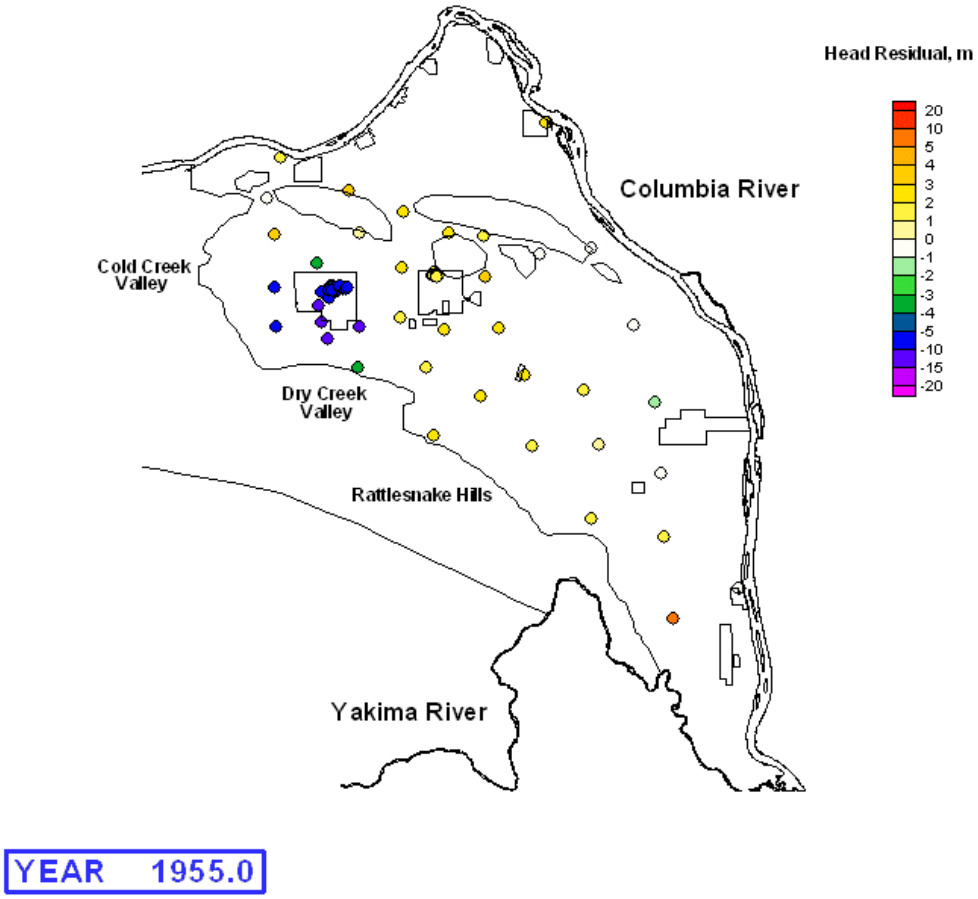

Figure D.2b. Hydraulic Head Residuals (simulated - measured water levels) for 1955

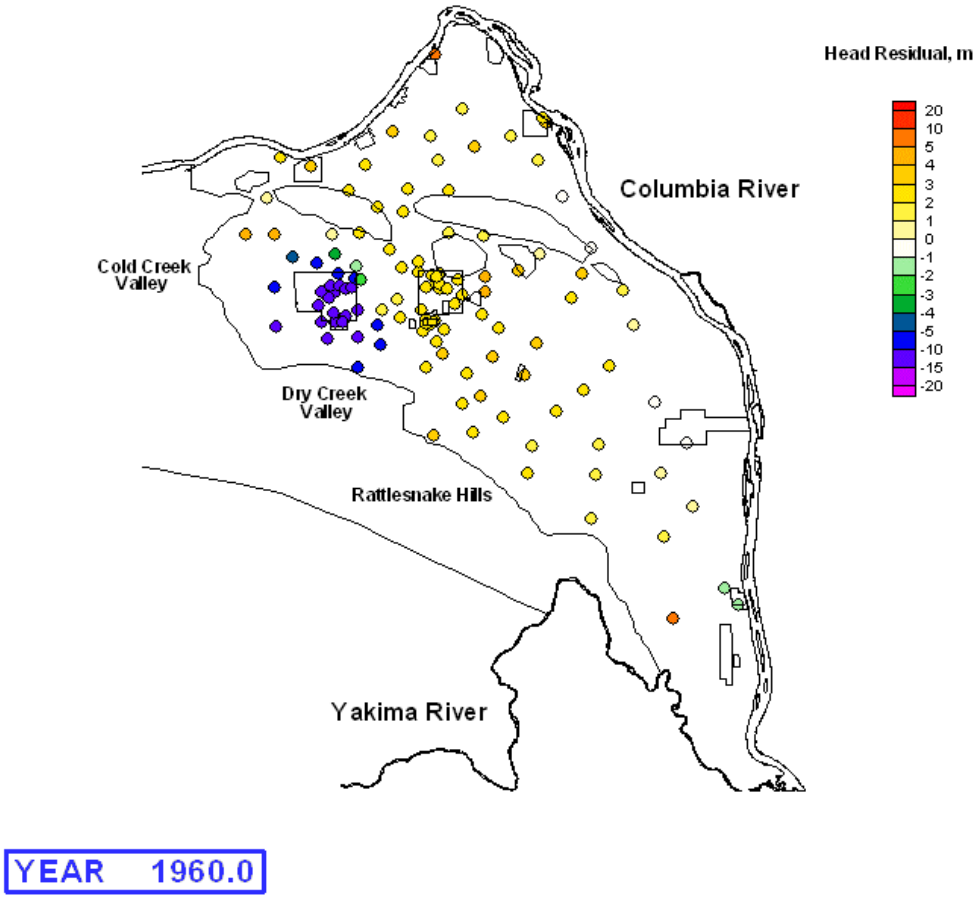

Figure D.2c. Hydraulic Head Residuals (simulated - measured water levels) for 1960

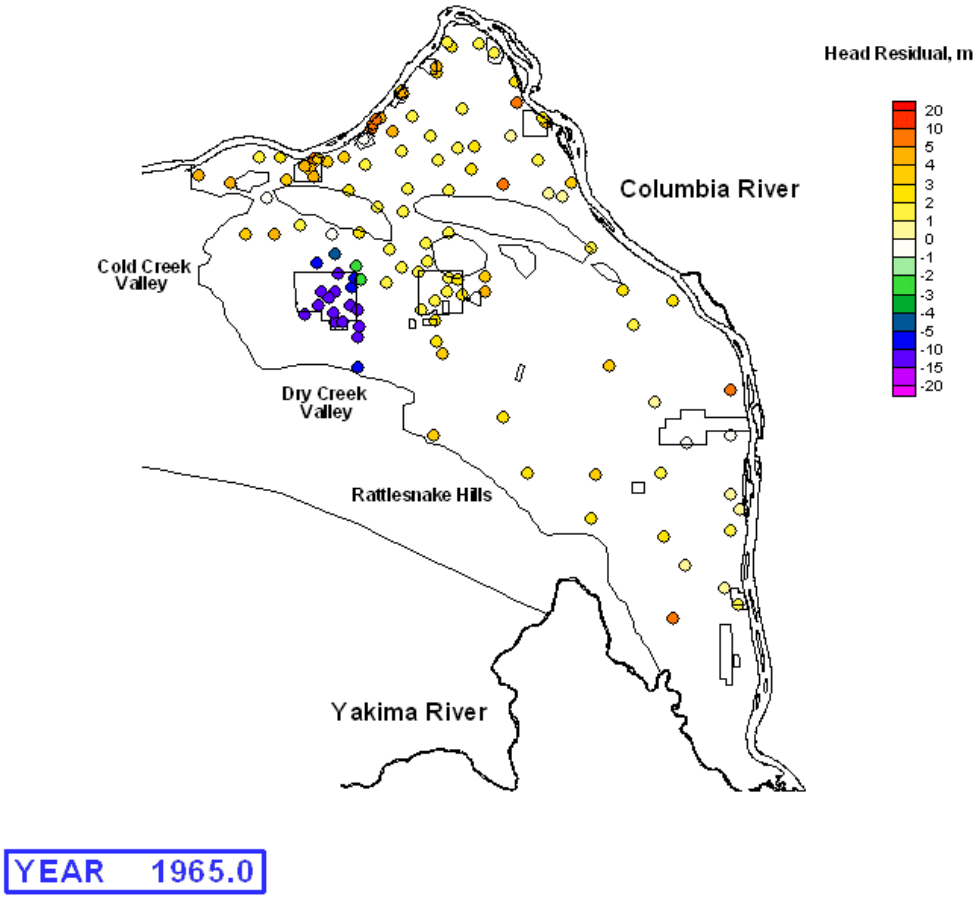

Figure D.2d. Hydraulic Head Residuals (simulated - measured water levels) for 1965

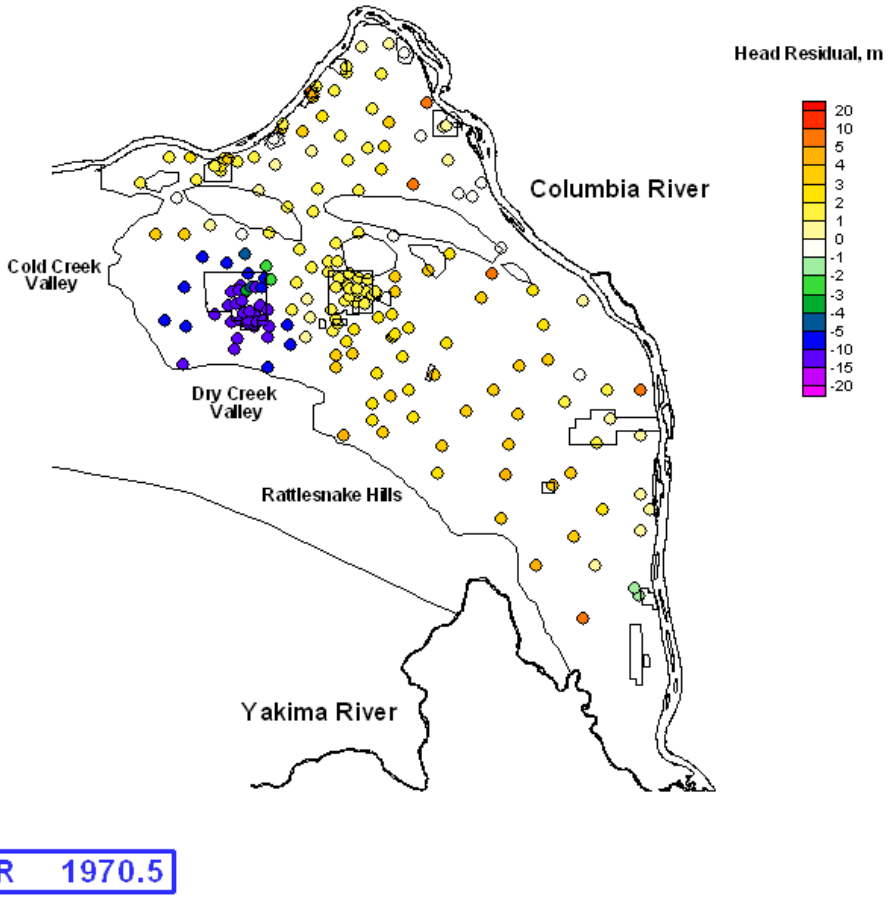

Figure D.2e. Hydraulic Head Residuals (simulated - measured water levels) for June 1970

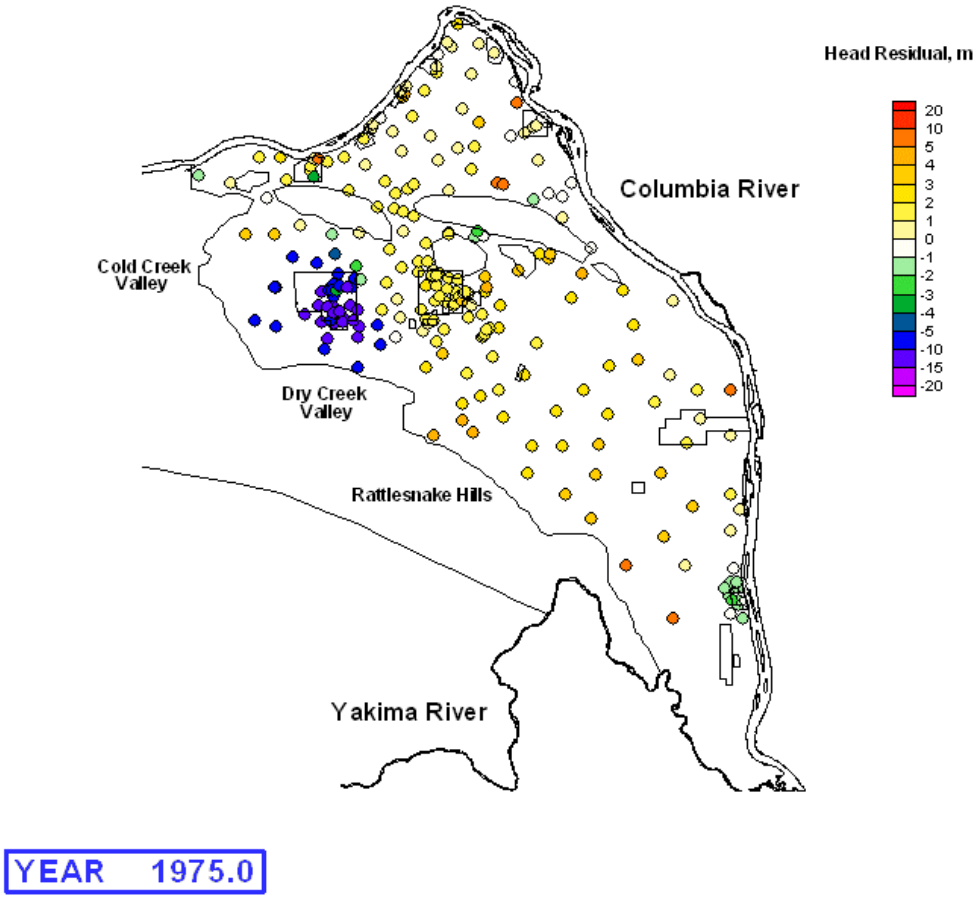

Figure D.2f. Hydraulic Head Residuals (simulated - measured water levels) for 1975

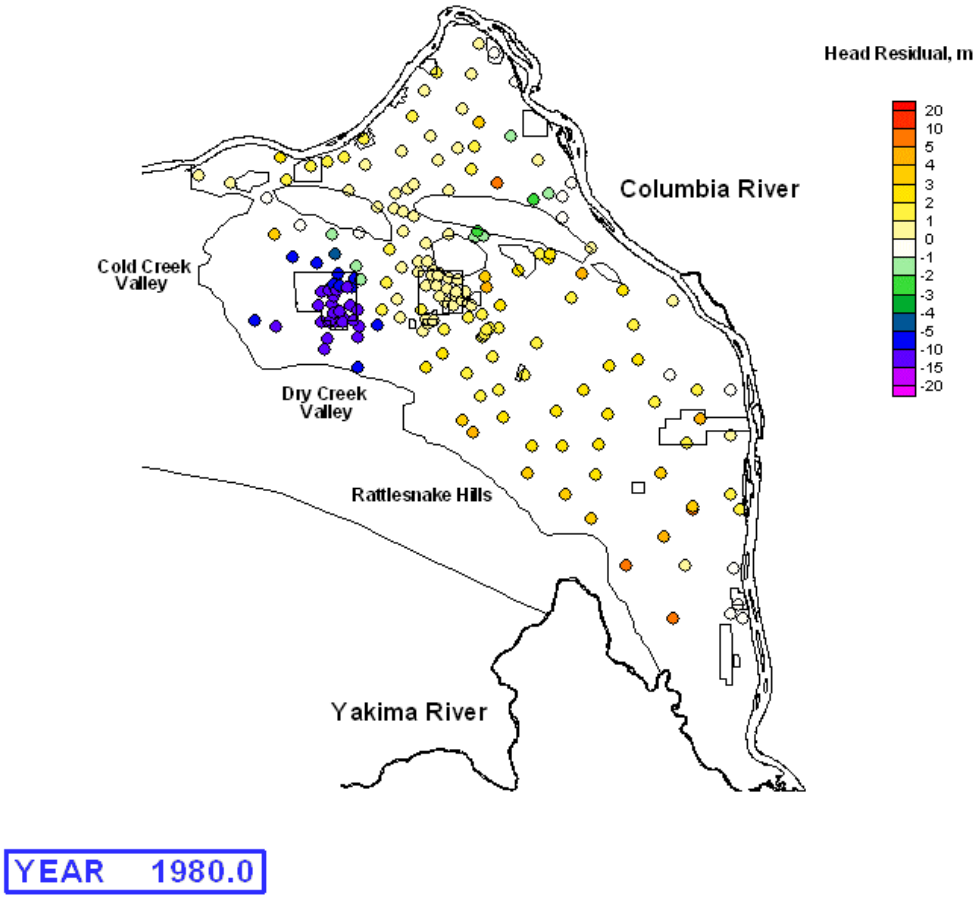

Figure D.2g. Hydraulic Head Residuals (simulated - measured water levels) for 1980

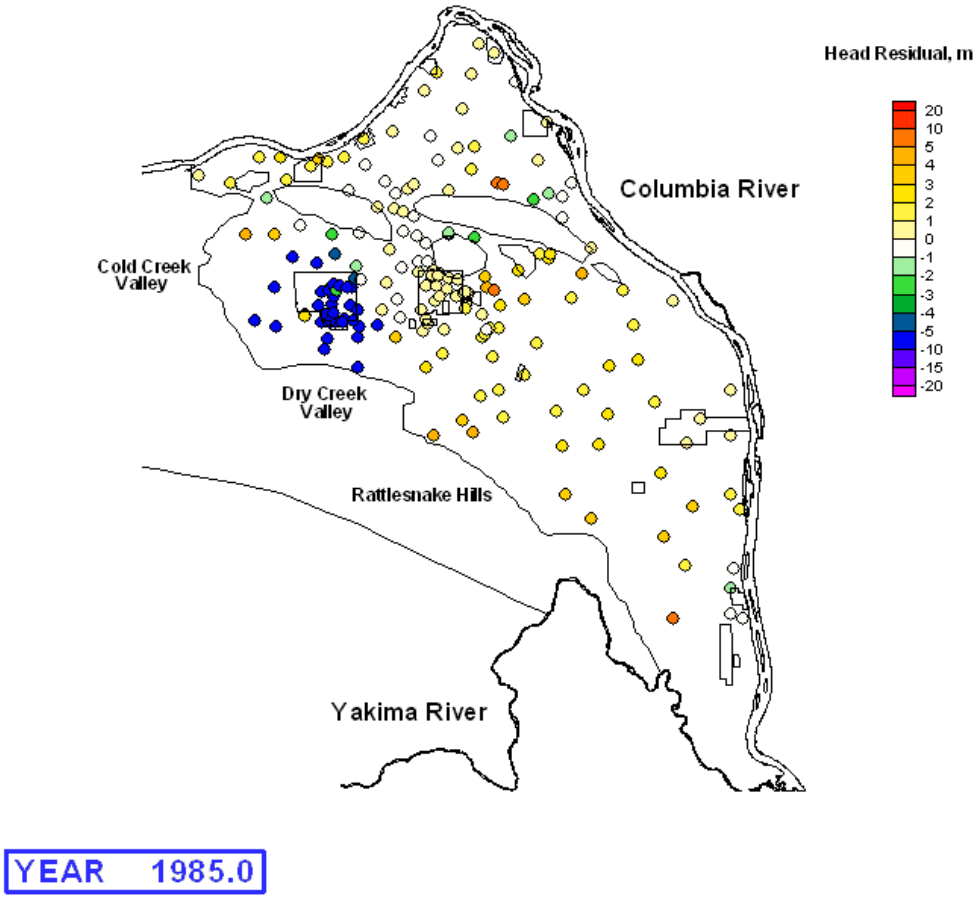

Figure D.2h. Hydraulic Head Residuals (simulated - measured water levels) for 1985

