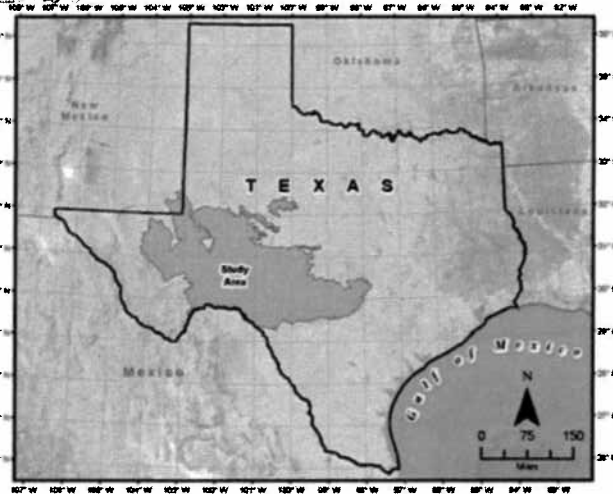
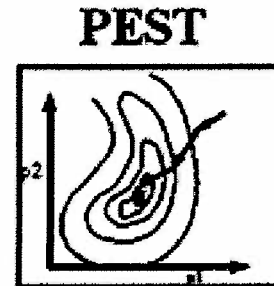
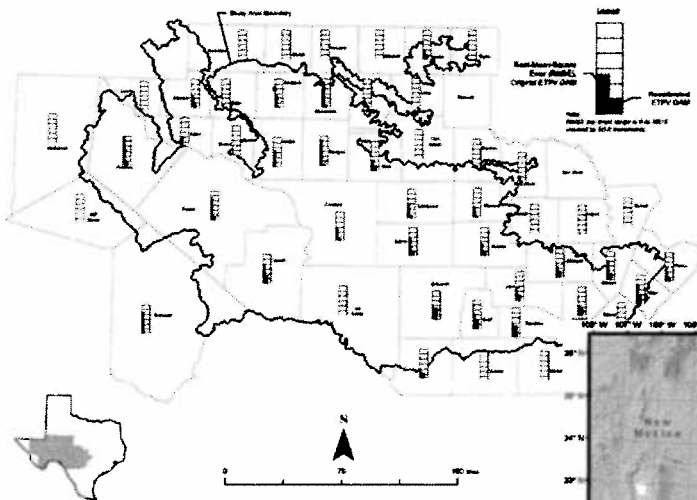


Final

Application of PEST to Re-Calibrate the Groundwater Availability Model for the Edwards-Trinity (Plateau) and Pecos Valley Aquifers



Report ###

by Steven C. Young, Ph.D., P.E., P.G.
John Doherty, Ph.D.
Trevor Budge, Ph.D., EIT
Neil Deeds, Ph.D., P.E.

Texas Water Development Board

P.O. Box 13231, Capitol Station, Austin, TX 78711-3231

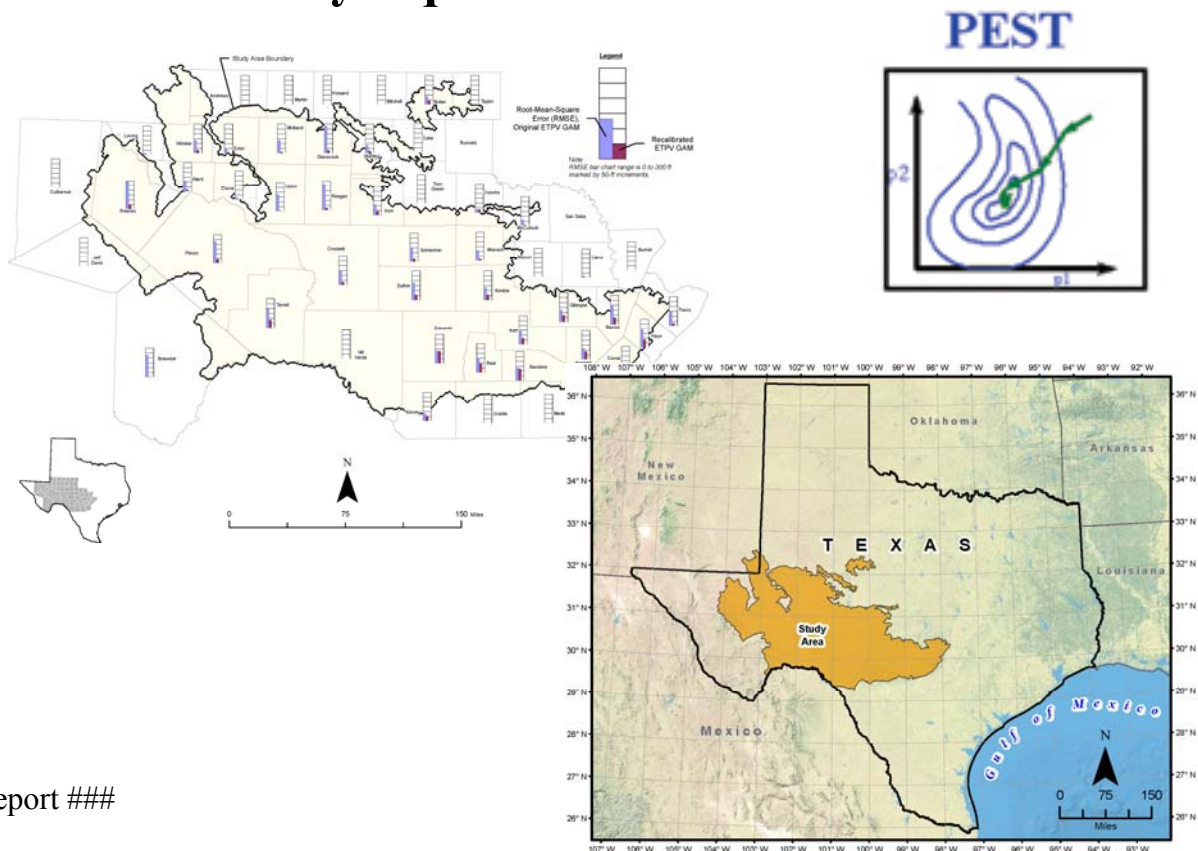
April 2010

2010 APR 14 AM 9:12

ADMINISTRATION

Final

Application of PEST to Re-Calibrate the Groundwater Availability Model for the Edwards-Trinity (Plateau) and Pecos Valley Aquifers



Report ###

by Steven C. Young, Ph.D., P.E., P.G.
John Doherty, Ph.D.
Trevor Budge, Ph.D., EIT
Neil Deeds, Ph.D., P.E.

Texas Water Development Board

P.O. Box 13231, Capitol Station, Austin, TX 78711-3231

April 2010

Texas Water Development Board Report

Final

Application of PEST to Re-Calibrate the Groundwater Availability Model for the Edwards-Trinity (Plateau) and Pecos Valley Aquifers

by Steven C Young, Ph.D., P.E., P.G.
URS Corporation

John Doherty, Ph.D.
Watermark Numerical Pty, Ltd.

Trevor Budge, Ph.D., EIT
URS Corporation

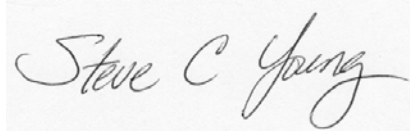
Neil Deeds, Ph.D., P.E.
Intera, Inc.

April 2010

Geoscientist seal

This report documents the work of the following Licensed Geoscientists:

Steve C. Young, P.G.

A handwritten signature in cursive script that reads "Steve C Young".

Dr. Young was the Project Manager for the work and was responsible for the technical oversight on the project and the final model recalibration.

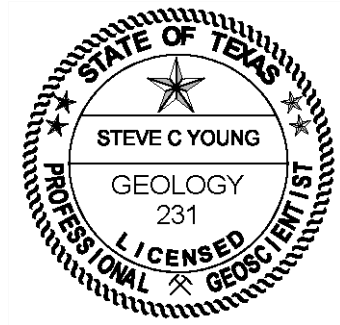


Table of Contents

1	Executive summary.....	1
2	Introduction.....	3
2.1	The Edwards-Trinity (Plateau) and Pecos Valley aquifers	3
2.2	Groundwater availability model (GAM).....	3
2.3	Recalibration of the ETPV GAM.....	3
2.4	Approach to recalibration using PEST.....	4
2.5	Approach to documenting the PEST application	5
3	Original and modified ETPV GAM files and aquifer zones.....	9
3.1	General	9
3.2	Original model.....	9
3.3	Changes to MODFLOW model.....	11
4	Observations and parameters for model calibration	25
4.1	Observation dataset	25
4.2	Parameterization of aquifer properties	27
4.3	Parameterization of recharge.....	32
4.4	Boundary condition parameterization	33
4.5	Parameterization of pumping rates.....	35
5	Calibration specifications.....	53
5.1	Introduction	53
5.2	PEST control file.....	53
5.3	Parameters	54
5.4	Observations.....	58
5.5	PEST input files.....	60
6	Implementation and outcomes	65

TWDB Report ##: Final – Application of PEST to Re-Calibrate
the Groundwater Availability Model for the Edwards-Trinity (Plateau) and Pecos Valley Aquifers

6.1	Implementation.....	65
6.2	Obtaining calibration outcomes.....	66
6.3	Model parameters achieved through model recalibration	69
6.4	Model solutions	71
6.5	Calibration statistics	74
7	Limitations of the model.....	103
7.1	Data and assumptions supporting the model.....	103
7.2	Implementation of model	103
7.3	Model applicability.....	104
7.4	Recommended model improvements	105
8	Summary and conclusions	107
9	References.....	109
	Appendix A Introduction to PEST.....	113
A.1	What is PEST?.....	113
A.2	Model independence.....	115
A.3	Short history of PEST.....	116
A.4	Why highly parameterized inversion?.....	121
	Appendix B The model calibration process.....	125
B.1	Some considerations.....	125
B.2	Practical calibration	127
B.3	Definition of parameters.....	129
	Appendix C Utility software provided with PEST	135
C.1	PEST utilities.....	135
C.2	Groundwater data utilities	139
C.3	Surface water utilities.....	144

Appendix D	Some of the theoretical foundations of PEST	145
D.1	Introduction	145
D.2	Classical parameter estimation	146
D.3	Highly parameterized inversion	150
D.4	Summary	157
Appendix E	File details of ETPV GAM.....	165
E.1	MODFLOW input and output files	165
E.2	Files employed in processing model outputs used in calibration process	166
E.3	Files used in hydraulic property parameterization of revised model	167
E.4	Executable programs used by composite model	171
Appendix F	Listing of wells used for the recalibration.....	173
Appendix G	Program documentation	185
G.1	General	185
G.2	CLIPARRAY1	185
G.3	REAL2CND	186
G.4	REALSTR2CND	187
G.5	STRFLBUD.....	188
G.6	VCONTCALC.....	189
G.7	WELLBUILD2.....	191
G.8	PTINGRID	193
G.9	NONLINOBS	194
Appendix H	The PEST control file.....	195
Appendix I	The model batch file.....	199
Appendix J	The HPC cluster	205
J.1	General hardware description.....	205

J.2 General software description 205

J.3 Running Parallel PEST on the cluster 207

List of Figures

Figure 2-1. Location of the Edwards-Trinity and Pecos Valley aquifers. 7

Figure 2-2. Spatial extents of the Edwards-Trinity (Plateau) and Pecos Valley aquifers and Hill Country part of the Trinity aquifer. 7

Figure 2-3. Conceptual model of the Edwards-Trinity (Plateau) and Pecos Valley aquifers and Hill Country part of the Trinity aquifer. Source: Anaya and Jones, 2009. 8

Figure 2-4. Block diagram of the Edwards-Trinity (Plateau), Trinity (Hill Country), and Pecos Valley aquifers. Source: Anaya and Jones, 2009. 8

Figure 3-1. Contents of name file *eddt_p_tr.nam*. 15

Figure 3-2. Active area of layer 1 of the model domain. Fixed head cells are colored red. 15

Figure 3-3. Active area of layer 2 of the model domain. Fixed head cells are colored blue. 16

Figure 3-4. Recharge distribution for the steady-state model representing predevelopment conditions 16

Figure 3-5. Pumping rate in layer 1 in stress period 21 17

Figure 3-6. Pumping rate in layer 2 in stress period 21 17

Figure 3-7. *General head boundary* cells in layer 1 colored by conductance. 18

Figure 3-8. *General head boundary* cells in layer 2 colored by conductance. 18

Figure 3-9. *Drain* cells in layer 1. Conductance is uniformly 1,000 ft²/day. 19

Figure 3-10. *Drain* cells in layer 2. Conductance is uniformly 1,000 ft²/day. 19

Figure 3-11. *Stream* cells in both layers, colored by conductance value. 20

Figure 3-12. Hydraulic conductivity in layer 1. The same zonation as that depicted in this figure is employed in the recalibration process described below; zone numbers are provided. 20

Figure 3-13. Horizontal conductivity in layer 2. The same zonation as that depicted in this figure is employed in the recalibration process described below; zone numbers are provided. 21

Figure 3-14. Log (to base 10) of vertical conductance between layers 1 and 2; greener indicates higher conductance. Log of conductance ranges between -8 and -4.6. 21

Figure 3-15. Specific storage assigned to layer 1 22

Figure 3-16. Specific storage assigned to layer 2 22

Figure 3-17. Specific yield assigned to layer 1 23

Figure 3-18. Specific yield assigned to layer 2 23

Figure 3-19. Name file *eddt_p_ss.nam* for steady state component of model. 24

Figure 3-20. Name file *eddt_p_tr.nam* for transient component of model. 24

Figure 4-1. Wells in layer 1 for which head observations were employed in the calibration process. 41

Figure 4-2. Wells in layer 2 for which head observations were employed in the calibration process. 41

Figure 4-3. Stream reaches (light blue) and streamflow measurement points (red) featured in the calibration dataset. 42

TWDB Report ##: Final – Application of PEST to Re-Calibrate
the Groundwater Availability Model for the Edwards-Trinity (Plateau) and Pecos Valley Aquifers