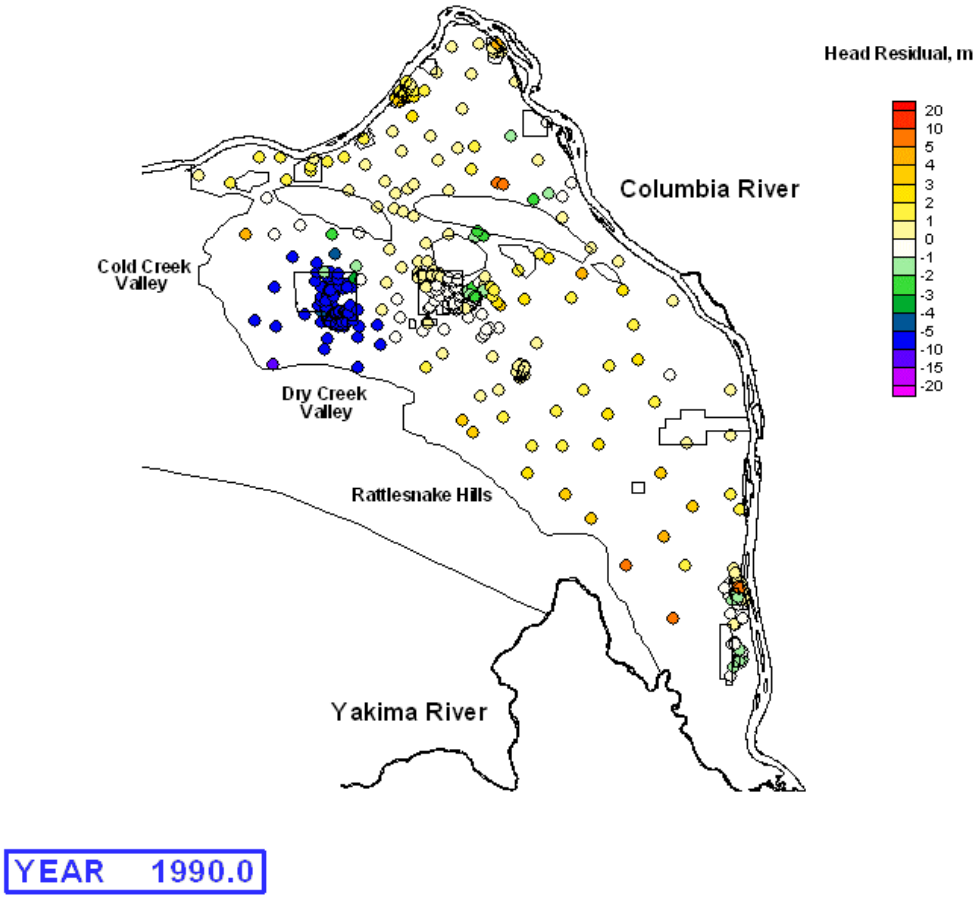

Figure D.2i. Hydraulic Head Residuals (simulated - measured water levels) for 1990

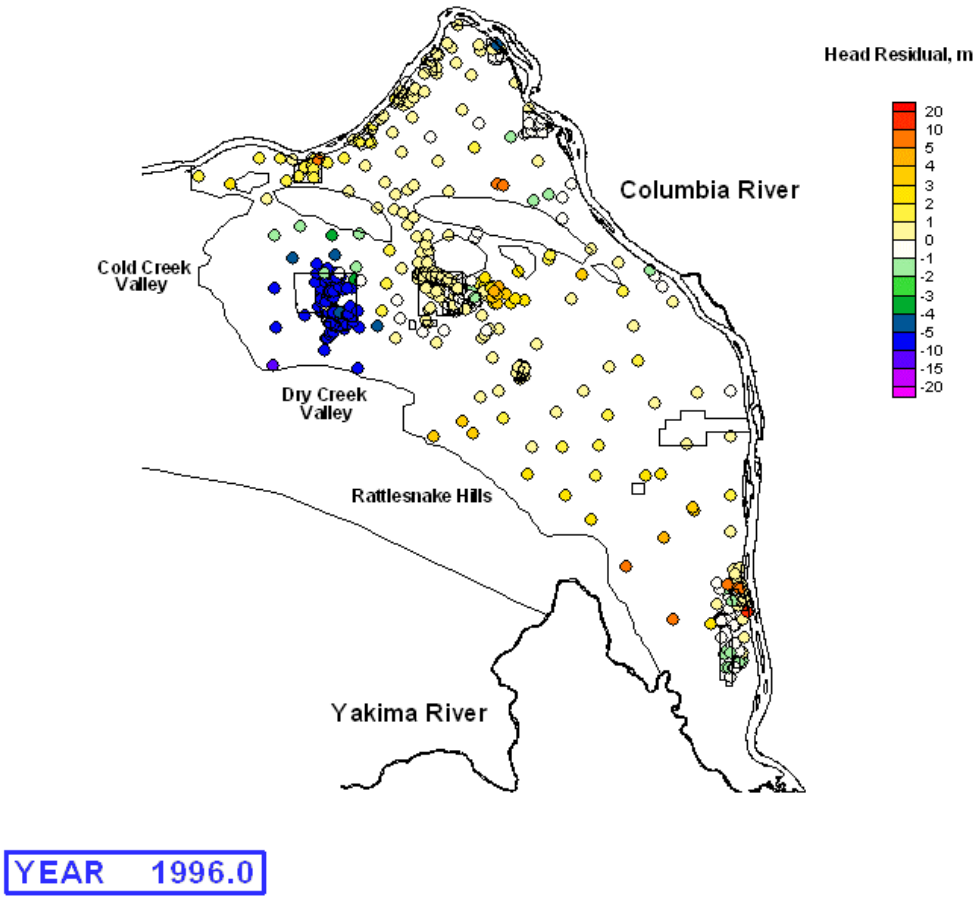

Figure D.2j. Hydraulic Head Residuals (simulated - measured water levels) for 1996