Figure 4-4.	Distribution of pumping test locations used in hydraulic property assignment of layer 1.	42
Figure 4-5.	Distribution of pumping test locations used in hydraulic property assignment of layer 2.	43
Figure 4-6.	Cumulative distribution for conductivity values in each zone.	43
Figure 4-7.	Distribution of hydraulic property pilot points used for parameterization of layer 1.	44
Figure 4-8.	Distribution of hydraulic property pilot points used for parameterization of layer 2.	44
Figure 4-9.	Recharge zonation and zone numbering scheme.	45
Figure 4-10.	Pilot points used in recharge parameterization.	45
Figure 4-11.	General head boundary pilot points in layer 1. Preferred conductances in ft ² /day are also depicted.	46
Figure 4-12.	General head boundary pilot points in layer 2. Preferred conductances in ft ² /day are also depicted.	46
Figure 4-13.	Pilot points used in stream conductance parameterization.	47
Figure 4-14.	Zones and zone numbers used in estimation of drain conductance for layer 1. ...	47
Figure 4-15.	Zones and zone numbers used in estimation of drain conductance for layer 2. ...	48
Figure 4-16.	“Standard” cell pumping rates in layer 1 used in five precalibration, 10-year stress periods.	48
Figure 4-17.	“Standard” cell pumping rates in layer 2 used in five precalibration, 10-year stress periods.	49
Figure 4-18.	Irrigated acres reported by the U.S. Agricultural Census from 1930 to 1964 and by TWDB from 1958 to 2000 for Glasscock, Midland, Pecos, and Reagan Counties.	49
Figure 4-19.	Irrigated acres reported by the U.S. Agricultural Census from 1930 to 1964 and by TWDB from 1958 to 2000 for Reeves, Schleicher, and Ward Counties. .	50
Figure 4-20.	Estimated pumping from 1930 to 2000 for Glasscock, Midland, Pecos, and Reagan Counties based on a merger of U.S. Agricultural survey data and TWDB irrigation survey data.	50
Figure 4-21.	Estimated pumping from 1930 to 2000 for Reeves, Schleicher, and Ward Counties based on a merger of U.S. Agricultural survey data and TWDB irrigation survey data.	51
Figure 4-22.	Zone definition for pumping rate build-up factor parameters.	51
Figure 6-1.	Parallel PEST run management file.	81
Figure 6-2.	Hydraulic conductivity in layer 1 from the PEST recalibration.	81
Figure 6-3.	Hydraulic conductivity in layer 2 from the PEST recalibration.	82
Figure 6-4.	Log (to base 10) of vertical conductance between layers 1 and 2 from the PEST recalibration.	82
Figure 6-5.	Specific storage in layer 1 from the PEST recalibration.	83
Figure 6-6.	Specific storage in layer 2 from the PEST recalibration.	83
Figure 6-7.	Specific yield in layer 1 from the PEST recalibration.	84
Figure 6-8.	Specific yield in layer 2 from the PEST recalibration.	84
Figure 6-9.	Recharge distribution for 1980.	85
Figure 6-10.	Conductance values (ft ² /day) for general head boundary cells in layer 1 from the PEST recalibration.	85

Figure 6-11.	Conductance values (ft ² /day) for general head boundary cells in layer 2 from the PEST recalibration.	86
Figure 6-12.	The conductance values (ft ² /day) for streams in layer 1 from the PEST recalibration.	86
Figure 6-13.	The conductance values (ft ² /day) for streams in layer 2 from the PEST recalibration.	87
Figure 6-14.	Simulated water levels for 1930 (presumed predevelopment conditions) in layer 1 from the PEST recalibration.	87
Figure 6-15.	Simulated water levels for 1930 (presumed predevelopment conditions) in layer 2 from the PEST recalibration.	88
Figure 6-16.	Water budget for the steady-state model for 1930. Inflows are shown in (a), and outflows are shown in (b).....	88
Figure 6-17.	Simulated water levels for 1980 in layer 1 from the PEST recalibration.	89
Figure 6-18.	Simulated water levels for 1980 in layer 2 from the PEST recalibration.	89
Figure 6-19.	Simulated water levels for 2000 in layer 1 from the PEST recalibration.	90
Figure 6-20.	Simulated water levels for 2000 in layer 2 from the PEST recalibration.	90
Figure 6-21.	Simulated water levels for 2000 in layer 1 from the ETPV GAM. Source: Anaya and Jones, 2009.	91
Figure 6-22.	Simulated water levels for 2000 in layer 2 from the ETPV GAM. Source: Anaya and Jones, 2009.	91
Figure 6-23.	Average annual water budget for the 1980 to 2000 simulation for the original and recalibrated ETPV GAM (inflow is on left and outflow is on right).	92
Figure 6-24.	Comparison of observed versus simulated water levels for layers 1 and 2 from the original ETPV GAM.	93
Figure 6-25.	Comparison of observed versus simulated water levels for layers 1 and 2 from the PEST recalibration of the ETPV GAM.....	94
Figure 6-26.	Comparison of RMSE calculated for the wells in each county for the original and the recalibrated model.....	95
Figure 6-27.	Calculated RMSE (ft) at each well location in layer 1 for the recalibrated model.....	96
Figure 6-28.	Calculated RMSE (ft) at each well location in layer 2 for the recalibrated model.....	97
Figure 6-29.	Comparison of groundwater contribution to stream flow from the original ETPV GAM (dashed black line) and the recalibrated ETPV GAM (solid grey line) and the river flow estimated from the cited river gage.....	98
Figure B-1.	The effect of history-matching as expressed by Bayes' equation.	133
Figure D-1.	Parameter estimation as a projection operation.	159
Figure D-2.	Components of parameter error.	159
Figure D-3.	Null space Monte Carlo generation of calibration constrained random fields. ..	160
Figure D-4.	Heterogeneous hydraulic property field within a model domain. Historical head measurements are available at wells located within the domain.....	160
Figure D-5.	Calibration based on zones of piecewise constancy.	161
Figure D-6.	Pilot-point-based calibration. Pilot points are shown in grey.....	161
Figure D-7.	Types of mathematical regularization used by PEST.	162

Figure D-8.	Use of super parameters to enhance the efficiency of the parameter estimation process.	162
Figure D-9.	The null-space Monte Carlo methodology can be used to compute many different parameter fields which all calibrate the model.	163
Figure G-1.	Example of an integer-limit-correspondence file.	186
Figure G-2.	An example STRFLBUD input file.	188
Figure G-3.	An example of a well stress period data file.	192
Figure G-4.	Example of a build-up factor file.	193
Figure J-1.	The cluster and its basic cabling.	209

List of Tables

Table 4-1.	Well locations used for the model calibration from 1980 to 2000.	25
Table 4-2.	Selected statistical values for the hydraulic conductivity estimates in each aquifer zone based on the cumulative distributions shown on Figure 4-6 (all values in ft/day).	28
Table 4-3.	Preferred values and bounds on horizontal hydraulic conductivity imposed during calibration (all values in ft/day).	29
Table 4-4.	Cut-off limits imposed on the hydraulic conductivity values assigned to each grid cell (all values in ft/day).	30
Table 4-5.	Preferred values and bounds on specific storage imposed during calibration (all values in ft ⁻¹).	30
Table 4-6.	Preferred values and bounds on specific yield imposed during calibration (all values are dimensionless).	31
Table 4-7.	Cut-off limits imposed on storage parameters assigned to each grid cell.	31
Table 4-8.	Preferred recharge factors and bounds for zones illustrated on Figure 4-9.	32
Table 4-9.	Summary of irrigation pumping for the model domain and the eight counties with the highest irrigation based on the TWDB surveys.	36
Table 4-10.	Application rates (acre-ft/acre) from the TWDB irrigation surveys.	37
Table 4-11.	Preferred values for pumping build-up factor parameters.	38
Table 5-1.	Names of sections that may be present within a PEST control file.	53
Table 5-2.	Zone-based aquifer hydraulic property parameters.	54
Table 5-3.	Details of template files cited in PEST control file used in calibration of the ETPV GAM.	60
Table 5-4.	Details of instruction files cited in PEST control file used in calibration of ETPV GAM.	61
Table 6-1.	Groups that contribute to the value of the objective function phi.	67
Table 6-2.	Some files generated by PEST; <i>casename</i> is the filename base of the PEST control file.	68
Table 6-3.	Summary statistics for aquifer parameters in the recalibrated ETPV GAM.	70
Table 6-4.	Preferred recharge factors and bounds for zones illustrated on Figure 6-9.	70
Table 6-5.	Water balance for steady-state recalibrated ETPV GAM for predevelopment conditions in 1930.	72
Table 6-6.	Water balance for the original ETPV GAM for transient conditions from 1980 to 2000.	73

TWDB Report ##: Final – Application of PEST to Re-Calibrate
the Groundwater Availability Model for the Edwards-Trinity (Plateau) and Pecos Valley Aquifers

Table 6-7.	Water balance for the steady-state recalibrated ETPV GAM for transient conditions from 1980 to 2000.	73
Table 6-8.	Water balance for layers 1 and 2 for the original and the recalibrated GAM for transient conditions from 1980 to 2000.	73
Table 6-9.	Calibration statistics for layer 1 for the original and recalibrated GAMs for transient conditions from 1980 to 2000.	75
Table 6-10.	Calibration statistics for layer 2 for the original and recalibrated GAMs for transient conditions from 1980 to 2000.	76
Table 6-11.	Calibration statistics for layers 1 and 2 for the original and recalibrated GAMs for transient conditions from 1980 to 2000.	77
Table 6-12.	Water balance for layers 1 and 2 for the original and recalibrated GAMs for transient conditions from 1980 to 2000.	79

1 Executive summary

The primary objective of this project was to recalibrate the Edwards-Trinity (Plateau) and Pecos Valley (ETPV) Groundwater Availability Model (GAM) using PEST and high-performance computing (HPC) clusters. At the onset of the project, a specific goal of the project was to improve the model calibration in Upton and Reagan Counties. PEST is a computer program that provides the capability to semi-automatically calibrate groundwater models and to integrate new information into the calibration process in a more robust and cost-effective manner than traditional trial-and-error manual approaches allow. An HPC cluster is a cluster that permits a single PEST application to run across a network of linked computers so that calibration runtimes can be greatly reduced.

The recalibration of the ETPV GAM includes matching 4,773 water level measurements in model layers 1 and 2 from the years 1980 to 2000. For the 4,773 water level measurements, the original GAM produces a root-mean square error (RMSE) of about 180 ft, and the recalibrated GAM produces a RMSE of about 60 ft. (The RMSE represents a measure of the average difference between measured and simulated water levels.) Compared to the original GAM calibration, the PEST-based GAM calibration provides significantly better model calibration statistics. Across the model domain, the recalibrated GAM reduces the RMSE by 122 ft, the mean error by 66 ft, and the absolute error by 106 feet. Across specific counties, the recalibrated GAM reduces the RMSE by more than a factor of 2 and 3 in 26 and 19 counties, respectively, and by more than 50 ft and 100 ft in 24 and 15 counties, respectively. The original ETPV has RMSEs of 176 ft and 269 feet for Upton and Reagan Counties, respectively. The recalibrated ETPV GAM has RMSEs of 9 ft and 23 feet for Upton and Reagan Counties, respectively.

Compared to the traditional practices of model calibration involving manual trial-and-error approaches, PEST provides a significantly improved methodology for determining optimal parameter values, investigating alternative data interpretations, and performing predictive uncertainty analyses. Over the long-term development of a model, which may involve several groups of modelers and several updates of data, the PEST has the potential to improve how GAMs are calibrated and applied.

From 1980 to 2000, the recalibrated ETPV GAM averages about 1.1 million acre-feet/year (AFY) of recharge compared to about 1.2 million AFY of recharge for the original ETPV GAM. The recalibrated ETPV GAM simulates about 1.8 million AFY of average total flow through the model domain compared to 1.7 million AFY in the original ETPV GAM. For the recalibrated GAM, the sources of groundwater include recharge (62%), storage (26%), rivers (8%), reservoirs (3%), and adjacent aquifers (1%). The mechanisms for groundwater discharge in the recalibrated GAM include rivers (27%), storage (25%), wells (19%), adjacent aquifers (14%), springs/seeps (11%), and reservoirs (4%).

All of the future model improvements and limitations discussed by Anaya and Jones (2009) regarding the original ETPV GAM equally apply to the recalibrated ETPV GAM. Among their key points are identifying benefits associated with collecting additional field data to support the model development, subdividing the ETPV GAM into two or more smaller regional models, and using a modeling approach that allows transmissivity to vary with saturated thickness.

TWDB Report ##: Final – Application of PEST to Re-Calibrate
the Groundwater Availability Model for the Edwards-Trinity (Plateau) and Pecos Valley Aquifers

2 Introduction

2.1 The Edwards-Trinity (Plateau) and Pecos Valley aquifers

As defined by Anaya and Jones (2009), the Edwards-Trinity (Plateau) and the Pecos Valley (ETPV) aquifers occupy an area of about 44,000 square miles of west central Texas, as shown on Figure 2-1. The Edwards-Trinity (Plateau) aquifer extends over an area of about 35,000 square miles beneath all or parts of 39 counties (Ashworth and Hopkins, 1995), and the Pecos Valley aquifer extends over an area of about 7,000 square miles beneath all or parts of 11 counties, of which some are in New Mexico (Figure 2-2).

2.2 Groundwater availability model (GAM)

To better understand groundwater flow in the ETPV aquifers, the Texas Water Development Board (TWDB) developed and calibrated a groundwater availability model (GAM) for these aquifers (Anaya and Jones, 2009). A GAM is a three-dimensional, numerical groundwater flow model capable of simulating regional-scale groundwater flow systems. The Texas state legislature mandated the development of these state-of-the-art, computer-based models for all of the major and minor aquifers in Texas. In 2005, House Bill 1763 mandated that groundwater conservation districts evaluate and develop the desired future conditions for aquifers within their groundwater management areas.

The EPTV GAM (Anaya and Jones, 2009) is based on the MODFLOW (McDonald and Harbaugh, 1988) groundwater code that simulates groundwater flow in the Pecos Valley, the Edwards-Trinity (Plateau), and the Hill County Trinity aquifers (Figures 2-3 and 2-4). The ETPV GAM consists of three layers that contain 400 columns and 300 rows. Each grid cell represents a square with an area of 1 square mile. The top layer, layer 1, is composed of 32,066 active cells to model the Pecos Valley aquifer and the Edwards hydrostratigraphic unit. The middle layer, layer 2, is composed of 31,332 active cells to model the Trinity hydrostratigraphic unit. The lower-most layer, layer 3, is an inactive layer that may be used to model interactions with the underlying aquifers. Henceforth, layer 3 is not discussed.