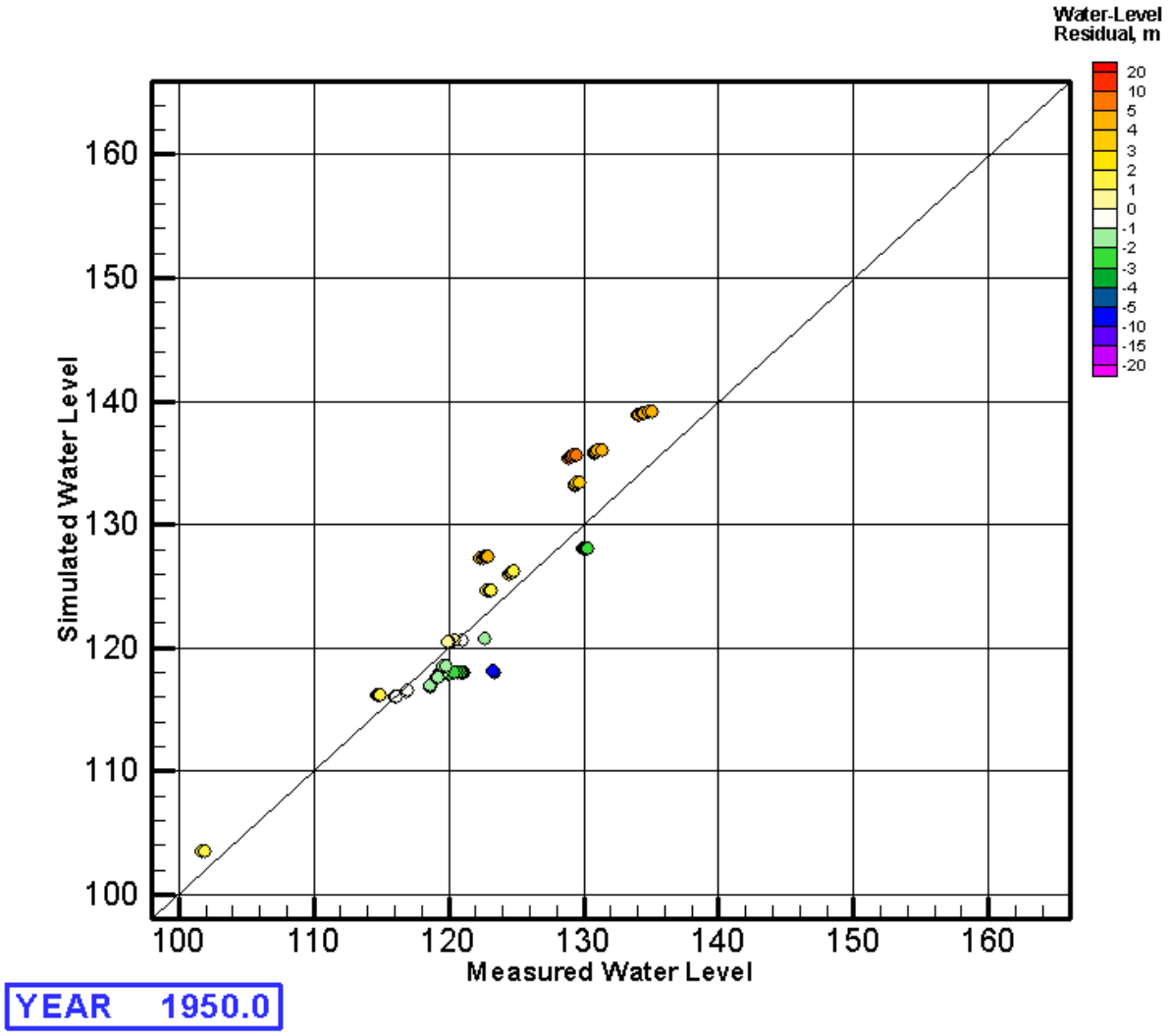

Figure D.3a. Comparison of Simulated versus Measured Water Levels for 1950

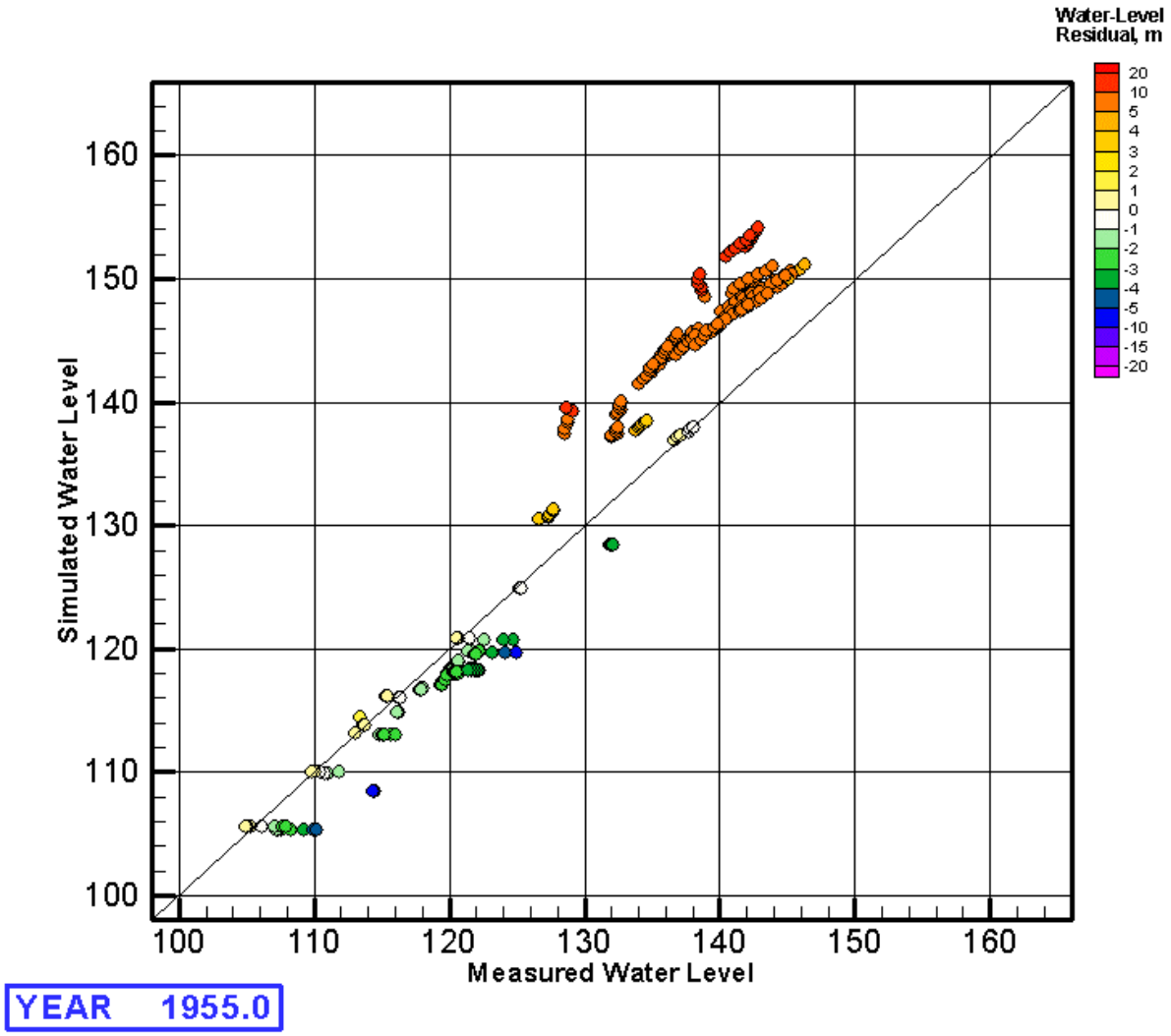

Figure D.3b. Comparison of Simulated versus Measured Water Levels for 1955

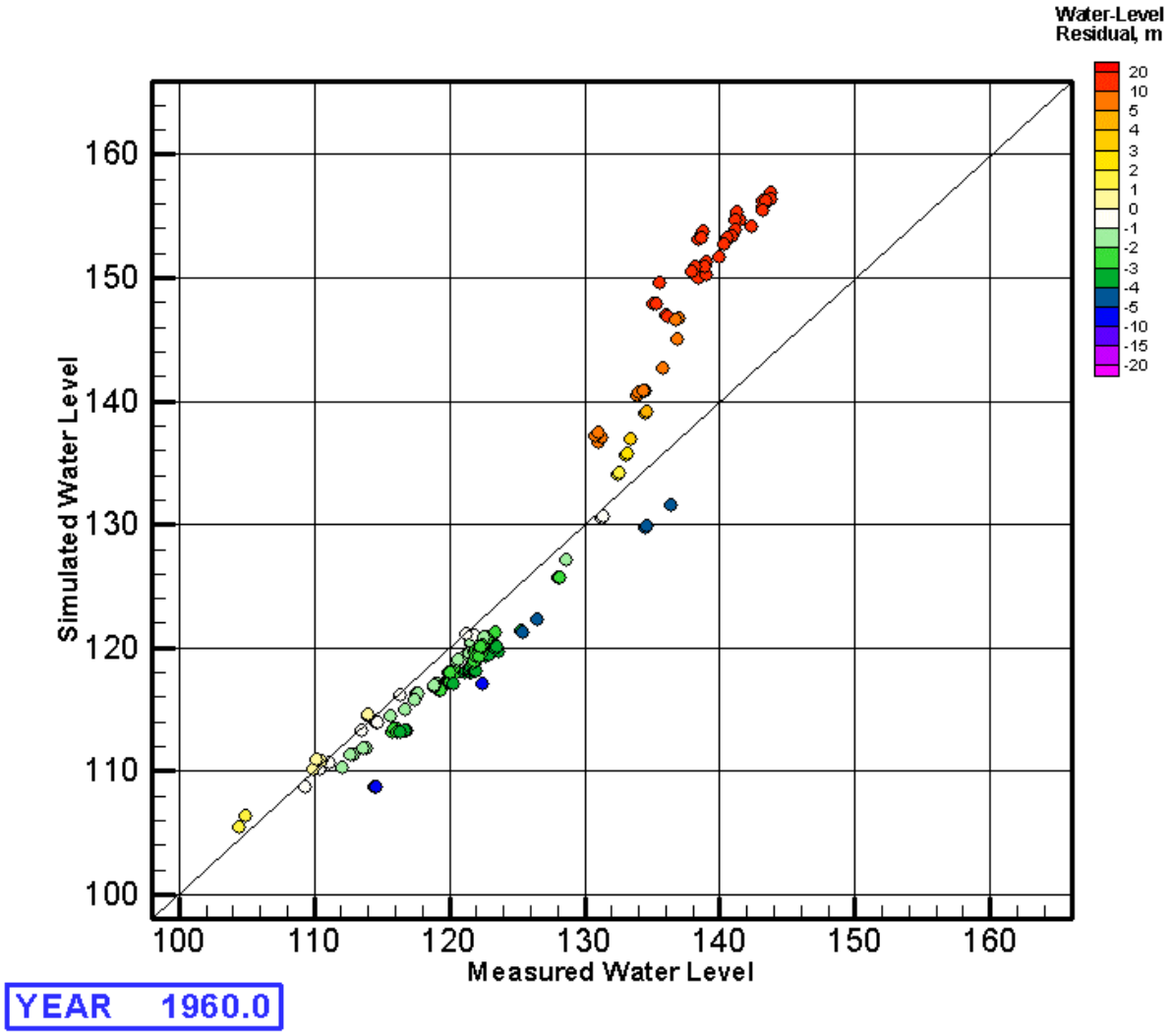

Figure D.3c. Comparison of Simulated versus Measured Water Levels for 1960

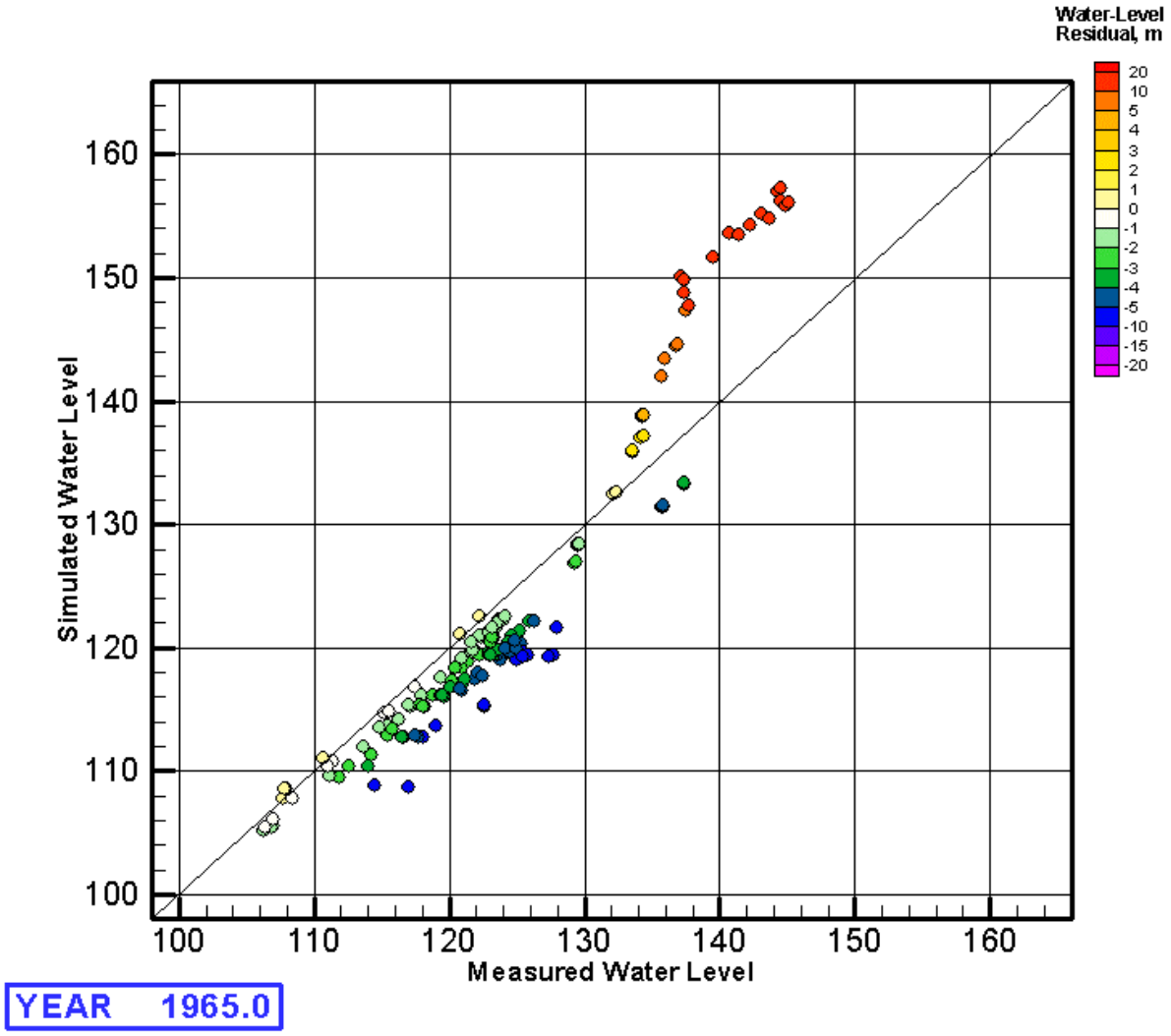

Figure D.3d. Comparison of Simulated versus Measured Water Levels for 1965

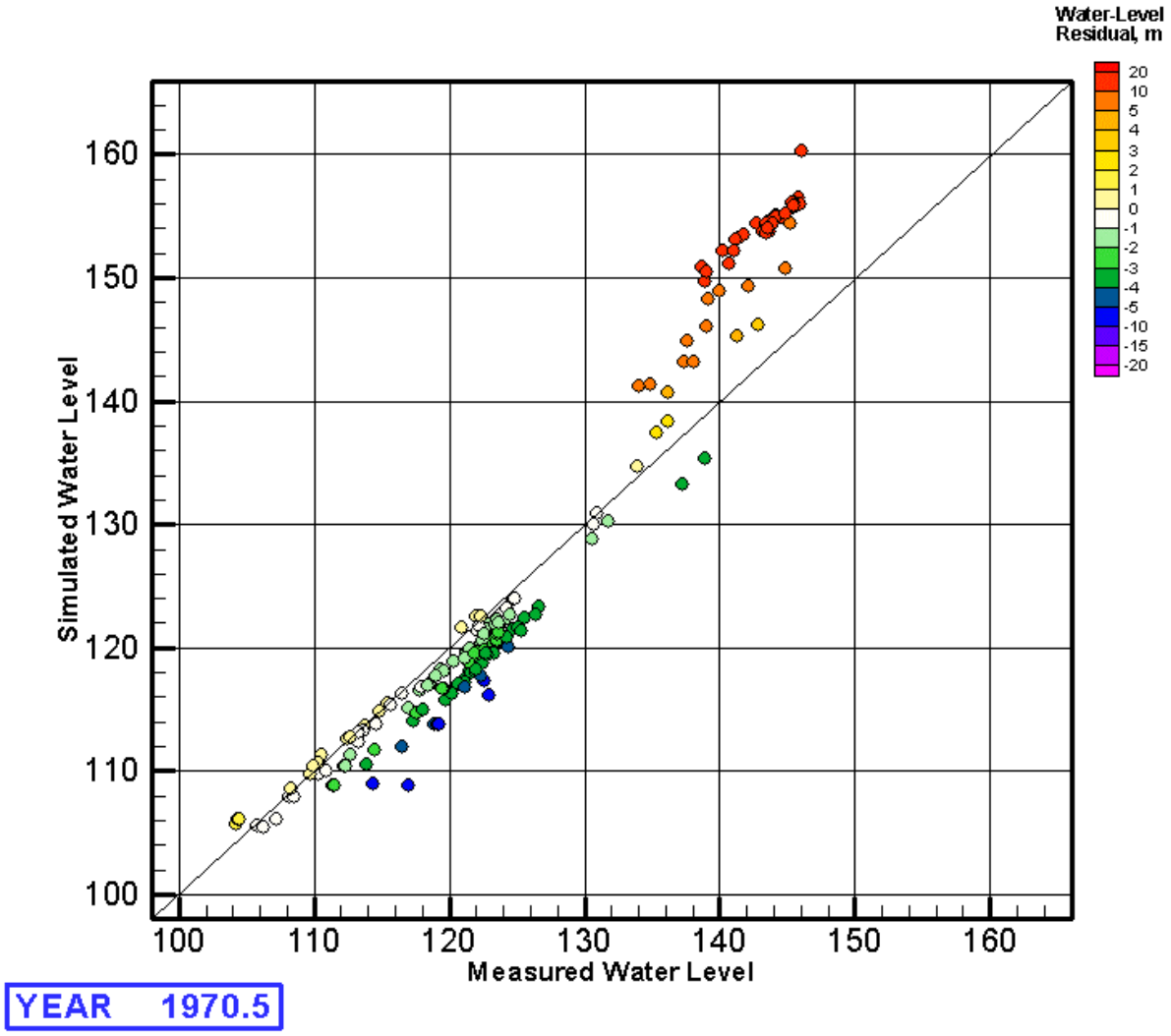

Figure D.3e. Comparison of Simulated versus Measured Water Levels for 1970

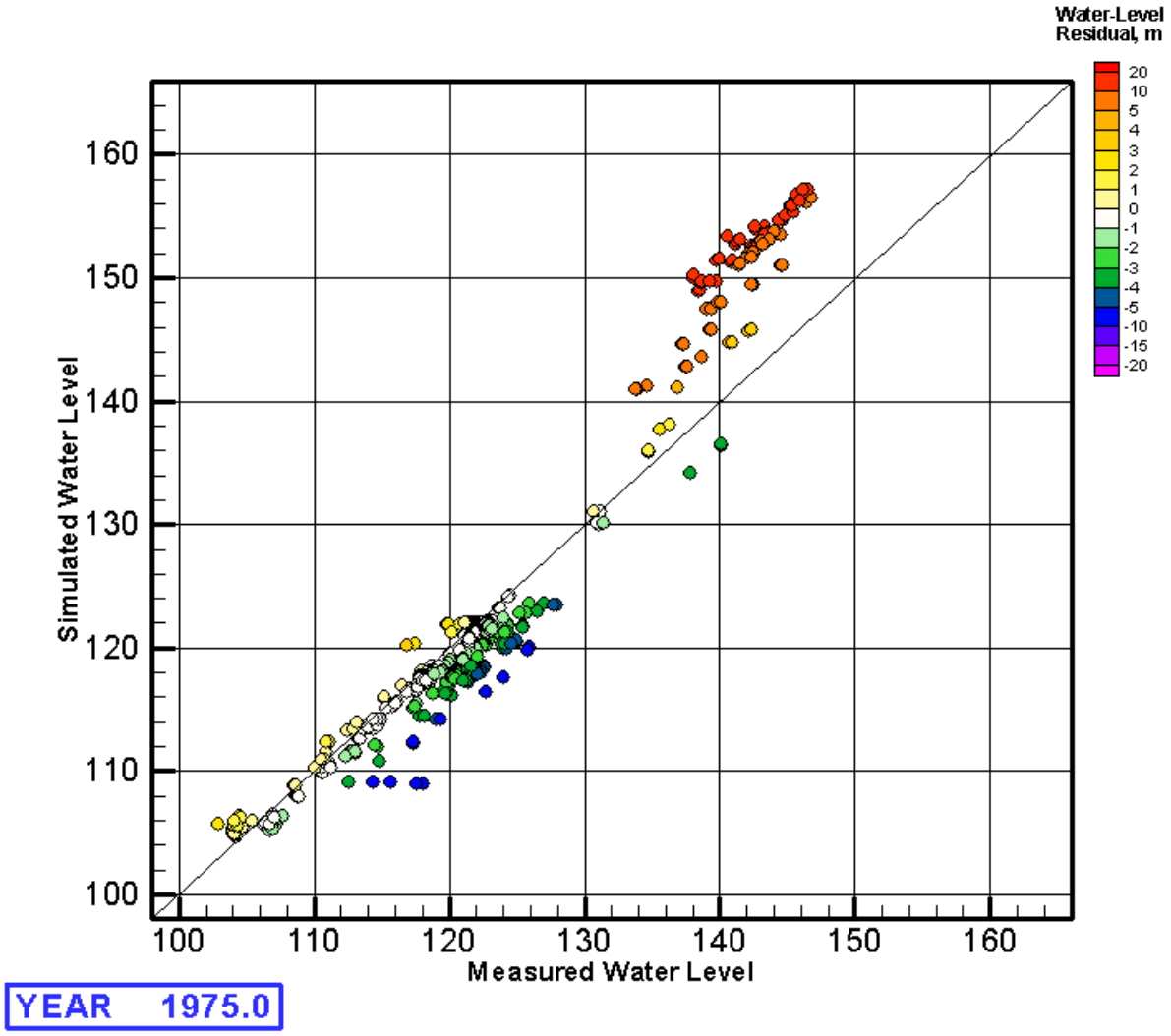

Figure D.3f. Comparison of Simulated versus Measured Water Levels for 1975

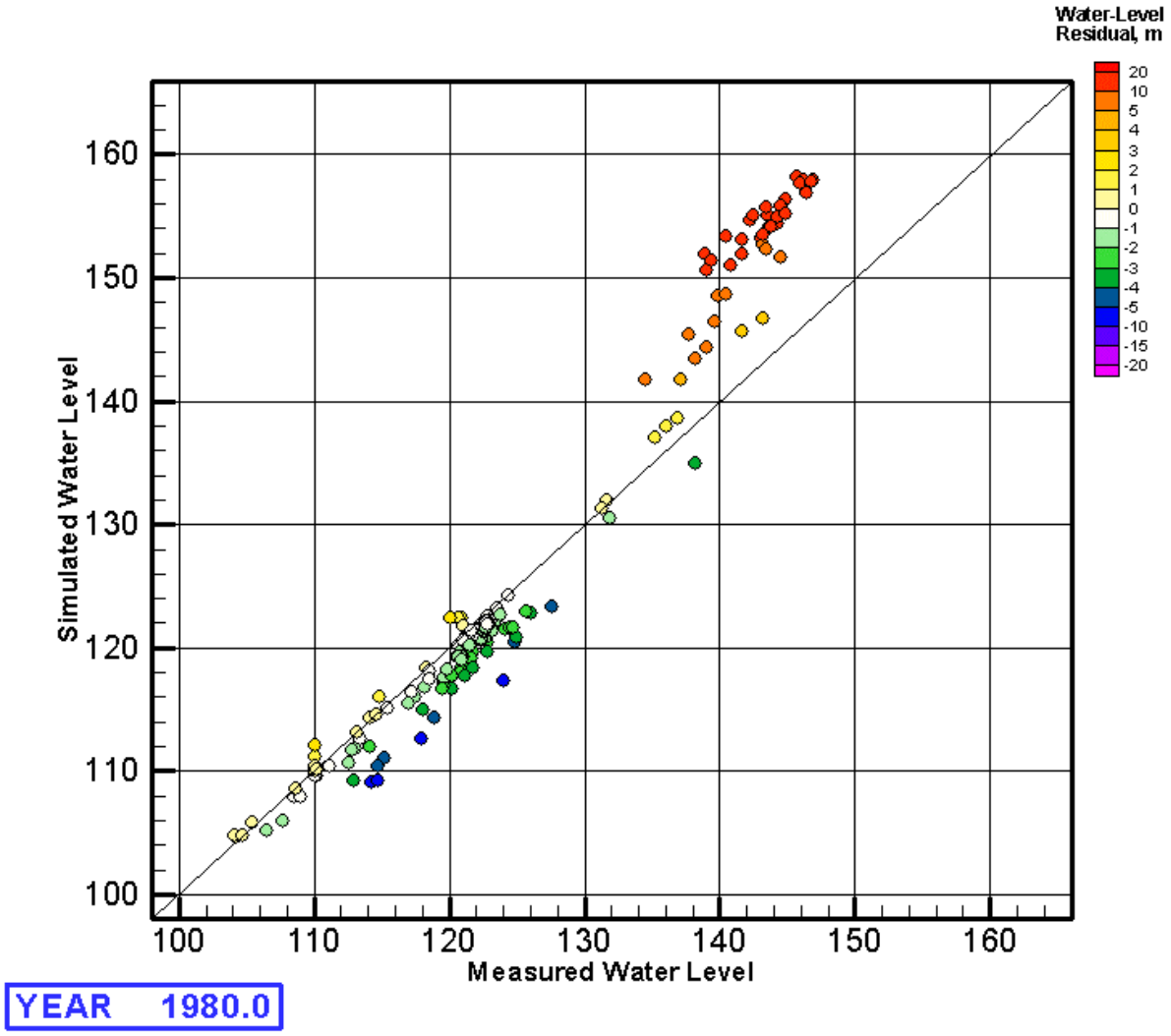

Figure D.3g. Comparison of Simulated versus Measured Water Levels for 1980

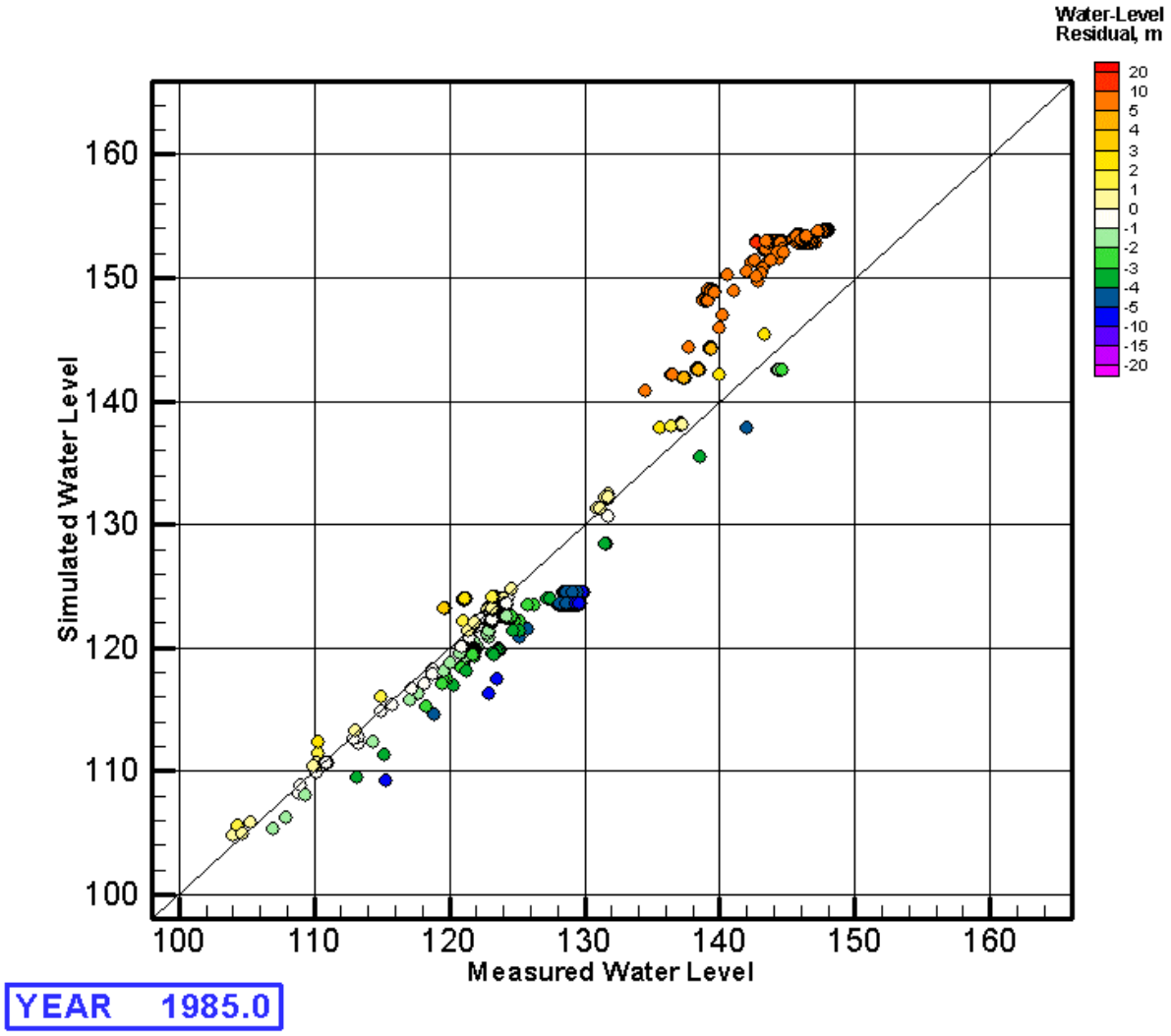

Figure D.3h. Comparison of Simulated versus Measured Water Levels for 1985

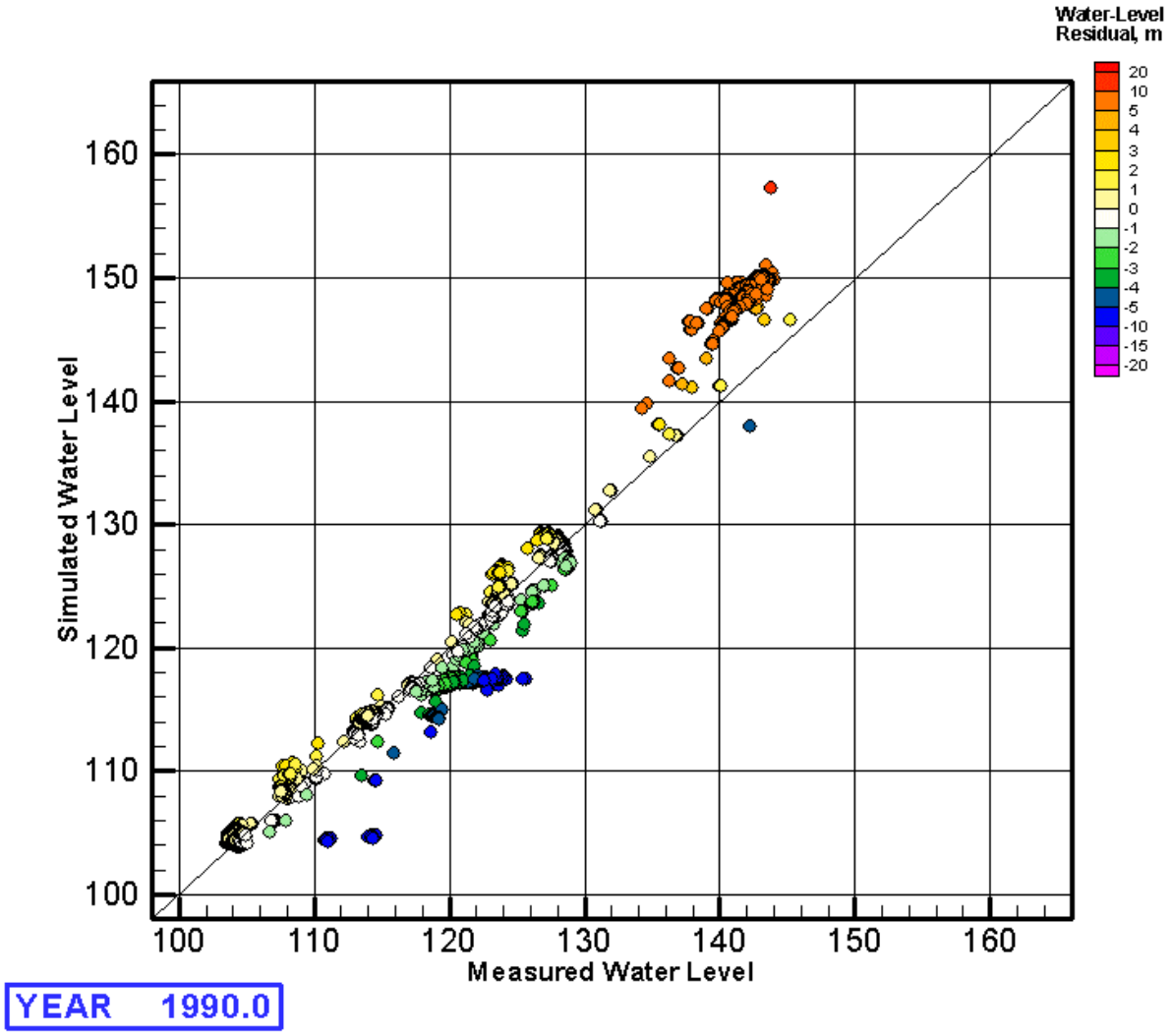

Figure D.3i. Comparison of Simulated versus Measured Water Levels for 1990

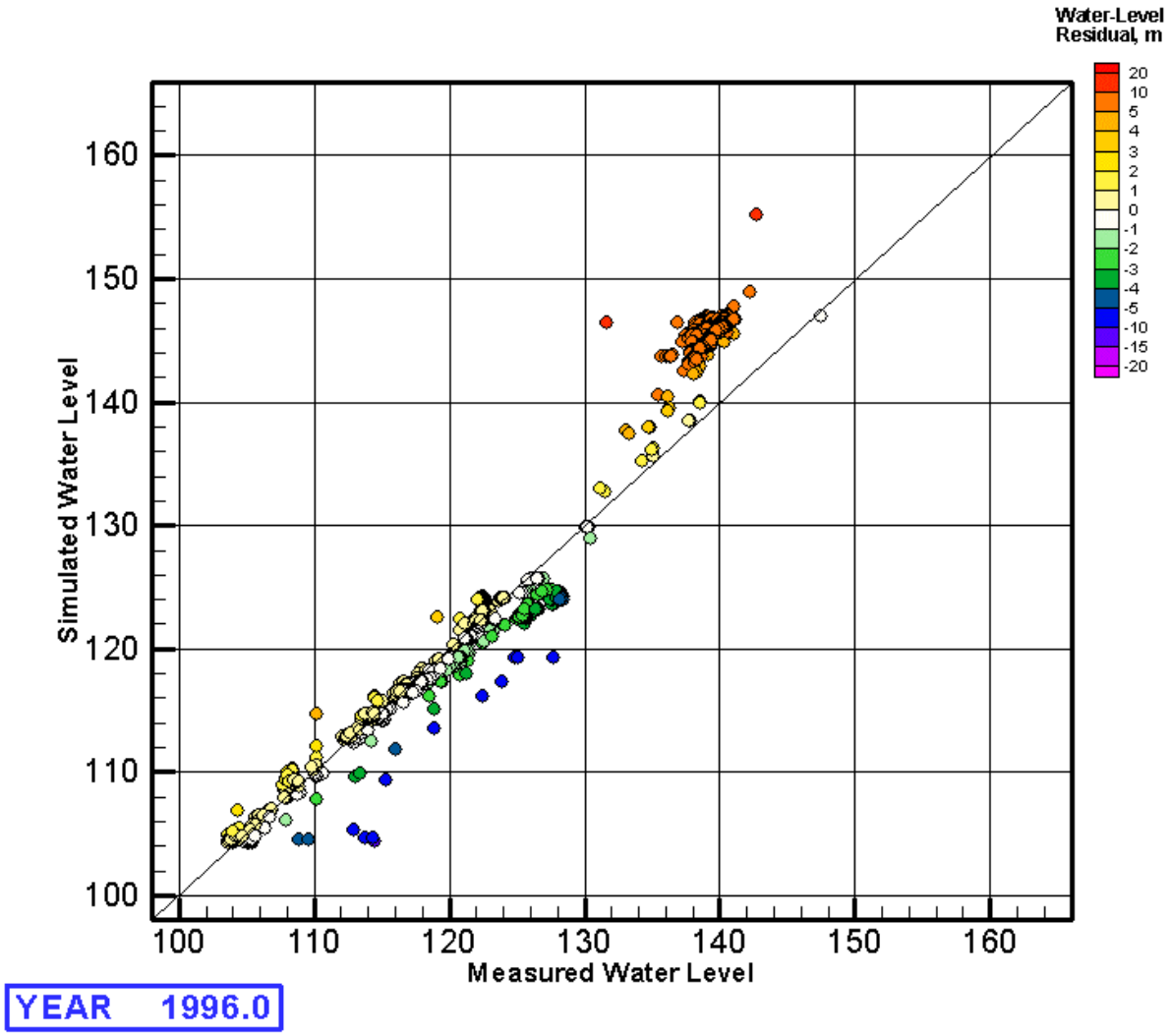

Figure D.3j. Comparison of Simulated versus Measured Water Levels for 1996

Appendix E

Selected Plots of Water Table Elevations and Head Residuals from Simulation of Hanford Wastewater Discharges (1943–1996) Using the Best-Fit Transient Inverse Model

Appendix E

Selected Plots of Water Table Elevations and Head Residuals from Simulation of Hanford Wastewater Discharges (1943-1996) Using Best-Fit Transient Inverse Model

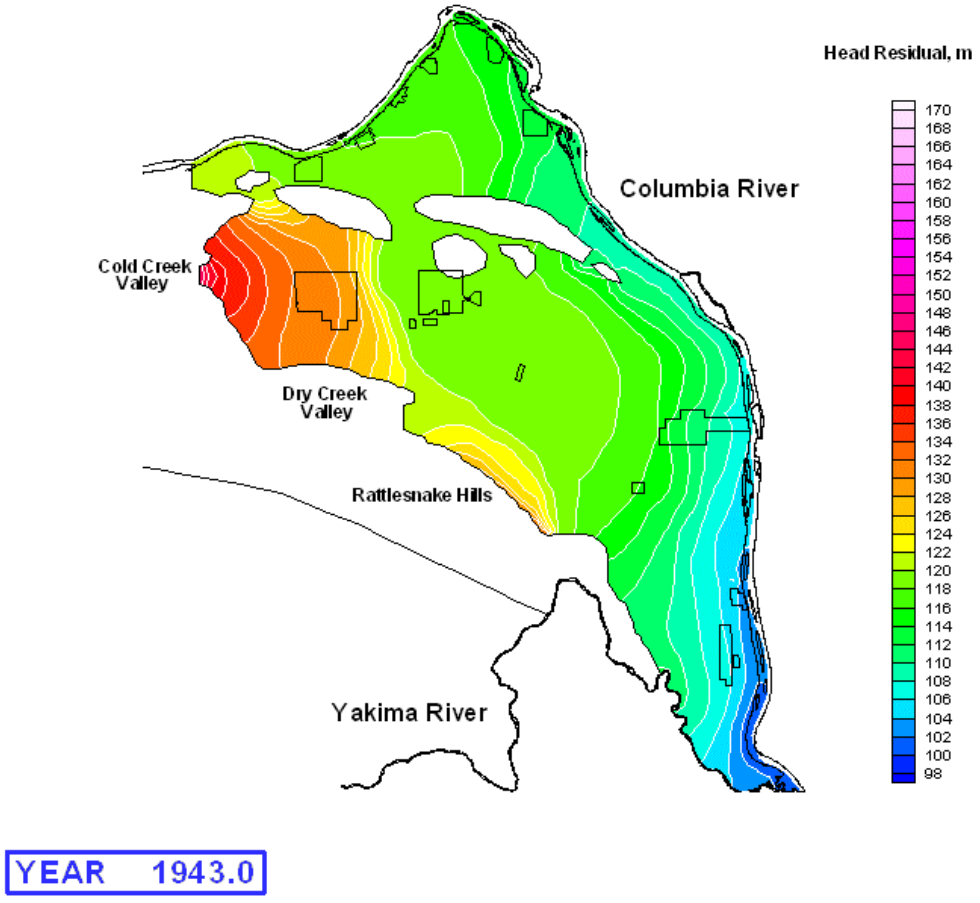

Figure E.1a. Simulated Water-Table Elevations for 1943

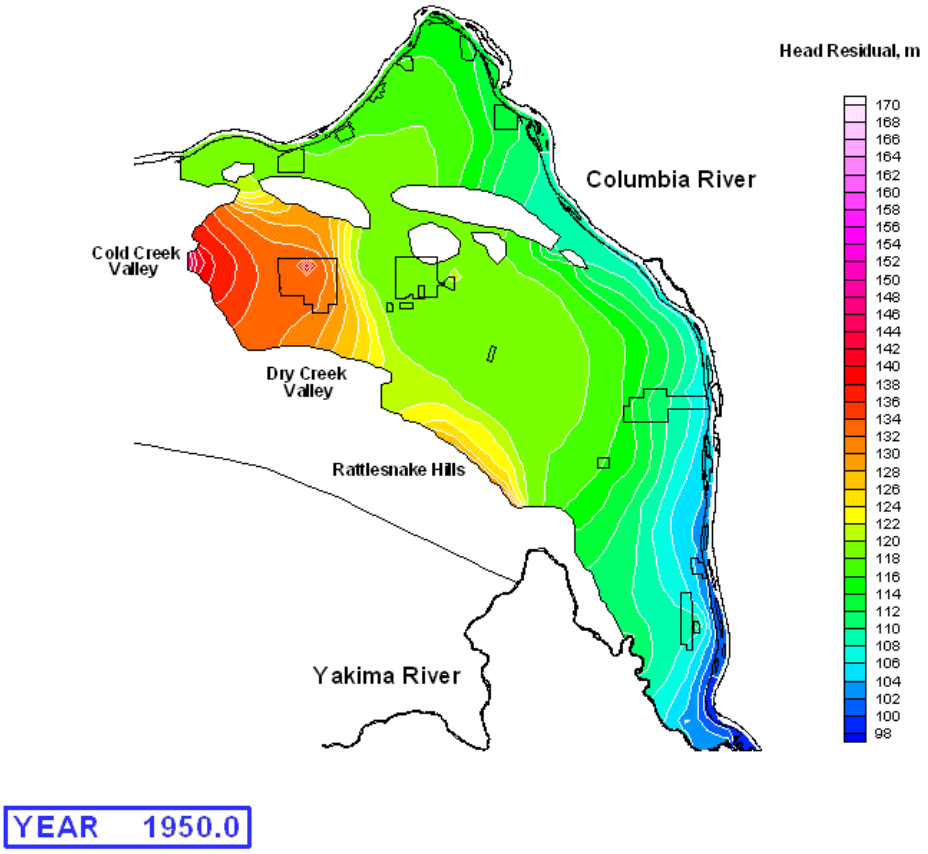

Figure E.1b. Simulated Water-Table Elevations for 1950

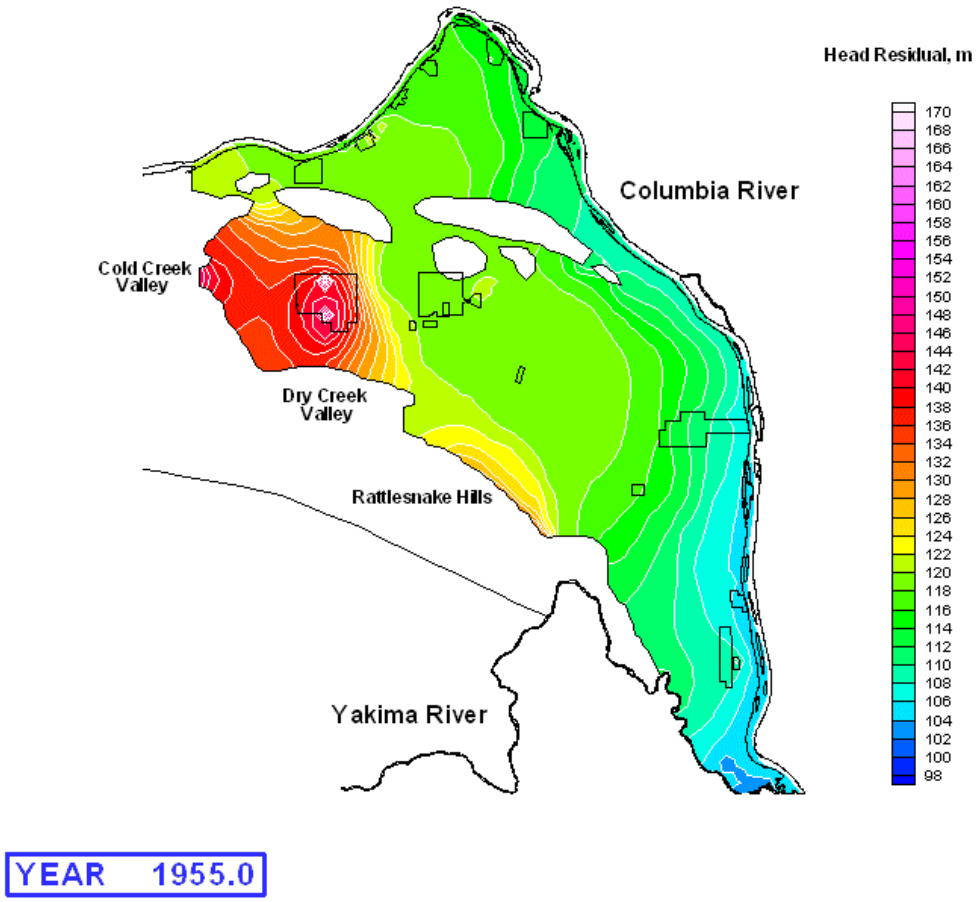

Figure E.1c. Simulated Water-Table Elevations for 1955

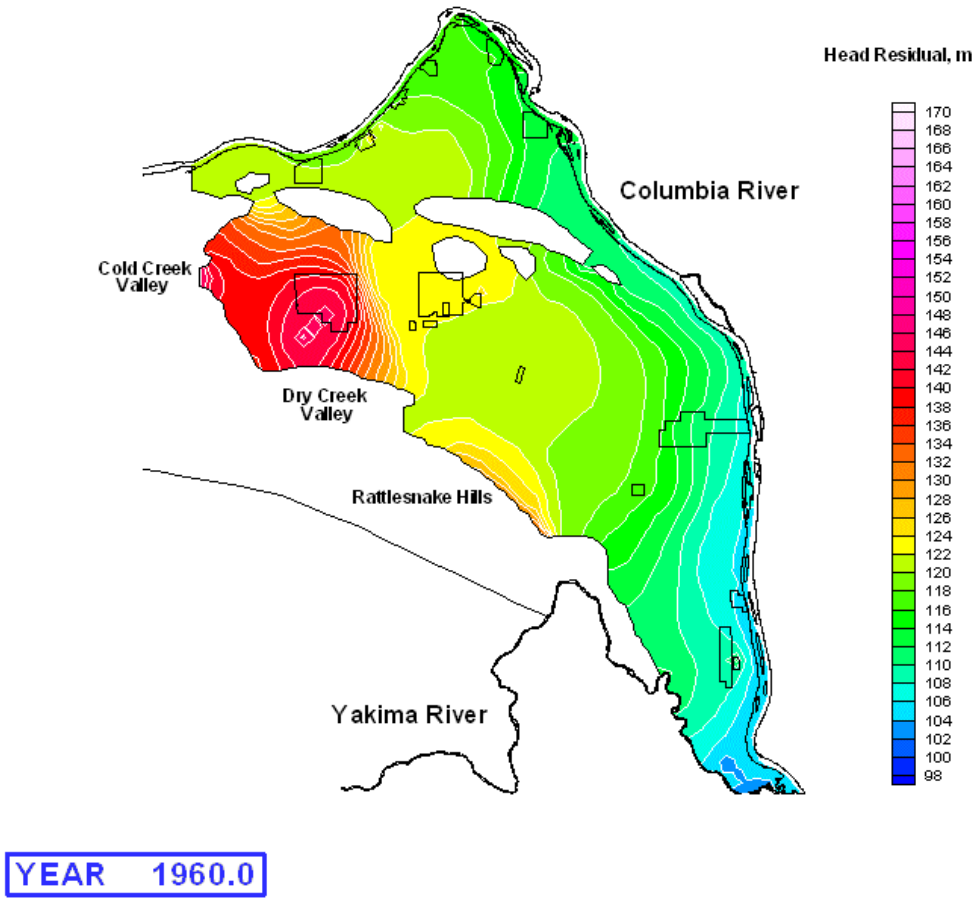

Figure E.1d. Simulated Water-Table Elevations for 1960

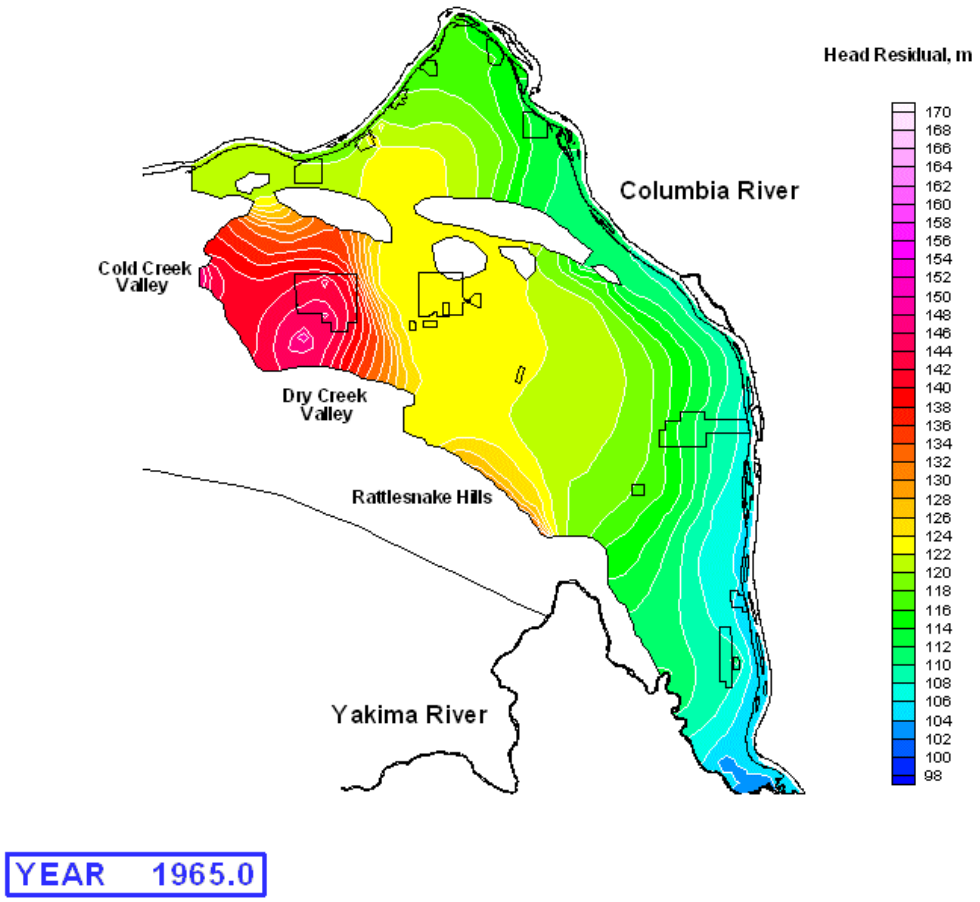

Figure E.1e. Simulated Water-Table Elevations for 1965

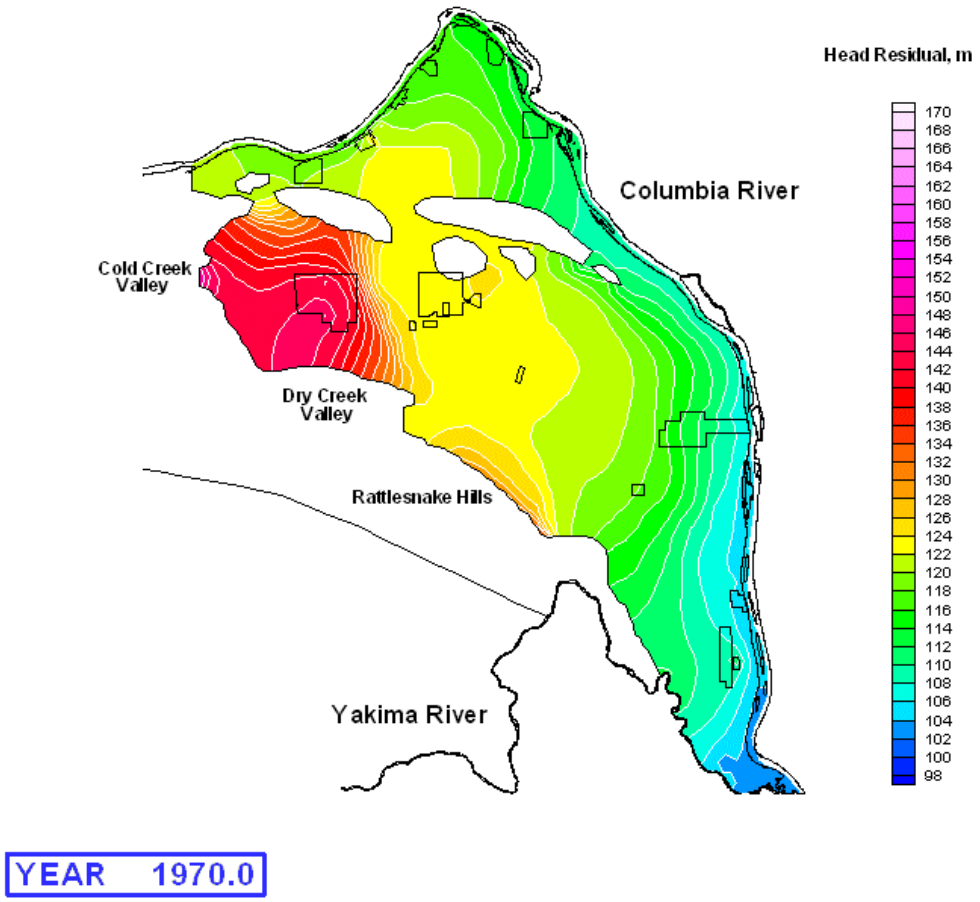

Figure E.1f. Simulated Water-Table Elevations for 1970

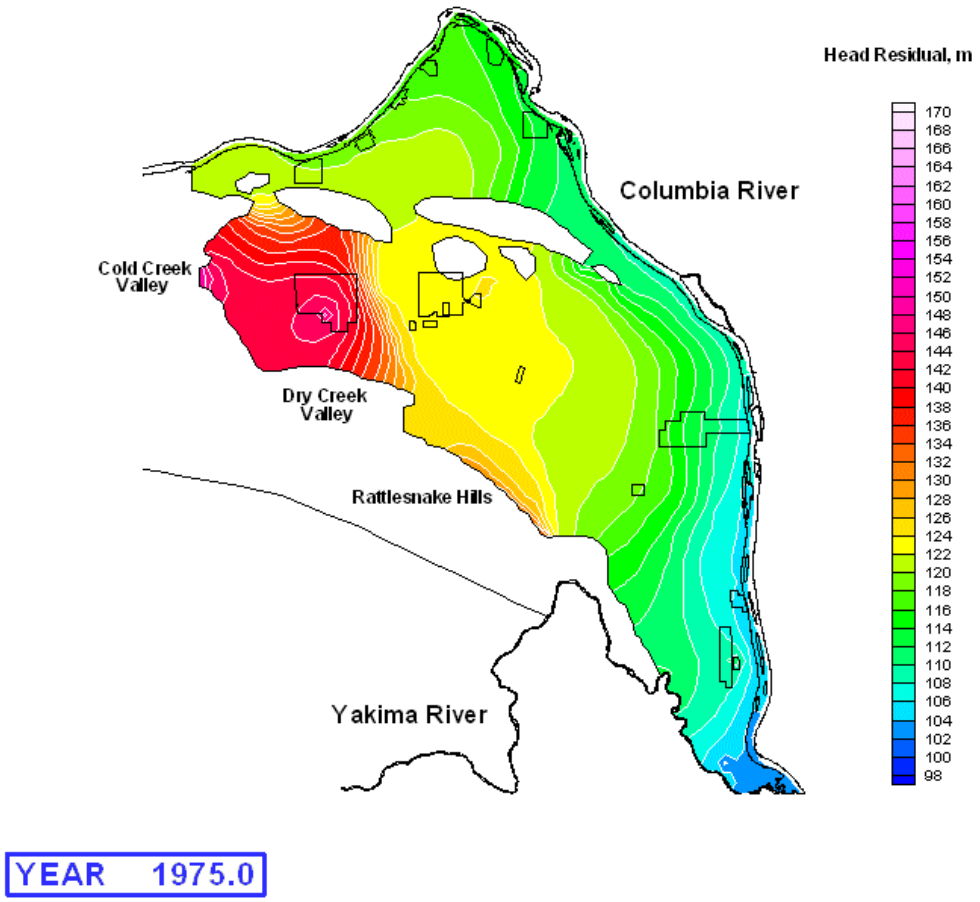

Figure E.1g. Simulated Water-Table Elevations for 1975