The ETPV GAM is comprised of a steady-state model and a transient model. The steady-state model reproduces the water levels for 1980. The transient model simulates the period from 1980 through 2000. The two model runs are coupled such that the water levels simulated by the steady-state model are used as the initial water levels for the transient simulations.

2.3 Recalibration of the ETPV GAM

The TWDB GAM program aims to produce state-of-the-art models that represent the best science available for predicting groundwater-level response to future changes in pumping and climatic conditions. The GAM program considers groundwater models to be works in progress, and therefore the program includes updates to the GAMs as new information and technologies become available. As part of a continual effort to improve the GAM program, the TWDB has identified PEST and high-performance computing (HPC) clusters as attractive options for improving the current TWDB capabilities and for updating and recalibrating GAMs.

PEST is a computer program that provides the capability to semi-automatically calibrate groundwater models and to integrate new information into the calibration process in a more robust and cost-effective manner than the traditional manual approaches allow. Initially, PEST was primarily suitable for relatively small-sized models, but recent improvements in the code have made it a viable option for use with larger and more complicated models such as the GAMs. One recent improvement is the capability to seamlessly use PEST in a parallel computing mode where PEST is run across a network of computers known as an HPC cluster. Some of the potential benefits that PEST offers compared to the traditional manual trial-and-error method are a reduction in time to develop, manage, and review model calibrations; more robust and credible model calibrations; and improved capabilities to quantify predictive uncertainty.

To evaluate the potential for PEST and HPC clusters to benefit the GAM program, the TWDB reviewed existing GAMs and selected the ETPV GAM as the best candidate for a demonstration project. Some of the reasons for selecting this GAM are that it is mathematically stable and the area around Upton and Reagan Counties would benefit greatly from additional model parameter adjustments to better simulate historical water levels.

2.4 Approach to recalibration using PEST

The primary objective of this project is to recalibrate the ETPV GAM using PEST techniques with a HPC cluster to determine the feasibility of using this approach and equipment in the groundwater availability and modeling program. To accomplish this objective, the authors retained the site conceptual model, the numerical grid, and locations of boundary conditions developed by Anaya and Jones (2009) for the ETPV GAM.

Because the current project focused on improving the mathematical and technical aspects of the model recalibration, this project accepted the limitations of the ETPV GAM discussed by Anaya and Jones (2009). Most of these limitations are associated with the lack of data available to characterize the stratigraphy, hydraulic boundaries, and recharge relationships for the different aquifers. A few of the limitations are associated with the problems and limitations of MODFLOW-96 (Harbaugh and McDonald, 1996), which was used to develop the ETPV GAM. One of the problems with MODFLOW-96 is its limited capability to accurately simulate the drying and resaturation of grid cells. To work around this limitation, Anaya and Jones (2009) modeled the groundwater system by fixing the transmissivity of the aquifer at each grid cell so that the transmissivity does not change with changes in the computed water level.

In addition to the incorporation of PEST, the authors investigated several options for improving the value and results of some outcomes of the recalibration process. This investigation promulgated two changes in the model recalibration process. One change involved using additional hydrographs in the model calibration. A hydrograph is comprised of measured water levels over time from a single well. The calibration data set used by Anaya and Jones (2009) for the transient model for the ETPV GAM included 10 hydrographs containing at least three water level measurements. The recalibration of the transient model using PEST included over 500 such hydrographs. A benefit of including more hydrographs in the calibration is that the overall

reliability of the model is improved and there is a greater likelihood of identifying where the uncertainty or inaccuracies with the model predictions may be the greatest. The inclusion of over 500 hydrographs was accomplished by inserting the digital information into the PEST control file. The PEST control file is the main input file to PEST and its structure is explained in detail in a later report section. As will be discussed in a subsequent report section, the inclusion of numerous hydrographs into a PEST calibration run incurs no logistical, numerical, or computational penalties.

Another change to the model calibration process is the manner in which the 1980 groundwater levels are developed to begin the transient model simulation. Our approach replaced the 1980 steady-state model with a 1930 steady-state model and a 1930-1980 transient model. The 1930 steady-state model includes no groundwater pumping and produces water levels to start the 1930-1980 transient model. The pumping rates for the 1930 to 1980 transient model were developed using historical values for irrigation acreages and application rates in combination with PEST's parameter estimation capabilities. As discussed later in this report, the irrigation data from several western counties indicate that pumping rates in 1980 may be 20% or less of their peak values in the 1940's and 1950's. Because groundwater levels may require decades to fully recover from large drawdowns caused by pumping, our approach provides a mechanism to help account for how historical pumping may have impacted 1980 water levels.

2.5 Approach to documenting the PEST application

The report is focused on describing the key steps and decisions associated with the PEST application. Given the technical nature of parameter estimation process, several of the sections discuss technical issues. For instance, to explain some of the fundamental principles inherent in the development and application of PEST, the report discusses technical points such as Bayes' theorem, highly parameterized inversion, pilot points, and mathematical regularization. Most groundwater modelers do not discuss these terms outside of graduate class, but they are nevertheless discussed in this report because of their importance to understanding the development and application of PEST.

With regard to documenting the PEST application, the report begins with the ETPV GAM MODFLOW-96 model files and then describes how the files were coupled to the PEST files. The report focuses on describing the type of information needed to set up and run a PEST file and minimizes discussions on how to construct the file. The details associated with PEST set up and file construction are available from the PEST manual (Doherty, 2009a; Doherty, 2009b). For those readers unfamiliar with MODFLOW-96 or parameter estimation techniques, the authors recommend that these readers complement their reading of this report with the documentation for MODFLOW-96 and PEST.

TWDB Report ##: Final – Application of PEST to Re-Calibrate
the Groundwater Availability Model for the Edwards-Trinity (Plateau) and Pecos Valley Aquifers

TWDB Report ##: Final – Application of PEST to Re-Calibrate the Groundwater Availability Model for the Edwards-Trinity (Plateau) and Pecos Valley Aquifers

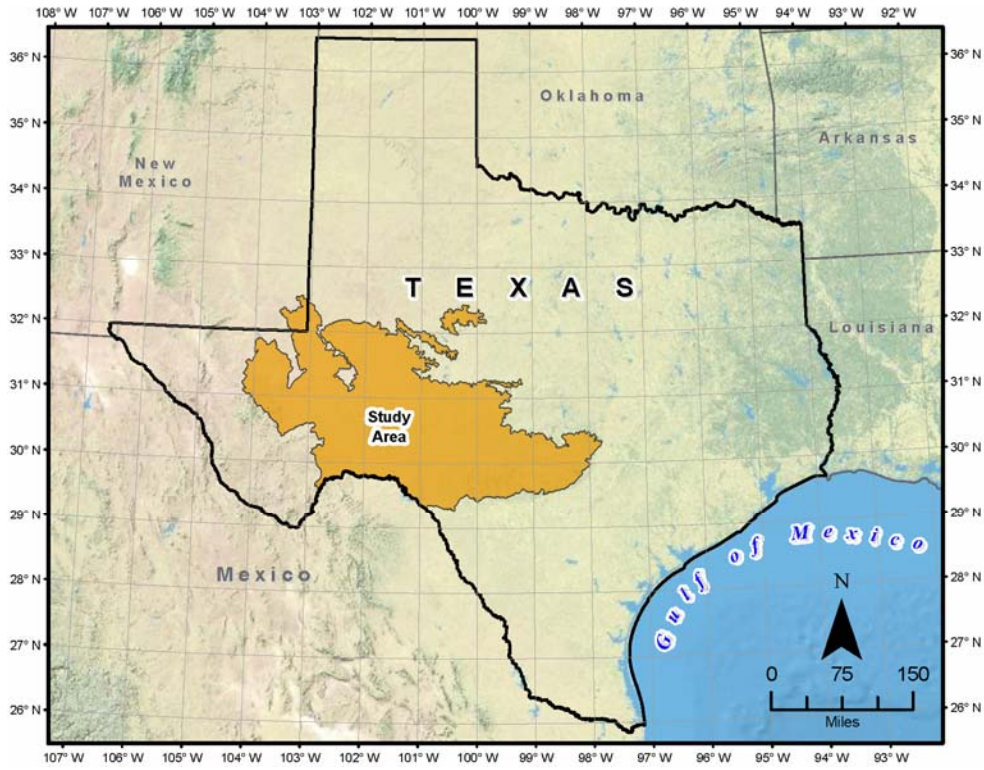


Figure 2-1. Location of the Edwards-Trinity and Pecos Valley aquifers.

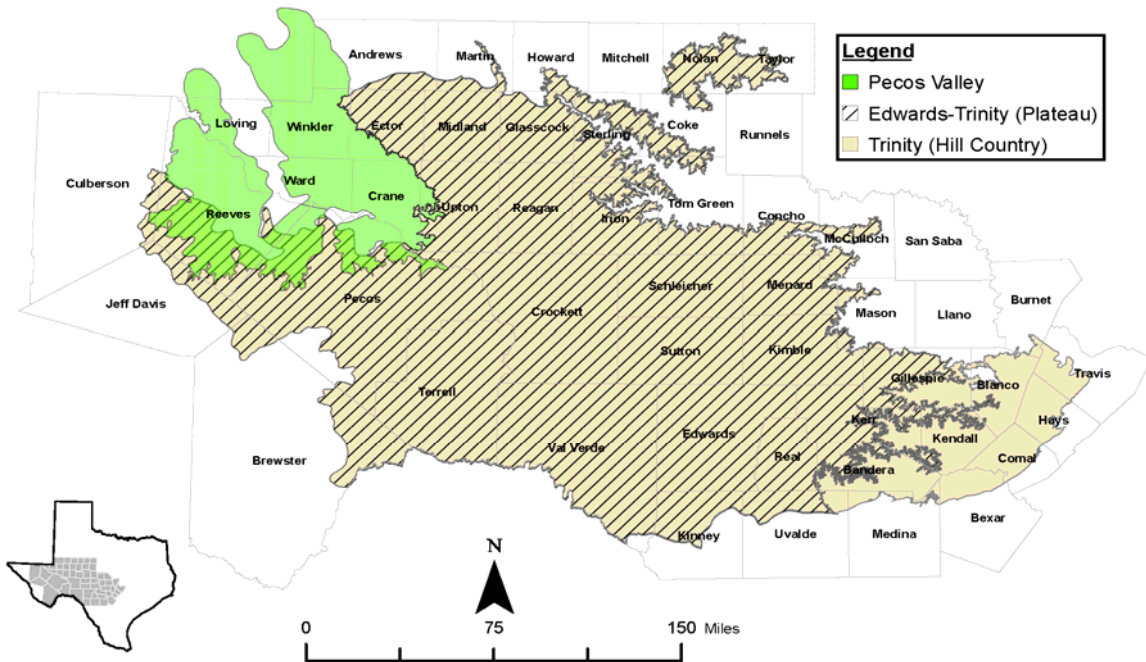


Figure 2-2. Spatial extents of the Edwards-Trinity (Plateau) and Pecos Valley aquifers and Hill Country part of the Trinity aquifer.

TWDB Report ##: Final – Application of PEST to Re-Calibrate the Groundwater Availability Model for the Edwards-Trinity (Plateau) and Pecos Valley Aquifers

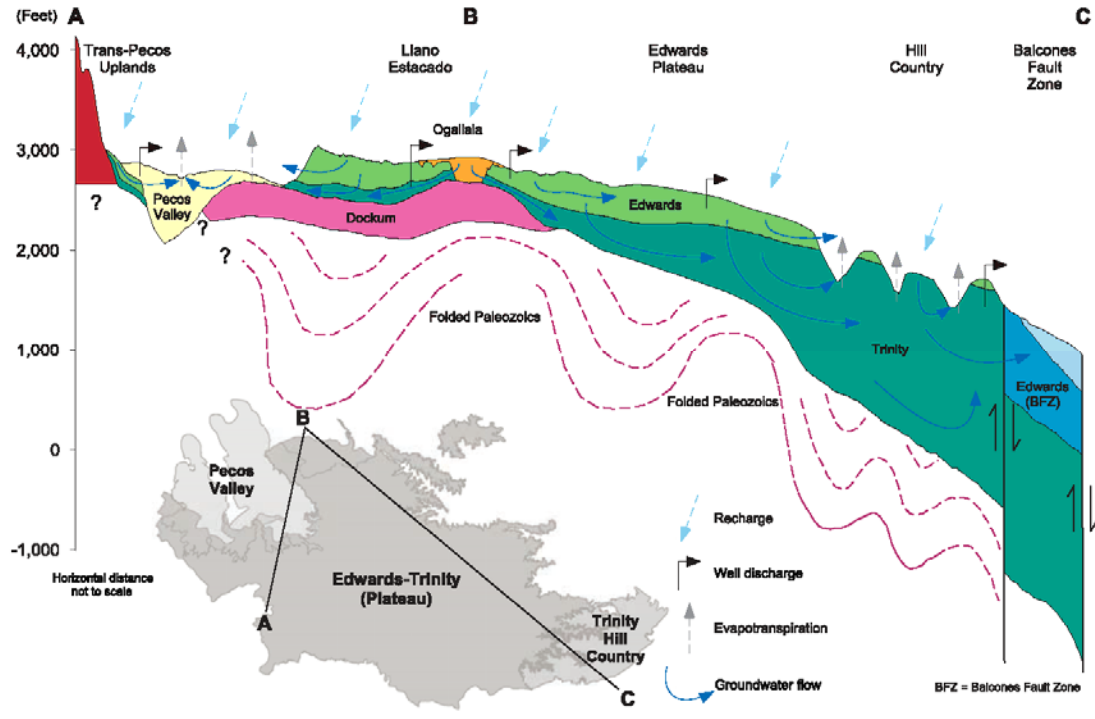


Figure 2-3. Conceptual model of the Edwards-Trinity (Plateau) and Pecos Valley aquifers and Hill Country part of the Trinity aquifer. Source: Anaya and Jones, 2009.