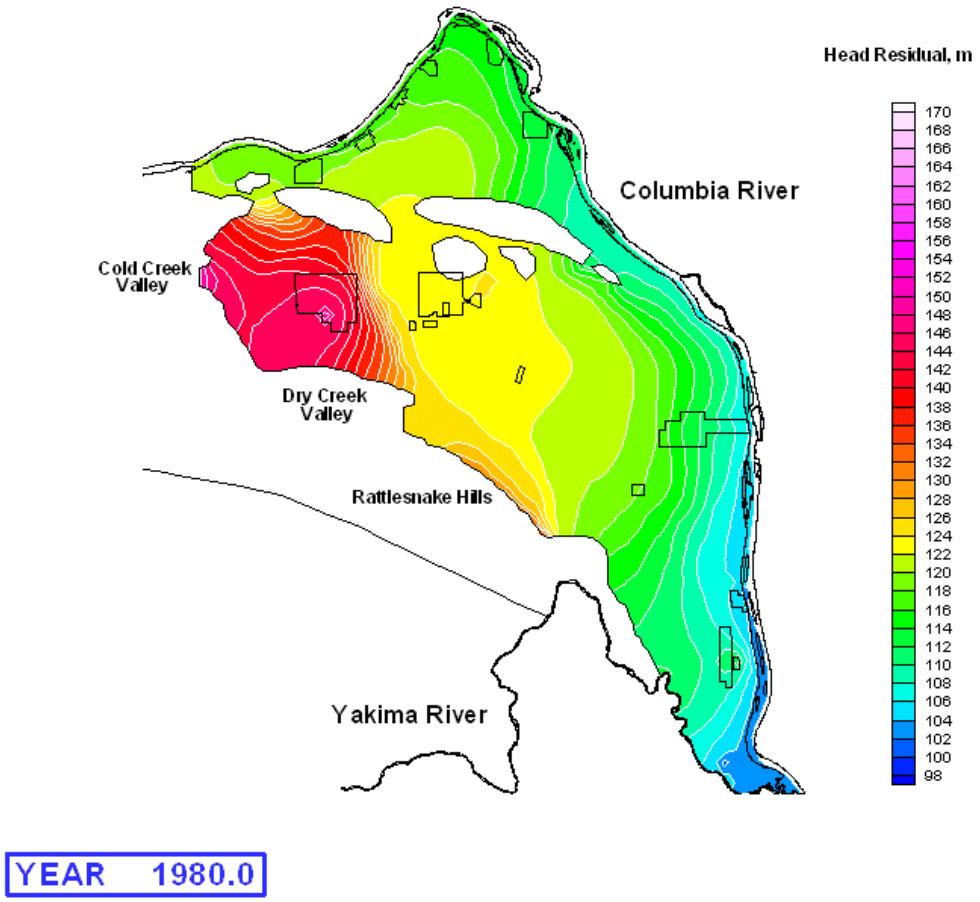

Figure E.1h. Simulated Water-Table Elevations for 1980

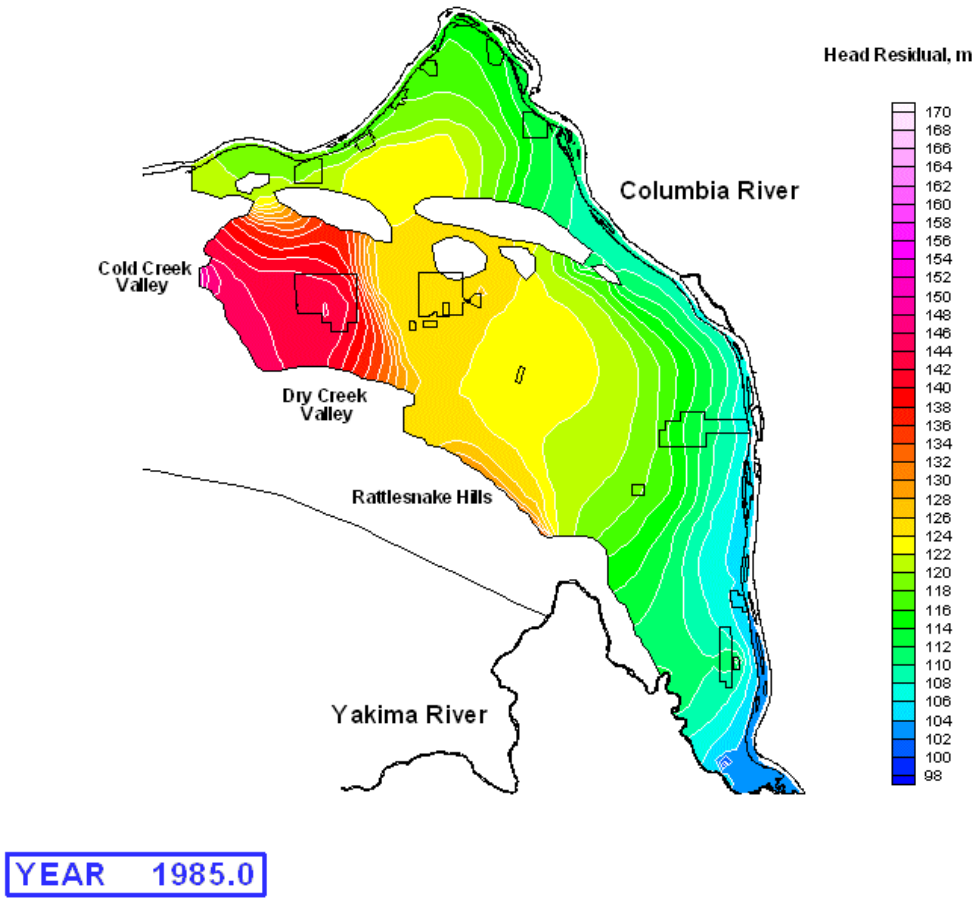

Figure E.1i. Simulated Water-Table Elevations for 1985

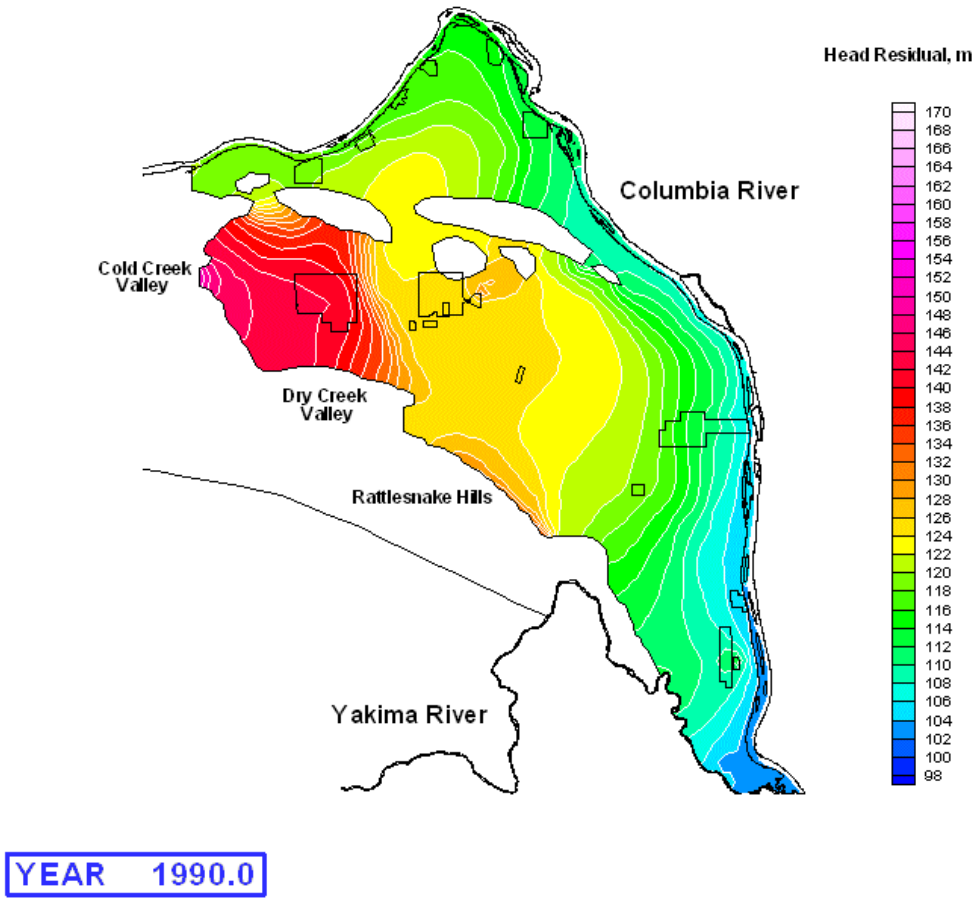

Figure E.1j. Simulated Water-Table Elevations for 1990

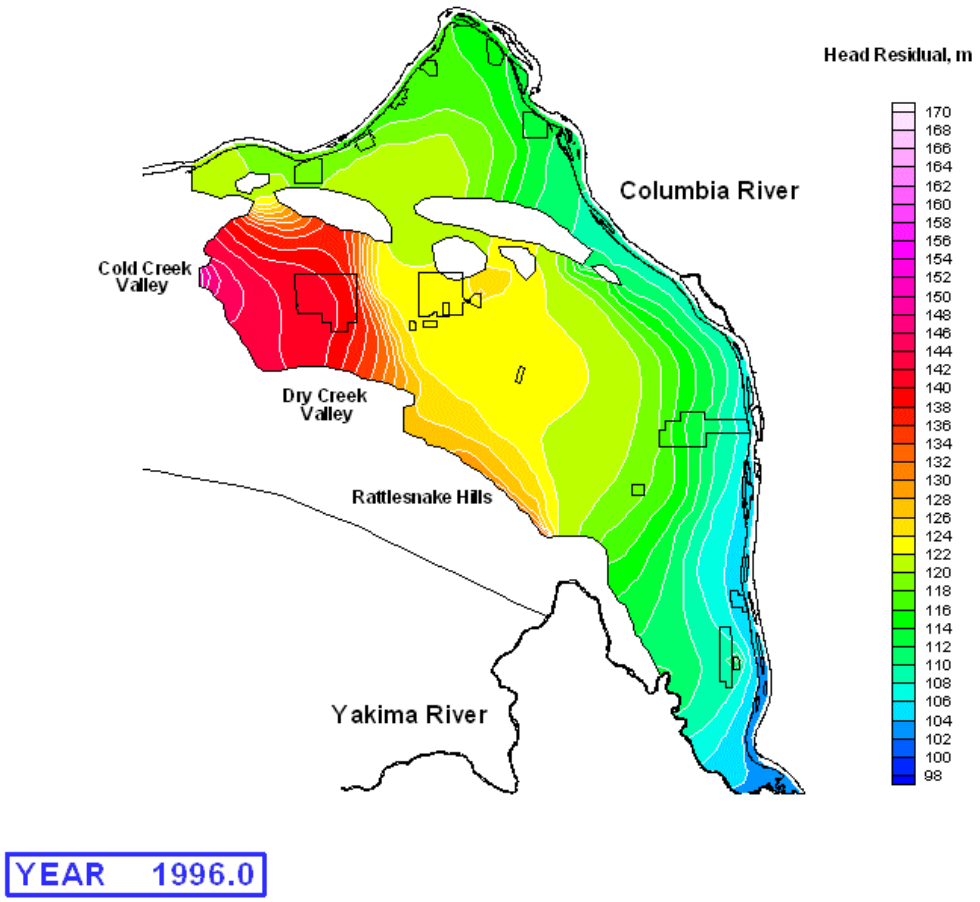

Figure E.1k. Simulated Water-Table Elevations for 1996

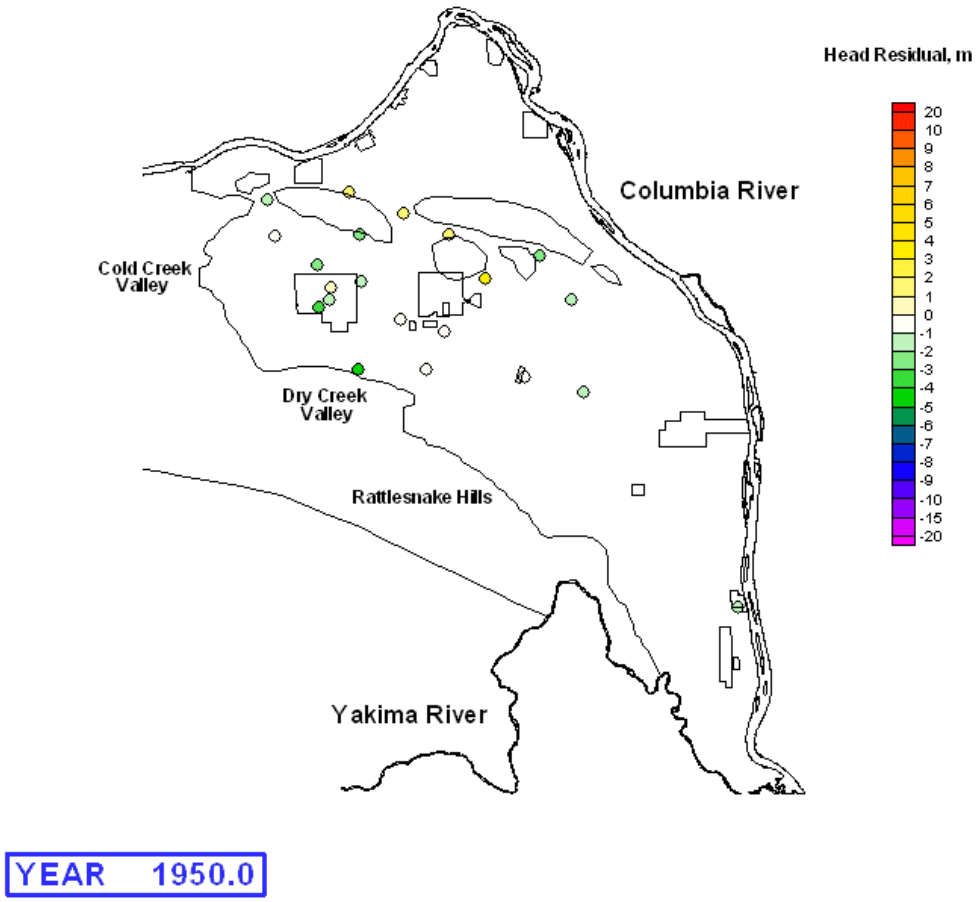

Figure E.2a. Hydraulic Head Residuals (simulated - measured water levels) for 1950

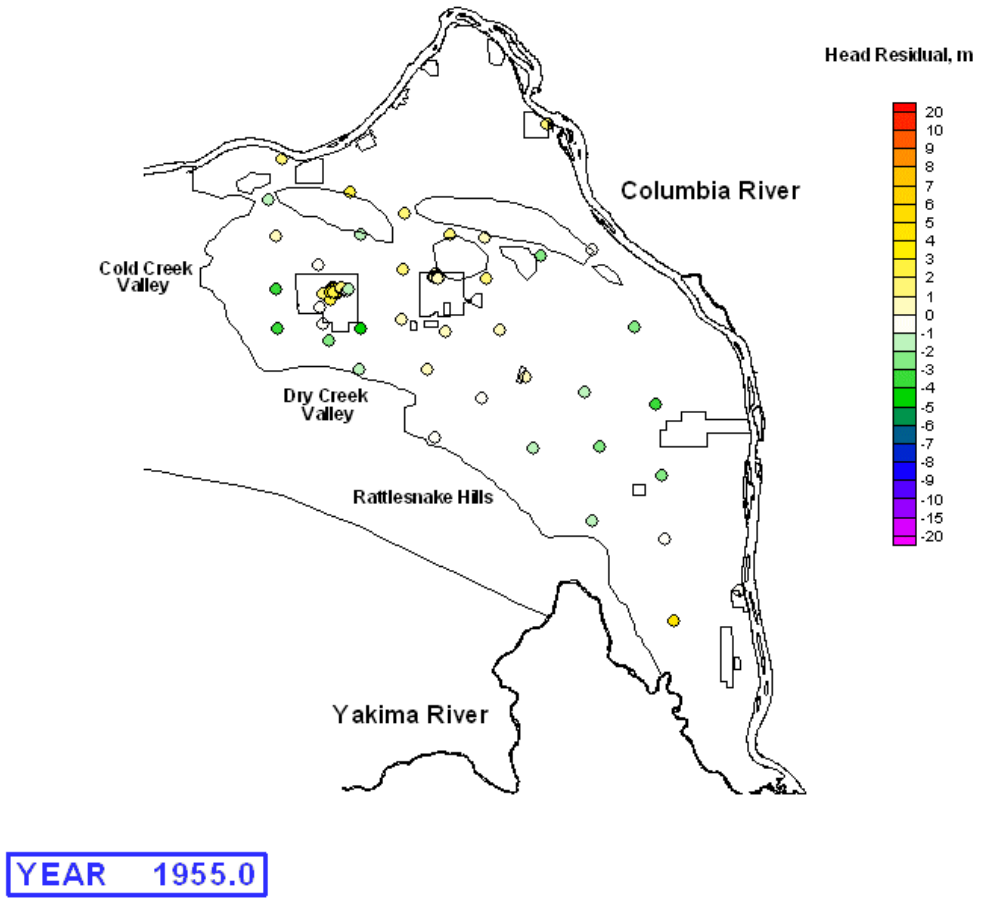

Figure E.2b. Hydraulic Head Residuals (simulated - measured water levels) for 1955

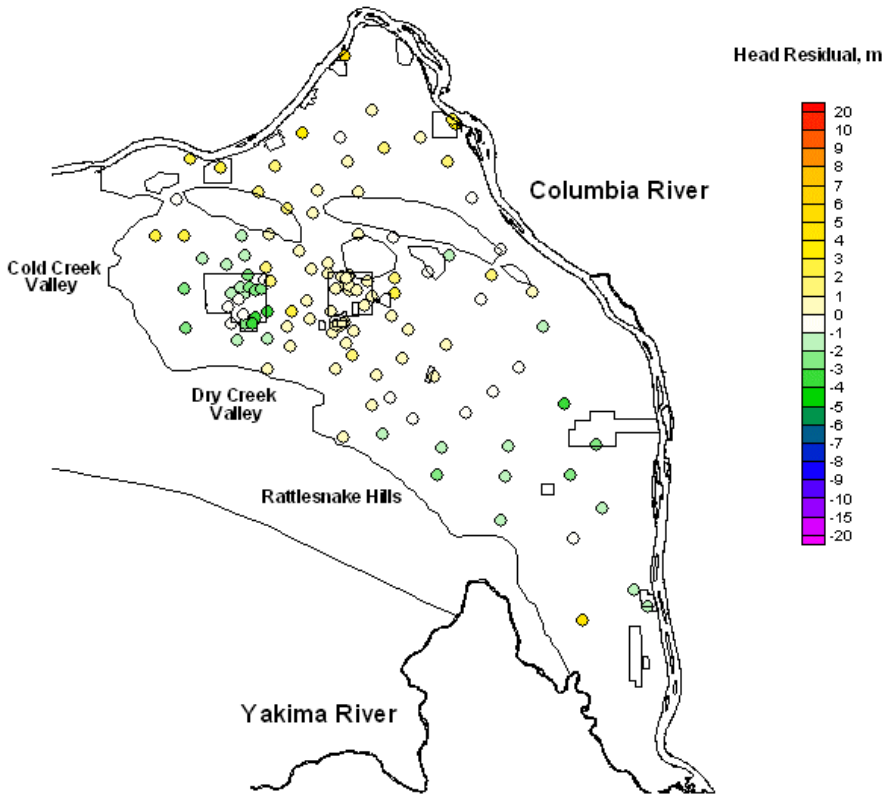

Figure E.2c. Hydraulic Head Residuals (simulated - measured water levels) for 1960

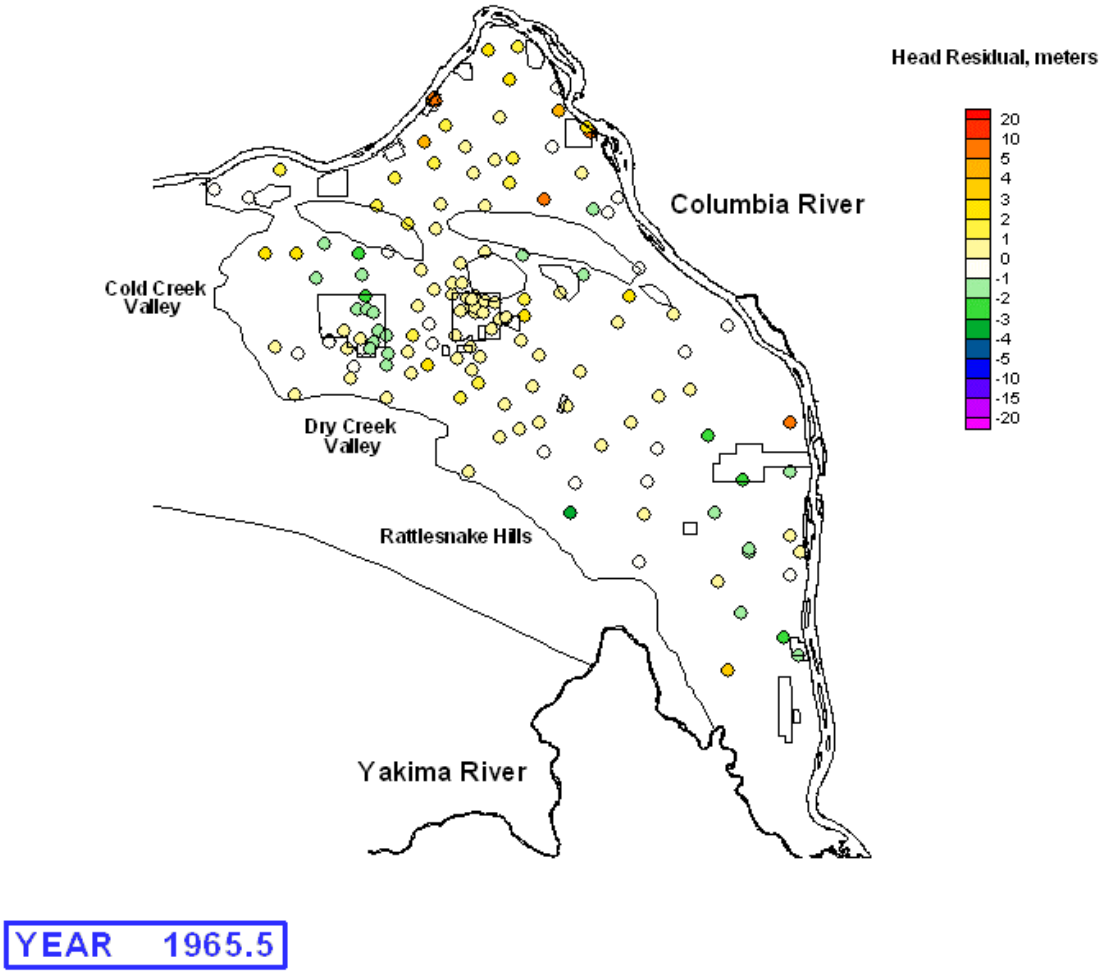

Figure E.2d. Hydraulic Head Residuals (simulated - measured water levels) for 1965