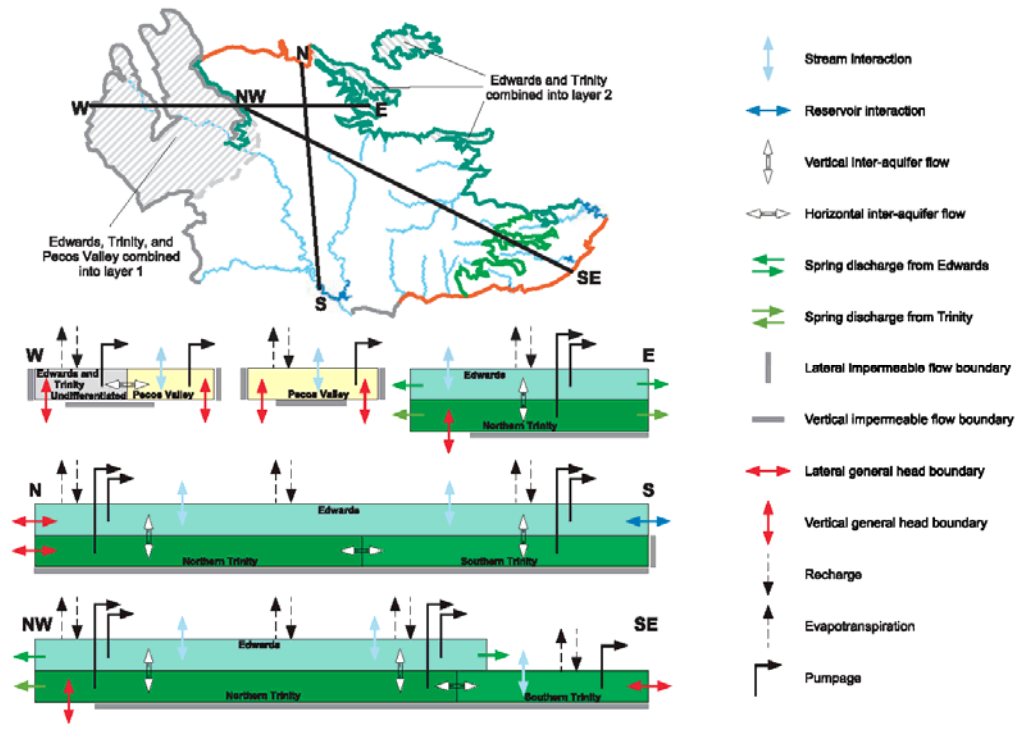


Figure 2-4. Block diagram of the Edwards-Trinity (Plateau), Trinity (Hill Country), and Pecos Valley aquifers. Source: Anaya and Jones, 2009.

3 Original and modified ETPV GAM files and aquifer zones

3.1 General

This section describes the steps taken in recalibrating the ETPV GAM (Anaya and Jones, 2009) using PEST. The conceptual basis of the model is described by Anaya and Jones (2009). The purpose of the present section is to describe the mechanics of the recalibration process.

3.2 Original model

3.2.1 Files used by the model

Appendix E. describes the MODFLOW files comprising the input dataset of the Edwards-Trinity groundwater model as supplied by TWDB. Collectively, these comprise a set of input files for use of the USGS MODFLOW96 model. (A number of other files were also supplied, but these do not comprise part of the model.) Output files generated by the model are also listed in Appendix E.1.

The name file *eddt_p_tr.nam* employed by the original ETPV GAM is displayed in Figure 3-1.

As is apparent from Figure 3-1, MODFLOW-generated text output is written to file *output.dat* while binary heads, drawdown and cell-by-cell flow term output are written to files *heads.dat*, *ddown.dat*, and *budget.dat* respectively. According to instructions provided in *oc.dat* however, no drawdown output is requested, so file *ddown.dat* is empty at the end of each model run.

3.2.2 Some model specifications

The model supplied by TWDB is programmed to run for 21 stress periods. The first stress period is of length 365250 days (1000 years). The next twenty stress periods are of length 365.25 days, and are thus of one year's duration each. These are intended to represent the years 1980 to 1999, these comprising the period over which calibration of the model was effected.

The model is provided with three layers, all of which are endowed with MODFLOW “type 2” status. If the head drops below the top surface of a Type 2 model layer, specific yield is used rather than storativity but the transmissivity is not changed to reflect a change in the saturated thickness. However, all cells in the lowermost layer are inactive, so that the number of layers represented in the model is effectively two. The grid is comprised of square cells with sides of length 5280 ft. Areas encompassing active cells within layers 1 and 2 are pictured in Figures 3-2 and 3-3. The actual model grid is not shown in these figures as it is too fine.

Recharge is provided using a series of arrays, with a different array used for each stress period. The recharge array pertaining to the initial, pseudo steady-state stress period is shown in Figure 3-4. Recharge ranges from zero to about 0.0001 ft/day (0.4 inch/year); darker coloration in Figure 3-4 indicates greater recharge.

Pumping rates provided in the MODFLOW *well* package input file are different for every stress period. Those for the final stress period are illustrated in Figures 3-5 and 3-6 for layers 1 and 2, respectively.

Cells to which a *general head boundary* condition are assigned are depicted in Figures 3-7 and 3-8 for layers 1 and 2, respectively. Conductance values within these cells are zoned. This zonation, and the conductance values with which they are associated, is illustrated in Figures 3-7 and 3-8. Heads and conductances are invariant throughout the simulation.

Cells belonging to layer 1 to which a MODFLOW *drain* status is assigned appear in Figure 3-9, while those belonging to layer 2 appear in Figure 3-10. All are assigned a conductance of 1000 ft²/day (which is high enough to guarantee little head drop across them.)

Cells to which a *stream* boundary condition is assigned are depicted in Figure 3-11; this diagram pertains to both layers. (Streams are continuous and are apparently assigned to the layer to which their elevation pertains.) Stream conductances are zoned, with each conductance being assigned one of four values, as shown in Figure 3-11. MODFLOW variables governing the *stream* boundary condition are invariant from stress period to stress period.

3.2.3 Model hydraulic properties

Horizontal hydraulic conductivity is assigned on a zonal basis in both the first and second layers. Zone dispositions and horizontal hydraulic conductivities are depicted in Figures 3-12 and 3-13. Note, however, that transmissivities, rather than hydraulic conductivities, are provided in the *bcf.dat* MODFLOW input file to conform with the input requirements of MODFLOW when employing type 2 layers.

The log of vertical conductance between layers 1 and 2 is contoured in Figure 3-14; note that the domain of this plot is restricted to regions where layers 1 and 2 overlap. It appears the distribution was constructed using the same zonation pattern as for horizontal hydraulic conductivity. The nonuniformity of the distribution suggests that vertical hydraulic conductivity (or vertical anisotropy) was assigned to layers 1 and 2 by the modelers and that this was converted to vertical conductance by the MODFLOW pre-processor that they employed; variations in layer thickness are therefore responsible for heterogeneity of the vertical conductance parameter field. Only vertical conductance appears in the MODFLOW *bcf.dat* input file.