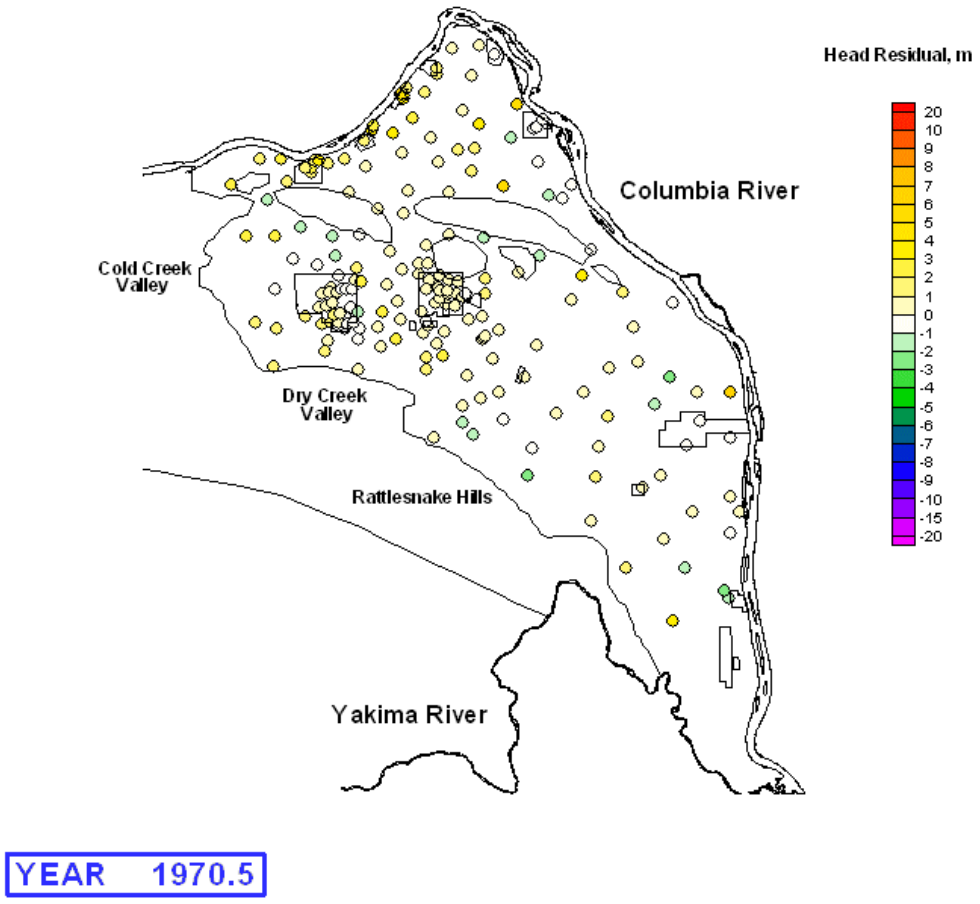

Figure E.2e. Hydraulic Head Residuals (simulated - measured water levels) for 1970

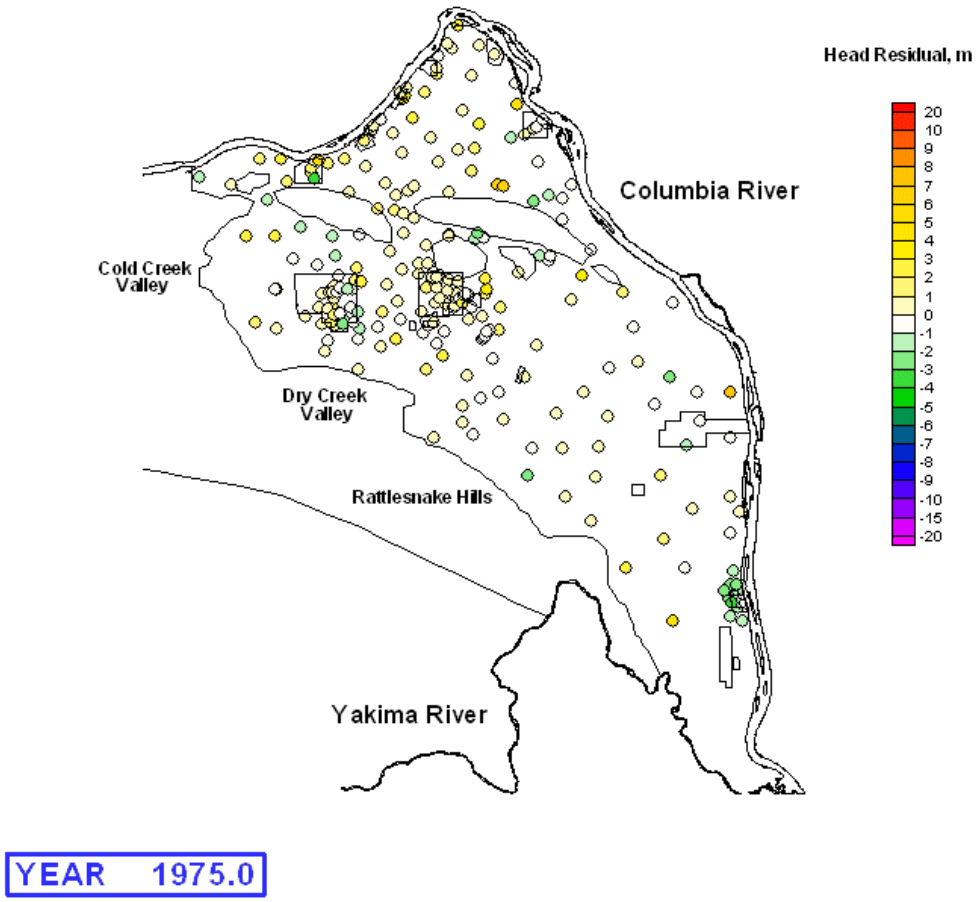

Figure E.2f. Hydraulic Head Residuals (simulated - measured water levels) for 1975

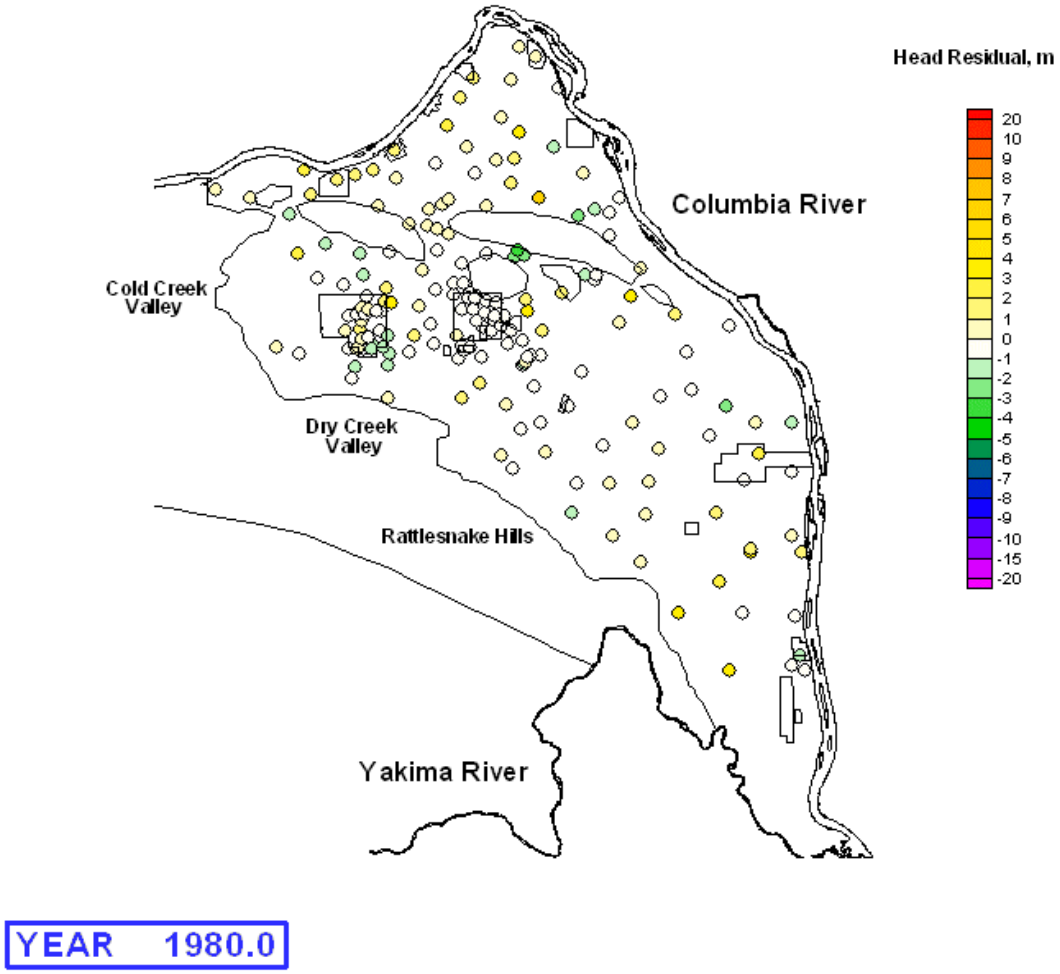

Figure E.2g. Hydraulic Head Residuals (simulated - measured water levels) for 1980

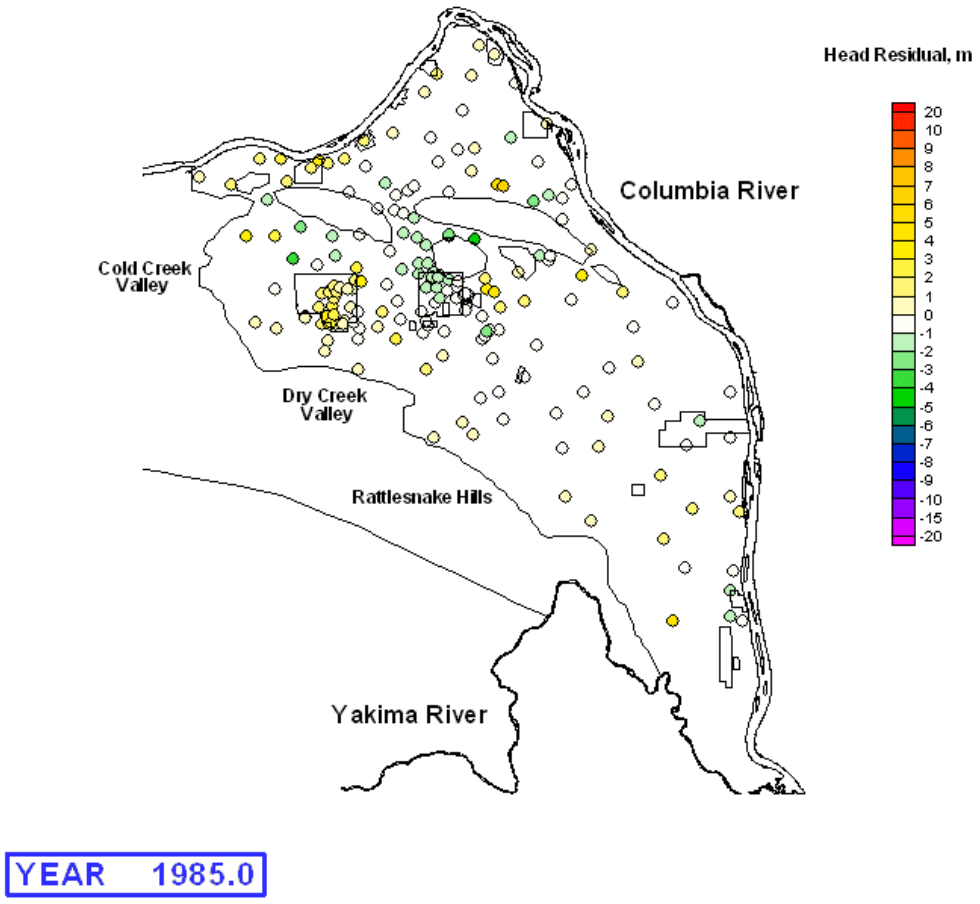

Figure E.2h. Hydraulic Head Residuals (simulated - measured water levels) for 1985

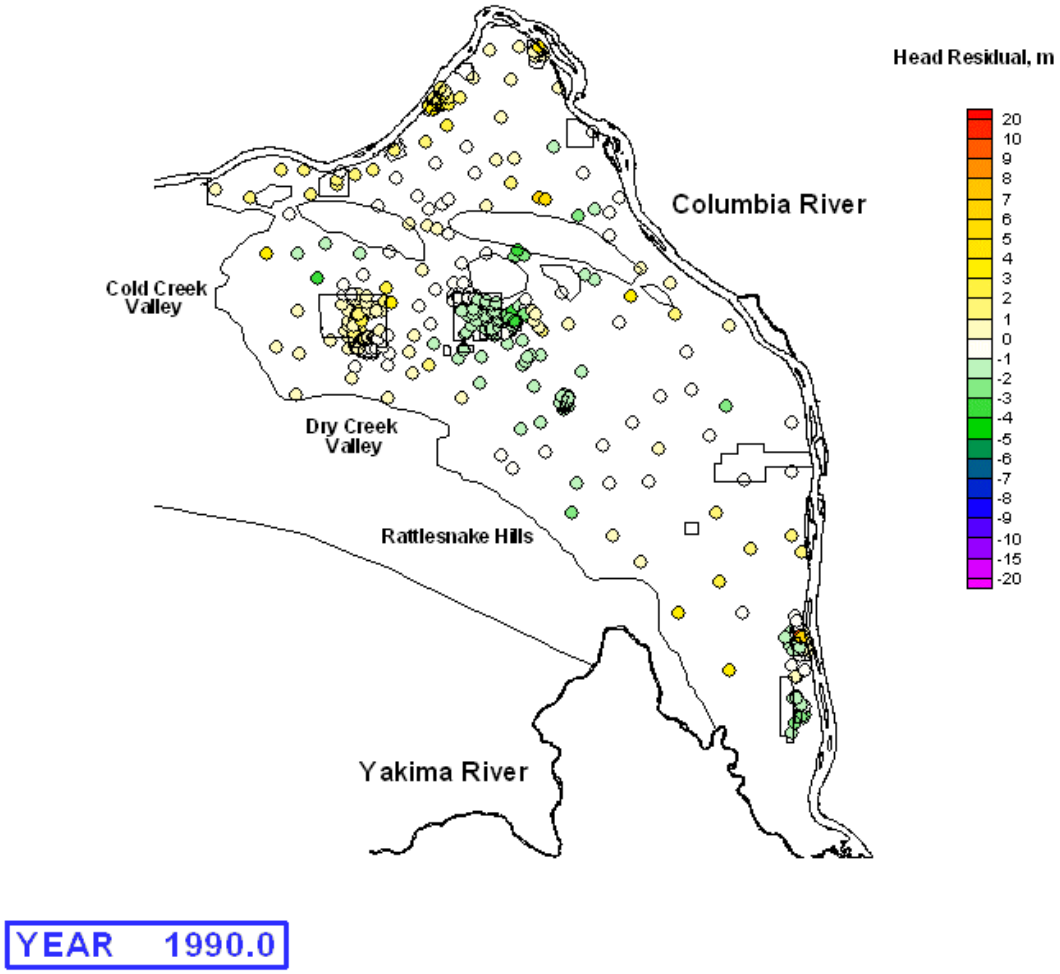

Figure E.2i. Hydraulic Head Residuals (simulated - measured water levels) for 1990

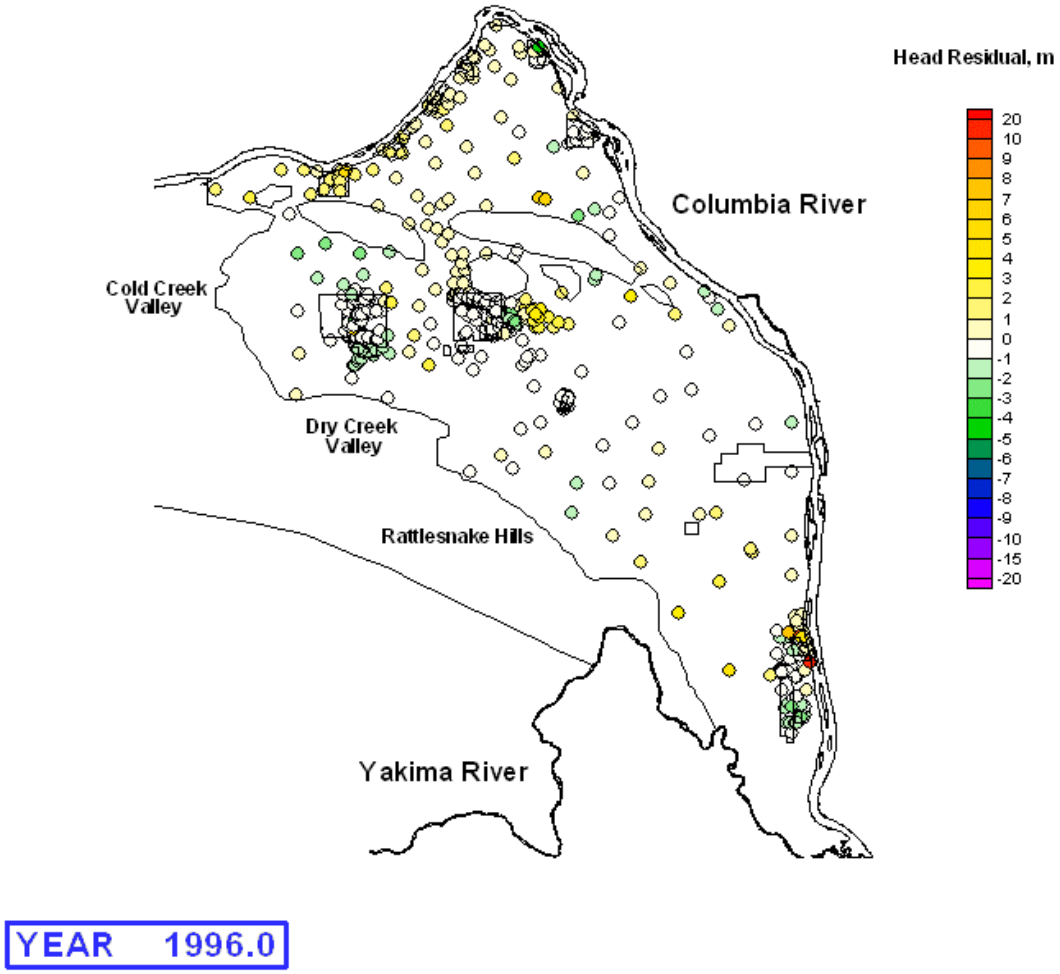

Figure E.2j. Hydraulic Head Residuals (simulated - measured water levels) for 1996

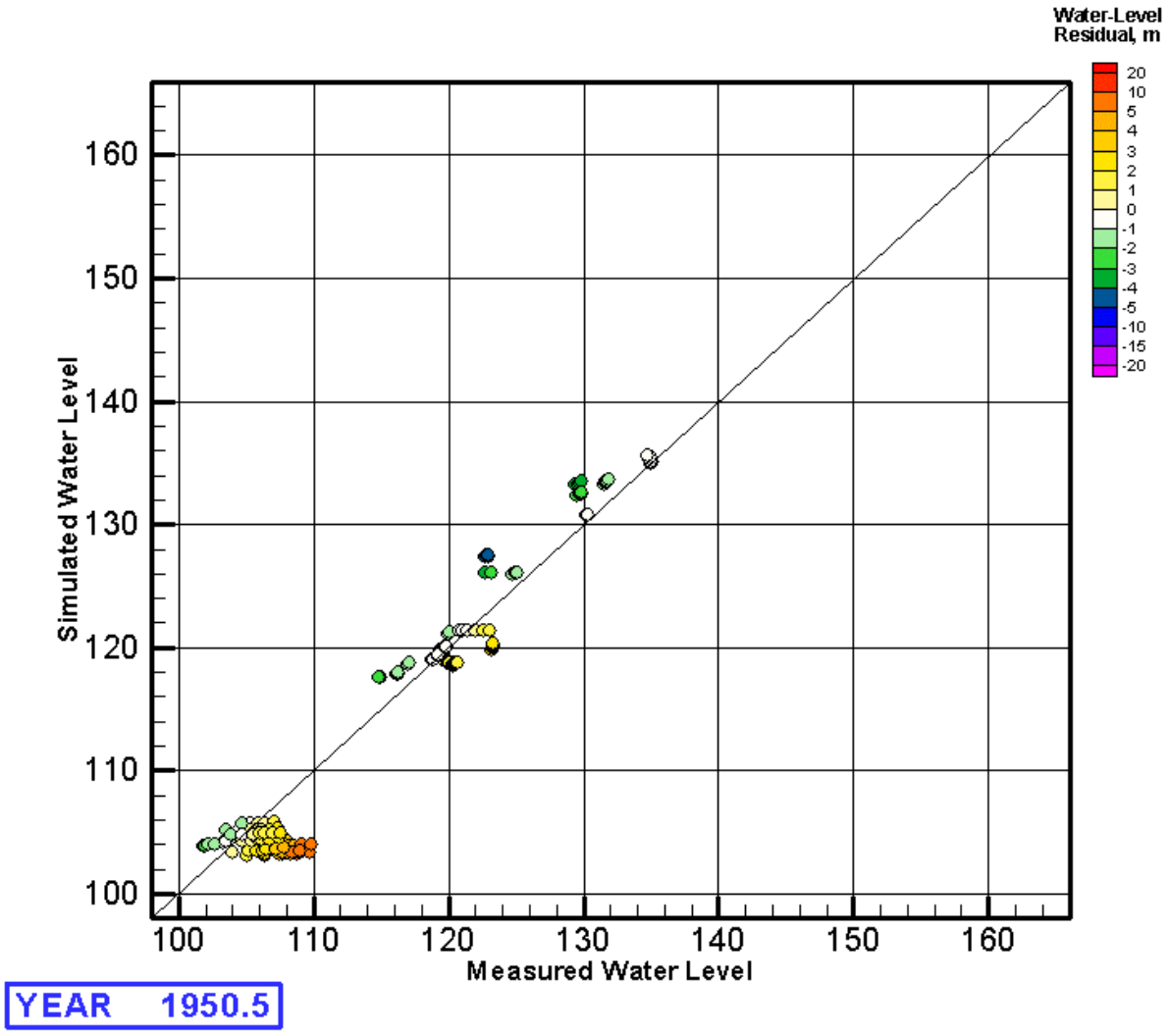

Figure E.3a. Comparison of Simulated versus Measured Water Levels for 1950

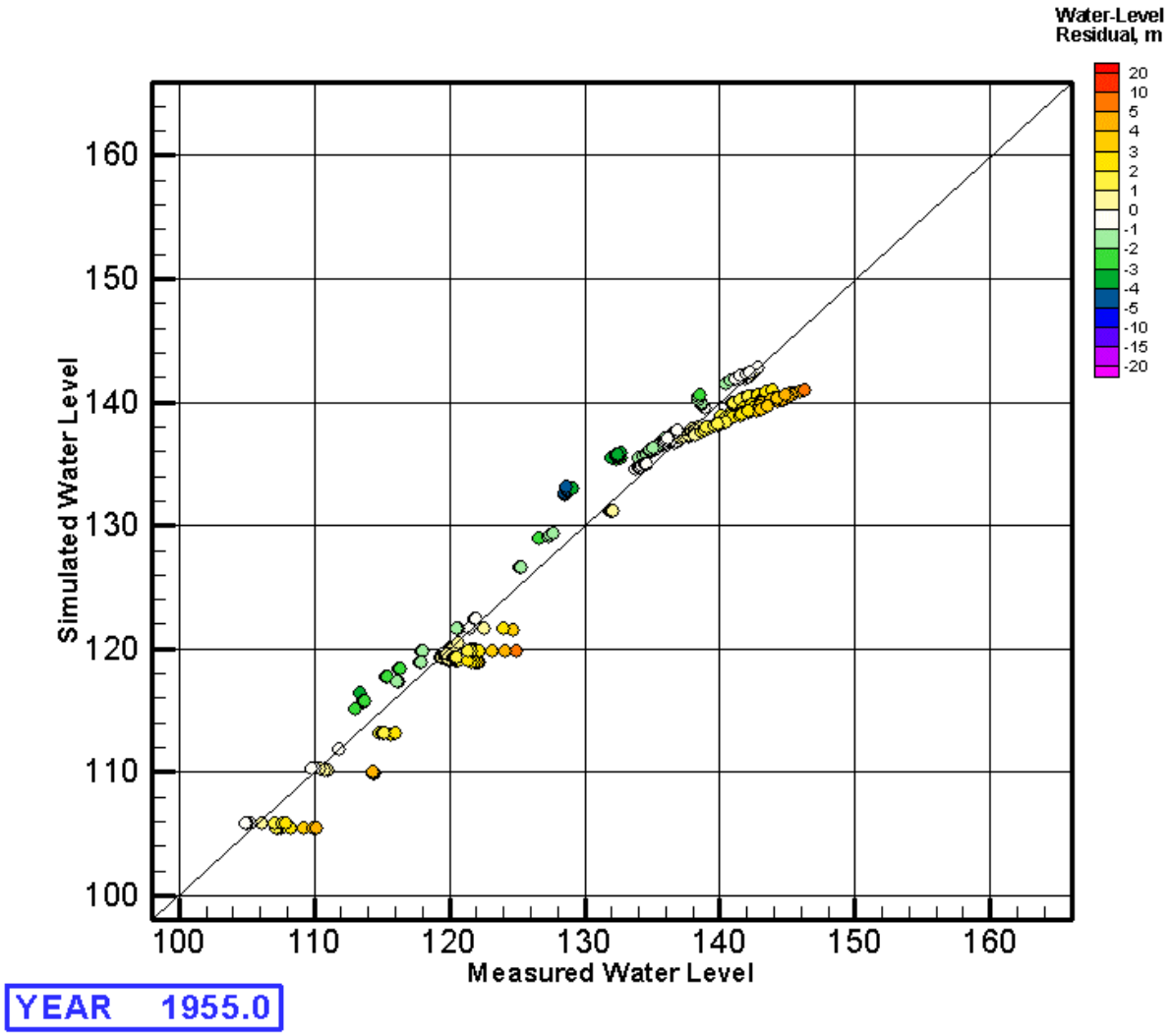

Figure E.3b. Comparison of Simulated versus Measured Water Levels for 1955

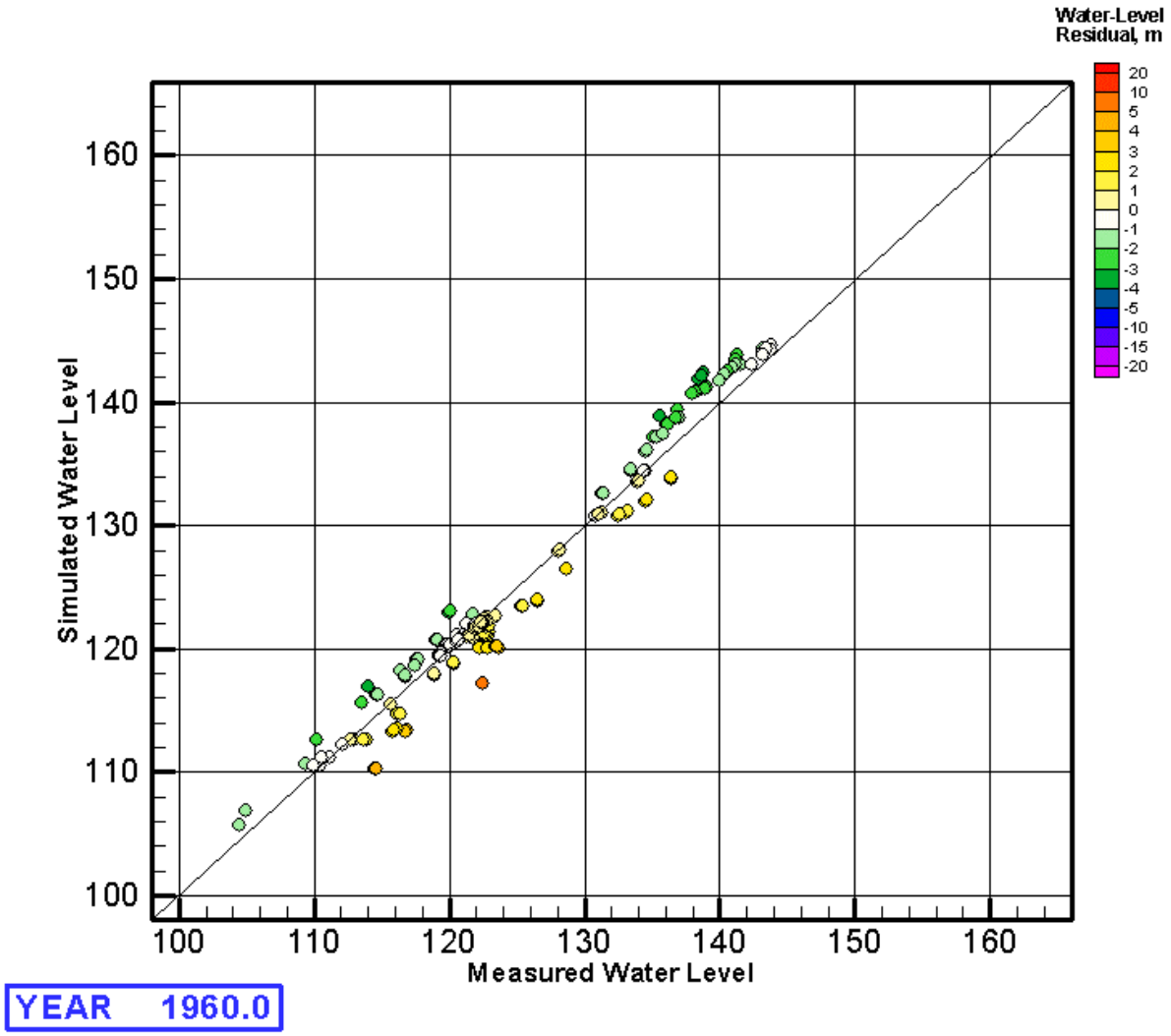

Figure E.3c. Comparison of Simulated versus Measured Water Levels for 1960

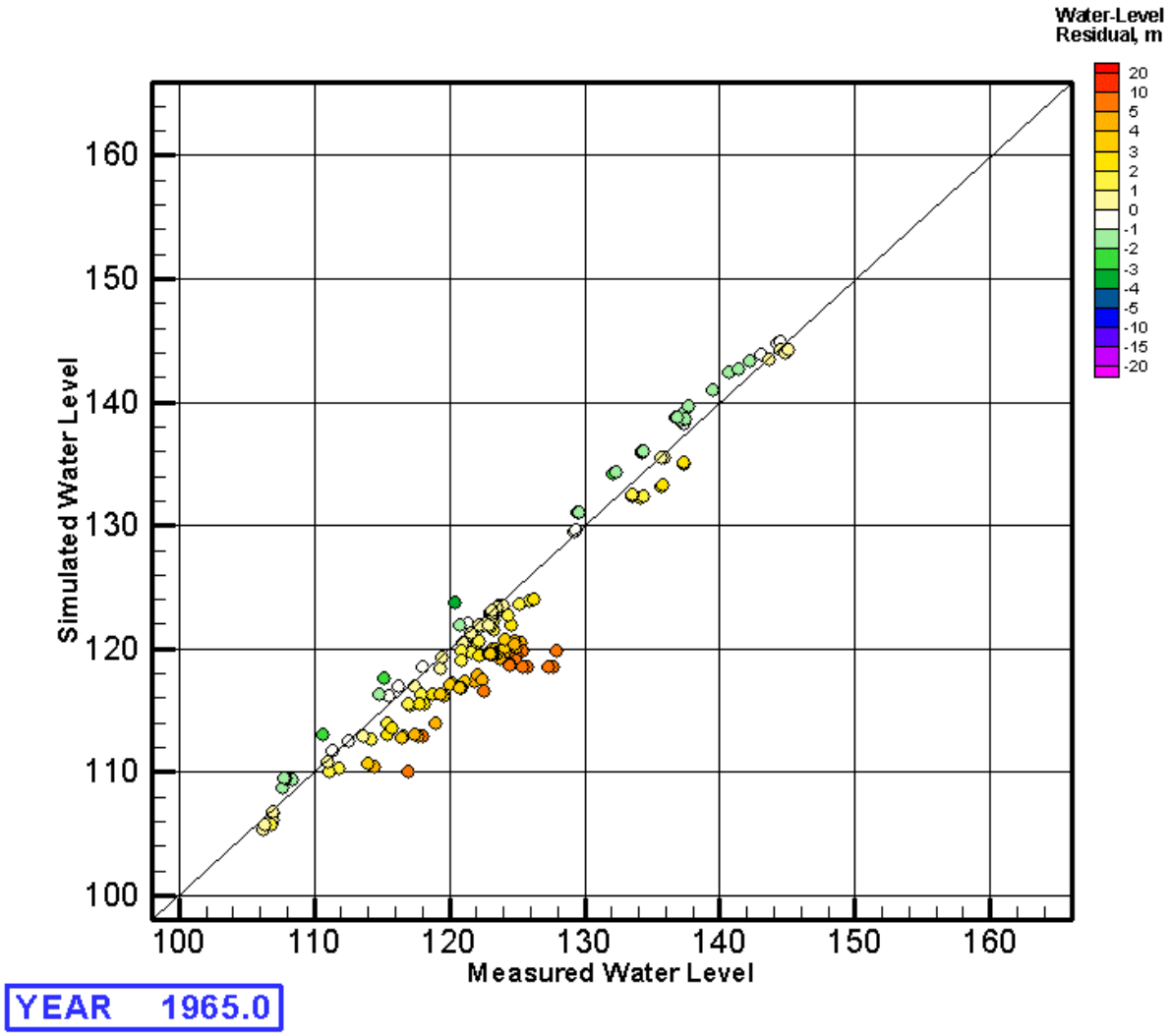

Figure E.3d. Comparison of Simulated versus Measured Water Levels for 1965

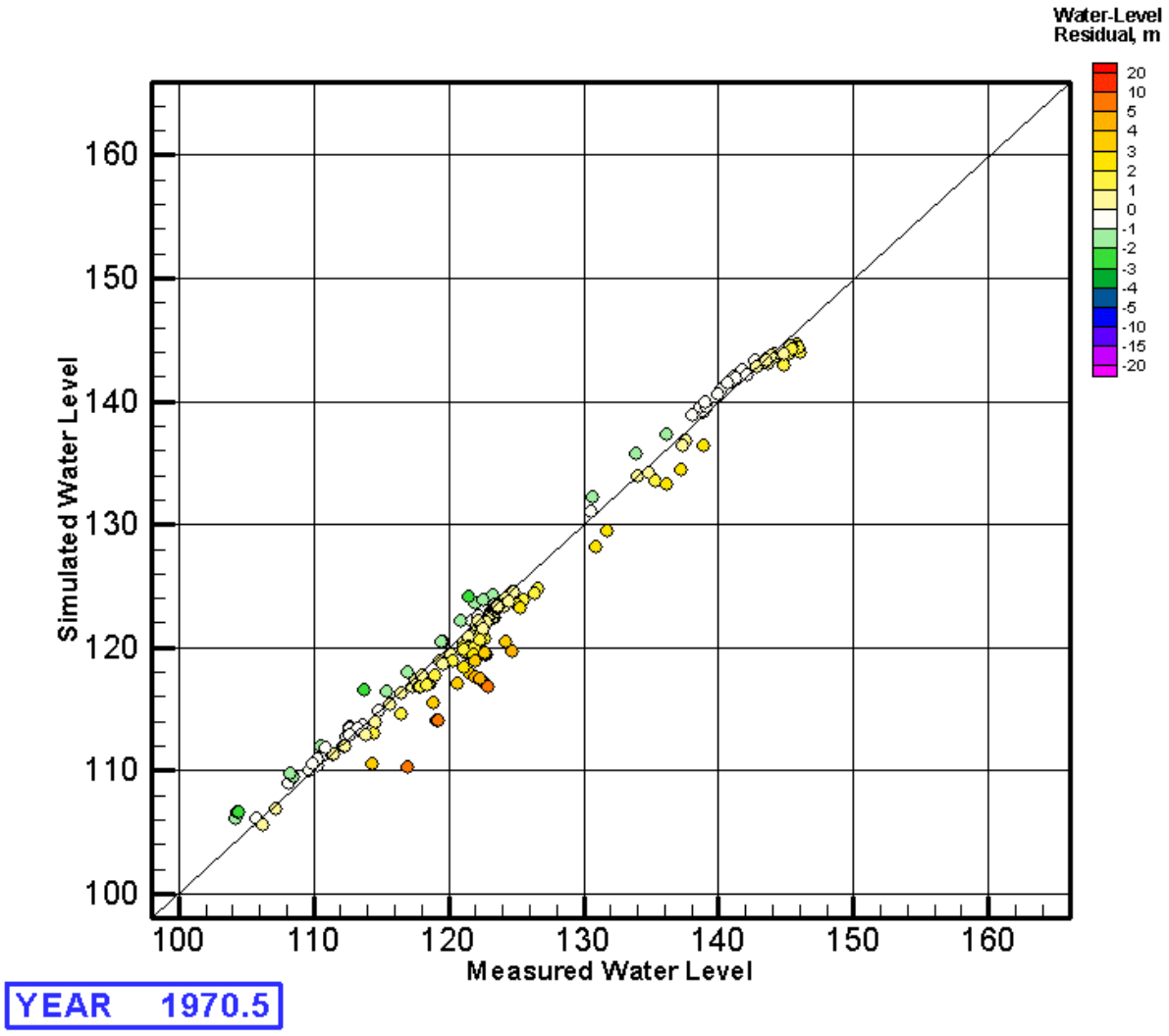

Figure E.3e. Comparison of Simulated versus Measured Water Levels for 1970

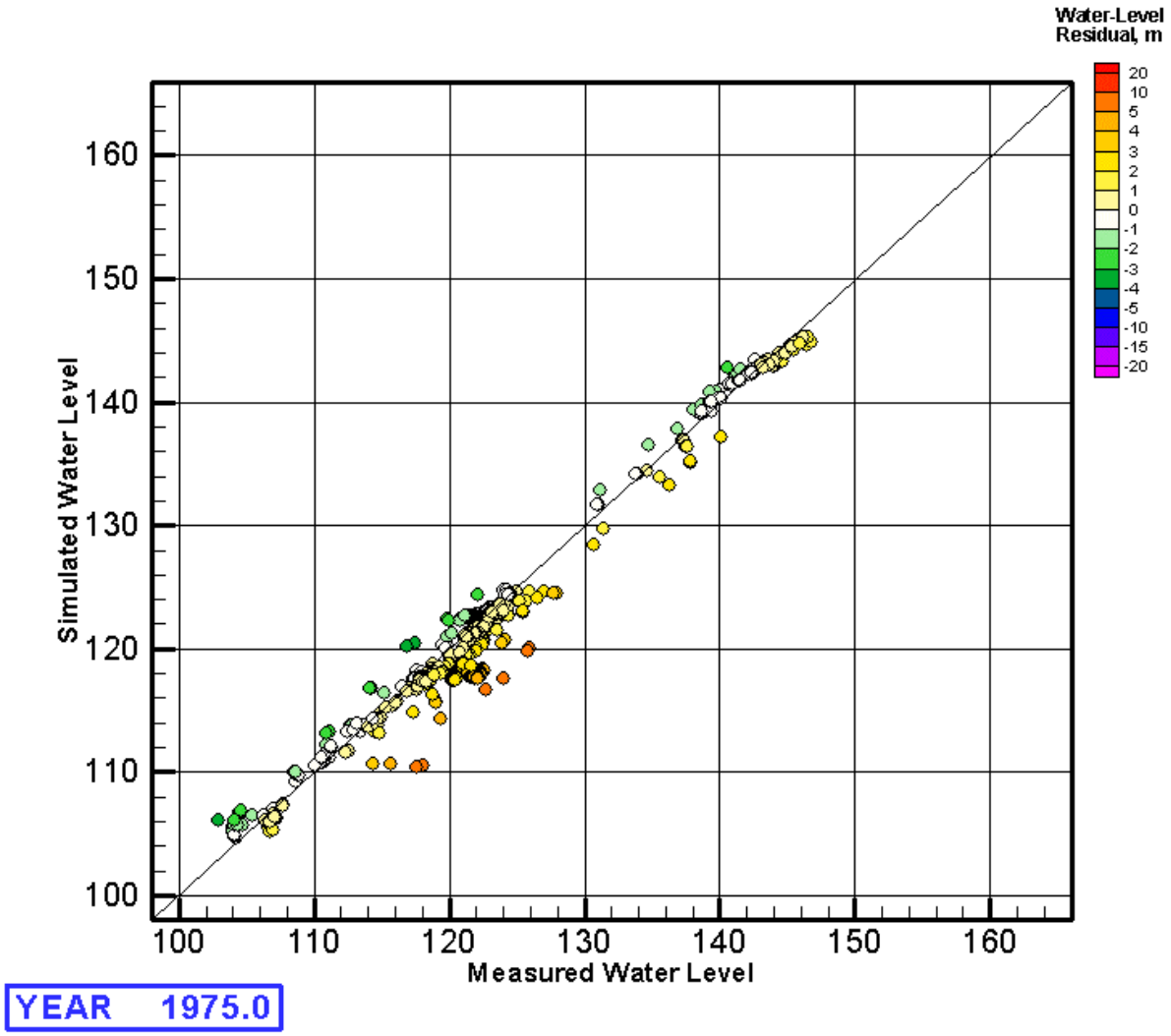

Figure E.3f. Comparison of Simulated versus Measured Water Levels for 1975

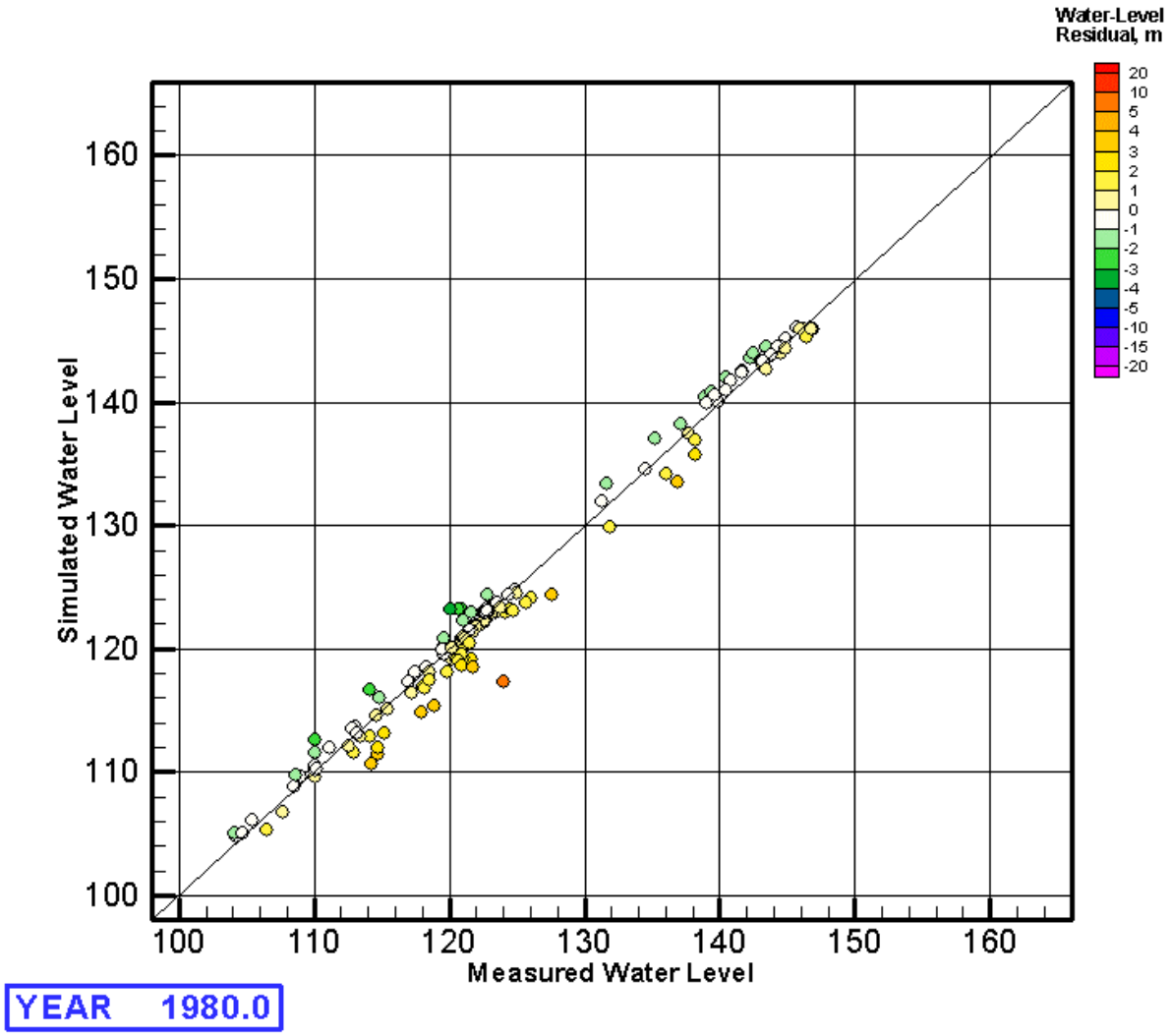

Figure E.3g. Comparison of Simulated versus Measured Water Levels for 1980

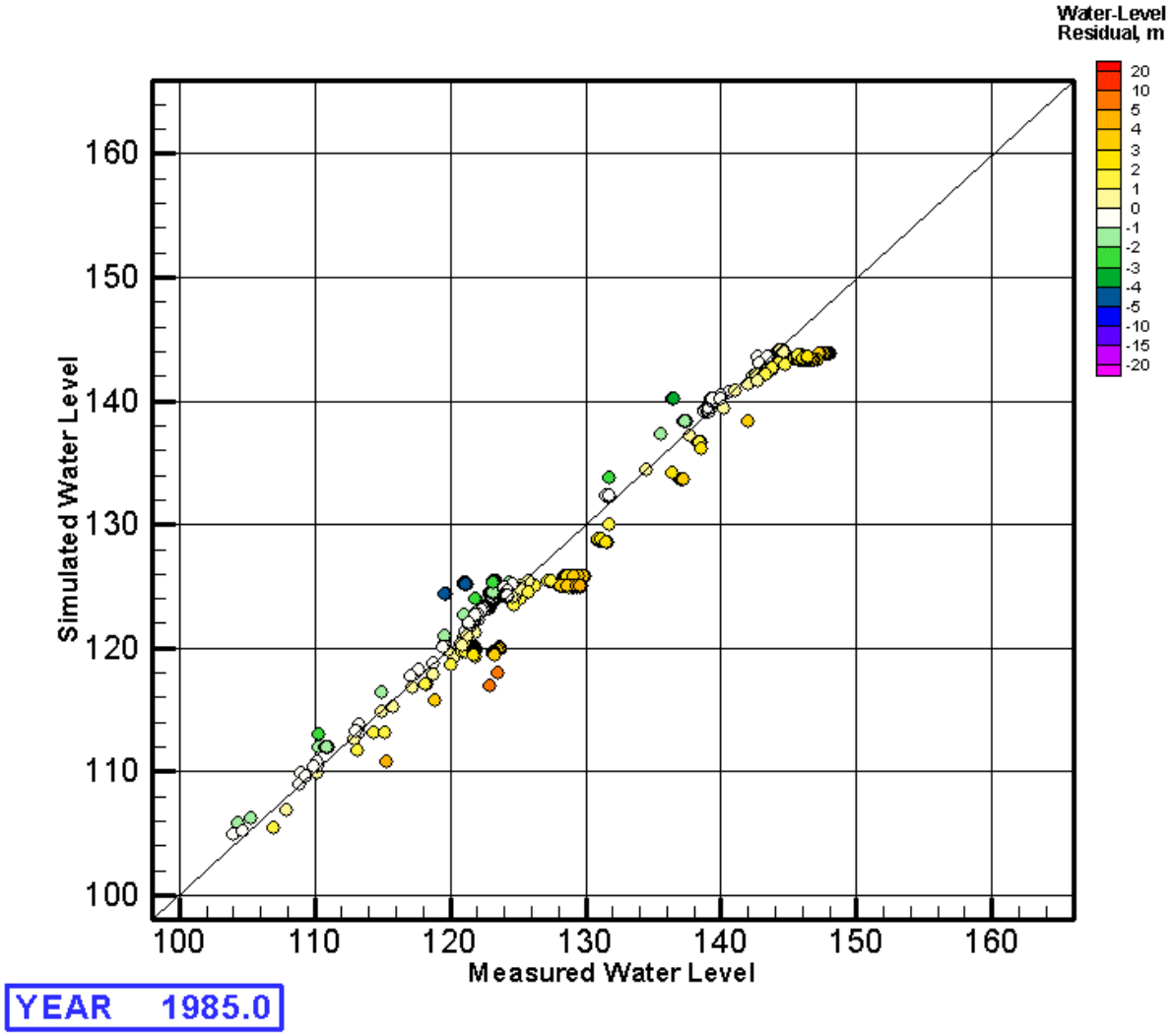

Figure E.3h. Comparison of Simulated versus Measured Water Levels for 1985

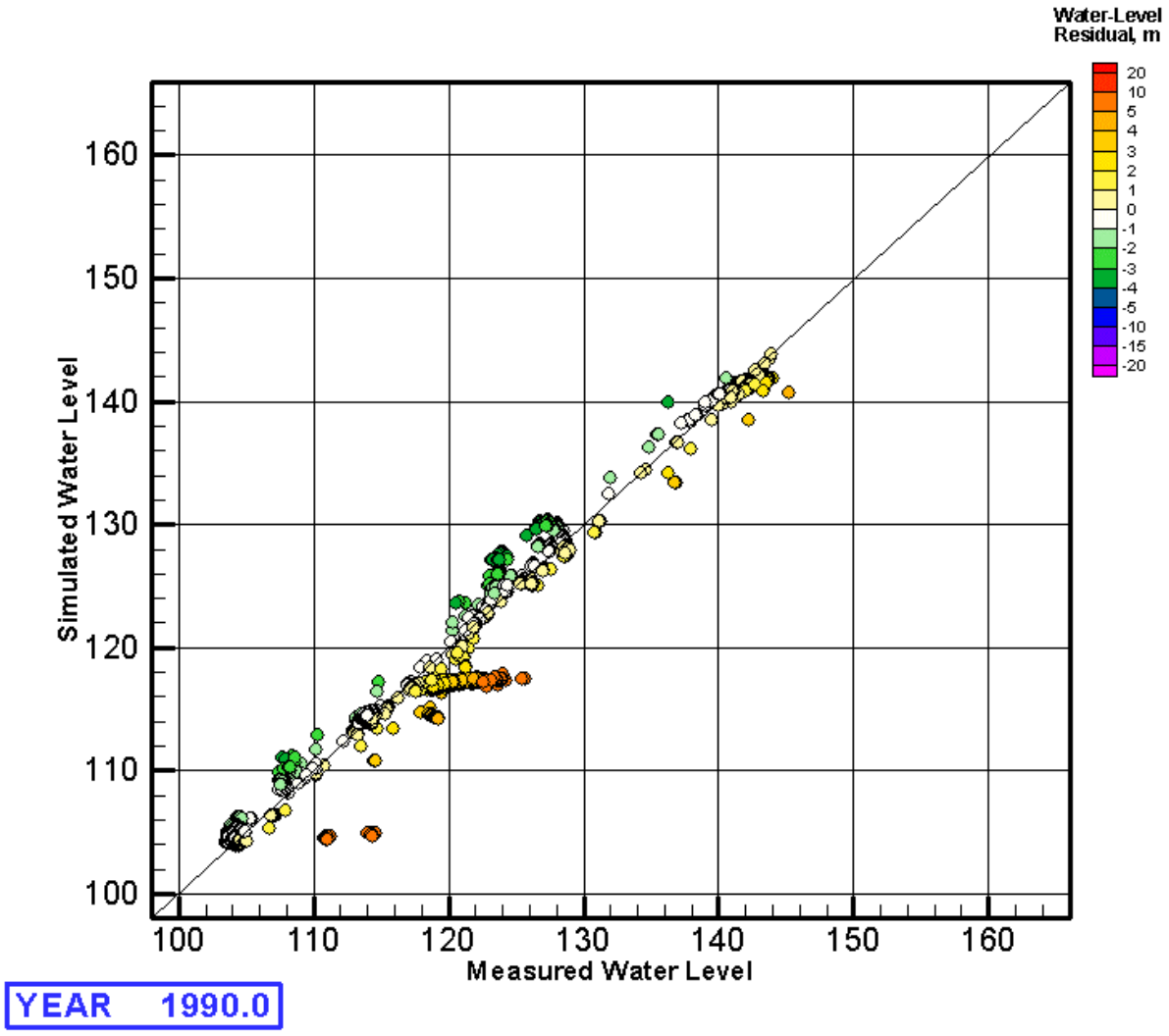

Figure E.3i. Comparison of Simulated versus Measured Water Levels for 1990

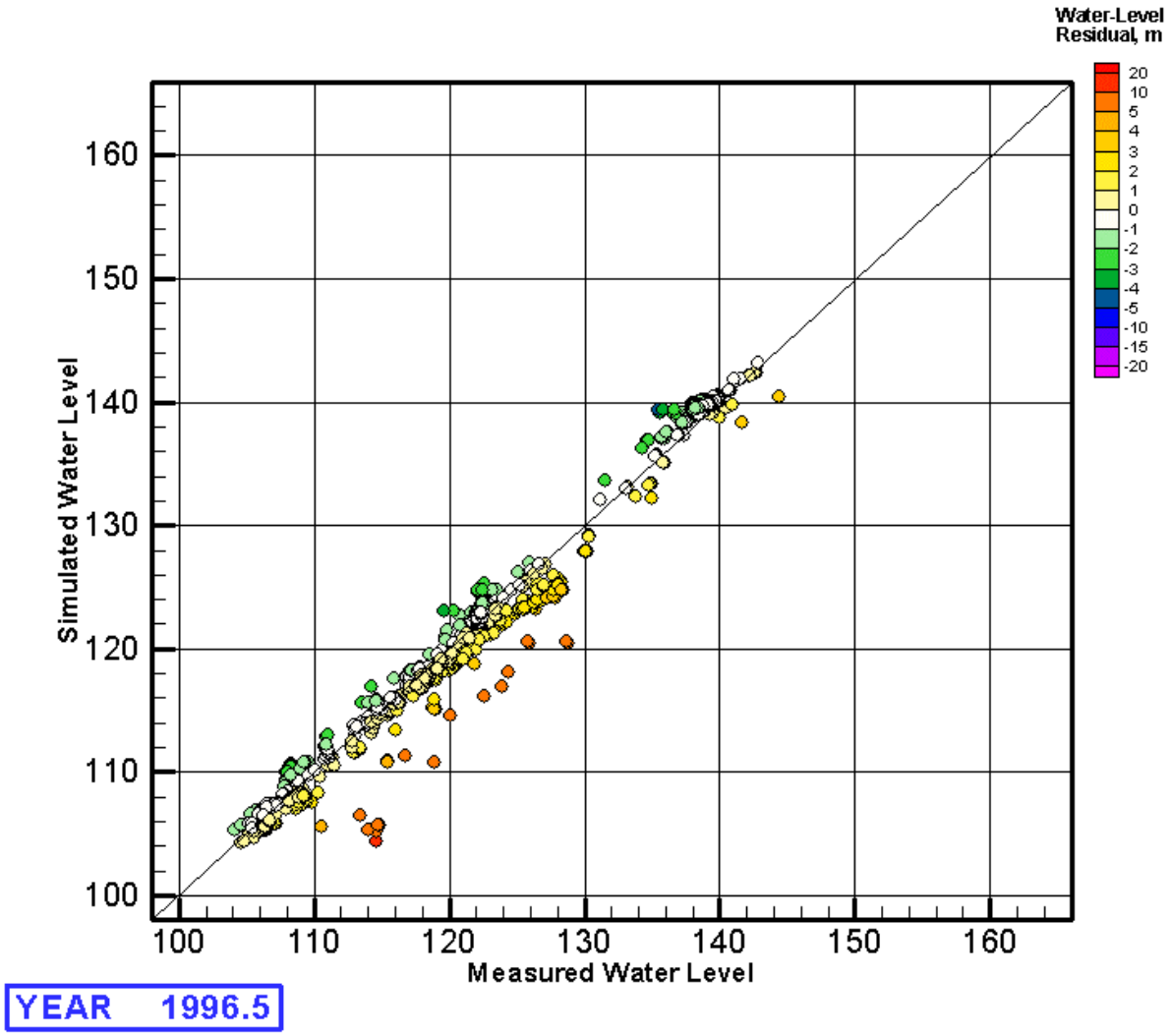

Figure E.3j. Comparison of Simulated versus Measured Water Levels for 1996

Distribution

No. of Copies

OFFSITE

S. Bede
U.S. Ecology
P.O. Box 638
Richland, WA 99352

M. L. Blazek
Oregon Department of Energy
625 Marion St. N.E.
Salem, OR 97310

R. Buck, Jr.
Wanapum Indian Band
P.O. Box 878
Ephrata, WA 98823

2 B. W. Drost
U.S. Geological Survey
1201 Pacific Avenue, Suite 600
Tacoma, WA 98402

3 M. Dunkleman
Department of Health
Division of Radiation Protection
Waste Management Section
P.O. Box 47827
Olympia, WA 98504-7827

2 D. N. Goswami
Washington State Department of Ecology
1315 West 4th
Kennewick, WA 99336-6018

2 S. Harris
Confederated Tribes of the Umatilla
P.O. Box 638
Pendleton, OR 97801

2 Nez Perce Indian Tribe
P.O. Box 365
Lapwai, ID 83540-0365
ATTN: D. L. Powaukee
Judit German Heins

No. of Copies

2 U.S. Department of Energy
Forrestal Building
1000 Independence Avenue, S.W.
Washington, D.C. 20585
ATTN: J. T. Bachmaier
A. Wallo

3 U.S. Department of Energy, Headquarters
19901 Germantown Road
Germantown, MD 20847-1290
ATTN: J. T. Bachmaier
M. K. Harmon
J. E. Rhoderick

2 Yakima Indian Nation
P.O. Box 151
Toppenish, WA 98948
ATTN: R. Jim
W. Riggsbee

ONSITE

5 **DOE Office of River Protection**

C. A. Babel	H6-60
N. R. Brown	H6-60
P. E. Lamont	H6-60
R. W. Lober	H6-60
R. M. Yasek	H6-60

16 **DOE Richland Operations Office**

W. W. Ballard	A5-12
B. E. Bilson	HO-12
P. F. Dunigan	A5-58
B. L. Foley	P7-62
R. F. Guercia	HO-12
R. D. Hildebrand (2)	HO-12
P. M. Knollmeyer	A5-11
J. G. Morse	A5-13
O. Robertson	HO-12

No. of Copies		
	G. H. Sanders	A5-15
	J. P. Sands	H6-12
	T. A. Shrader	H0-12
	A. C. Tortoso	HO-12
	DOE Public Reading Room (2)	H2-53
2	**Fluor Federal Services**	
	R. Khaleel	B4-43
	R. Puigh	B4-43
2	**Fluor Daniel Hanford**	
	D. R. Hall	S8-03
	M. I. Wood	H8-44
3	**CH2M-HILL Group, Inc.**	
	A. J. Knepp	HO-22
	F. M. Mann	H0-22
	D. A. Myers	H0-22
3	**CH2M-HILL Hanford, Inc**	
	W. J. McMahon	H9-03
	V. J. Rohay	H0-19
	C. C. Swanson	H9-02
	Jacobs Engineering	
	P. M. Rogers	H0-22
3	**Bechtel Hanford, Inc.**	
	L. R. Curry	H0-19
	B. H. Ford	HO-21
	M. J. Graham	HO-09
	U.S. Environmental Protection Agency	
	D. R. Sherwood	B5-01

No. of Copies		
68	**Pacific Northwest National Laboratory**	
	B. Barnett	K6-81
	M.P. Bergeron (5)	K9-36
	B. W. Bryce	K6-75
	C. R. Cole (10)	K9-36
	P. E. Dresel	K6-96
	P. W. Eslinger	K3-54
	M. J. Fayer	K9-33
	M. D. Freshley	H0-21
	E. J. Freeman	K9-36
	J. Fruchter	K6-96
	G. W. Gee	K9-33
	G. Guench	K9-33
	M. J. Hartman	K6-96
	F. N. Hodges	K6-81
	V. G. Johnson	K6-96
	C. T. Kincaid	K9-33
	J. W. Lindberg	K6-81
	S. P. Luttrell	K9-96
	M. McKinley	K8-41
	R. B. Mercer	K6-96
	P. D. Meyer	BPO
	C. J. Murray	K6-81
	S. M. Narbutovskih	K6-96
	D. R. Newcomer	K6-96
	W. E. Nichols	K9-33
	M. Oostrum	K9-33
	S. Orr (2)	K9-33
	T. L. Page	K9-18
	B. Peterson	K6-96
	S. Reidel	K6-81
	M. Richmond	K9-33
	T. D. Scheibe	K9-33
	C. S. Simmons	K9-33
	R. M. Smith	K6-96
	M. D. Sweeney	K6-81
	P. D. Thorne (5)	K9-33
	E. C. Thornton	K6-96
	L. W. Vail	K9-33
	W. D. Webber	K6-96
	M. D. White	K9-36
	B. A. Williams	K6-81
	M. D. Williams	K9-36
	S. K. Wurstner (5)	K9-36
	S. B. Yabusaki	K9-36
	Hanford Technical Library (2)	P8-55