MODFLOW type 2 layers require values for primary storage coefficient and specific yield. The former is obtained by multiplying specific storage by layer thickness. From an examination of file *bcf.dat* as supplied for the ETPV GAM, it is apparent that specific storage was supplied on a zonal basis to the graphical user interface employed in construction of the model; this package then converted these specific storage values to storage coefficient. Specific storage (back-calculated from storage coefficient) for layers 1 and 2 is depicted in Figures 3-15 and 3-16, respectively. Note that specific storage values of 0.0 were inadvertently assigned so some cells in layer 2.

Specific yield values assigned to layers 1 and 2 are shown in Figures 3-17 and 3-18.

In using the model, both with TWDB parameters described above and with parameters optimized to provide as good a fit as possible with the calibration dataset (see below), it is found that layer 2 becomes partially desaturated. The specific yield assigned to this layer therefore assumes an important role in the parameter estimation process. Incorrect assignment of elevations to the top of this layer, and to the top of layer 1 (which becomes completely saturated in places) can thrust upon specific yield some of the role required of specific storage, and vice versa. Values assigned to these quantities as documented above suggest that this may have occurred to some extent during the TWDB calibration process. The same may have occurred when repeating the calibration process with PEST as described below. However, no attempt was made to correct any mis-assignment of layer interface elevations in the recalibration process described below; hence the surrogate roles that are possibly played by these parameters are simply accepted.

3.3 Changes to MODFLOW model

3.3.1 Model timing

The original ETPV GAM comprising a long, one-thousand-year transient stress period followed by twenty 1-year stress periods was replaced by two models, one running under steady-state conditions and the other running under transient conditions. The latter employs 25 stress periods, the first five of which are 10 years long, while the last 20 of which are each 1 year long. These last 20 years cover the calibration period of the model, i.e., the years 1980 to 1999. The steady state model simulates pre-development conditions. Table 3-1 shows the stress periods and the corresponding time periods for the re-calibrated model.

Table 3-1. Stress periods and corresponding time periods for the re-calibrated model

Stress Period		Time Period (Years)	
Number	Duration (days)	Start	End
1	3652.5	1930	1940
2	3652.5	1940	1950
3	3652.5	1950	1960
4	3652.5	1960	1970
5	3652.5	1970	1980
6	365.25	1980	1981
7	365.25	1981	1982
8	365.25	1982	1983
9	365.25	1983	1984
10	365.25	1984	1985
11	365.25	1985	1986
12	365.25	1986	1987
13	365.25	1987	1988
14	365.25	1988	1989
15	365.25	1989	1990
16	365.25	1990	1991
17	365.25	1991	1992

Table 3-1. Stress periods and corresponding time periods for the re-calibrated model (Continued)

Stress Period		Time Period (Years)	
Number	Duration (days)	Start	End
18	365.25	1992	1993
19	365.25	1993	1994
20	365.25	1994	1995
21	365.25	1995	1996
22	365.25	1996	1997
23	365.25	1997	1998
24	365.25	1998	1999
25	365.25	1999	2000

3.3.2 Name files

The names and brief descriptions of the MODFLOW input and output files for the steady state and transient sub-models are provided in Appendix E. Name files for the steady state and transient model components are provided in Figures 3-19 and 3-20, respectively; these files are named *eddt_p_ss.nam* and *eddt_p_tr.nam* respectively.

As is apparent from the name files, both the steady-state and transient model components employ the same MODFLOW packages except for the *well* package; the steady-state model does not employ this package. The names of model input files reflect the packages that they represent. A suffix of “_ss” distinguishes input files for the steady-state model component from those of the transient model component. However, some files are shared by both model components; these possess no suffix.

While a binary drawdown file is cited in each of these name files, drawdown output is disabled in respective *output control* package input files, so that these files are empty at the end of a model run. Note also that cell-by-cell flow term output is disabled for all but the *streamflow routing* package. This minimizes the size of the budget output files; only stream budget data is required from these files when the model is run for calibration purposes.

3.3.3 MODFLOW data arrays

MODFLOW provides a number of different strategies for the reading of data arrays. These can be read directly from input files that hold all other data pertinent to each package. Alternatively they can be read from external files that hold these arrays alone. If the latter option is chosen, MODFLOW can be directed to a file which holds a single array by including an OPEN/CLOSE statement followed by the name of the file containing the array at pertinent locations in MODFLOW package input files. This strategy is employed extensively in the revised ETPV GAM. Table 3-2 lists array filenames used by the revised ETPV GAM, together with the array types that these files contain. (Note that the Groundwater Data Utilities convention of providing real formatted arrays with an extension of “.ref” is adopted.)

As described in documentation of the Groundwater Data Utilities, the MOD2ARRAY program facilitates reconstruction of MODFLOW input datasets to make use of external array storage in this manner. External storage of arrays allows construction and/or manipulation of these arrays to be undertaken prior to a model run without risk of damage to other aspects of a MODFLOW input dataset.

Table 3-2. External arrays employed by revised MODFLOW model.

Array filename	Data housed in array
<i>sh[1-3].ref</i>	Initial heads for transient model; cited in <i>bas_tr.dat</i> .
<i>trans[1-3].ref</i>	Transmissivity of layers 1, 2 and 3; cited in <i>bcf_ss.dat</i> and <i>bcf_tr.dat</i> .
<i>vcont[1-2].ref</i>	Vertical conductance under layers 1 and 2; cited in <i>bcf_ss.dat</i> and <i>bcf_tr.dat</i> .
<i>s[1-3].ref</i>	Primary storage coefficient for layers 1 to 3; cited in <i>bcf_tr.dat</i> .
<i>sy[1-3].dat</i>	Specific yield for layers 1 to 3; cited in <i>bcf_tr.dat</i> .
<i>top[1-3].ref</i>	Elevation of top of layers 1 to 3; cited in <i>bcf_ss.dat</i> and <i>bcf_tr.dat</i> .
<i>rech_ss.ref</i>	Steady state recharge; cited in <i>rch_ss.dat</i> .
<i>rech_av_19[30-70]s.ref</i>	Recharge averaged over 10 year periods spanning the 1930s 1940s, 1950s, 1960s and 1970s; cited in <i>rch_tr.dat</i> .
<i>rech_19[80-99].dat</i>	Recharge in one of years 1980 to 1999; cited in <i>rch_tr.dat</i> .

3.3.4 Solution convergence criterion

The HCLOSE convergence criterion applied to the PCG2 solver was set to a relatively small value of 10^{-4} ; in the original ETPV GAM, it was set to the relatively high value of 2.0. The RCLOSE convergence criterion (previously set at 100,000) was reduced to 1000. When calibrating a model using PEST, tight convergence of iterative solvers is required to enhance numerical precision of model outputs. This, in turn, raises the numerical integrity of finite-difference derivatives of model outputs with respect to adjustable parameters. These derivatives form the basis of the PEST-implemented inversion process.

3.3.5 MODFLOW executable program

The USGS version of MODFLOW96 was recompiled to use double precision arithmetic for all numerical computations. Experience (including that gained in undertaking the current modeling project) has demonstrated that this reduces the likelihood of MODFLOW convergence difficulties especially when the convergence criterion is set tighter than normal as is discussed above. The recompiled version of MODFLOW96 is named *mf96_d.exe*. This was compiled with the Lahey-Fujitsu LF95 compiler. A single precision version, compiled with the same compiler (named *mf96.exe*) is also provided.

Use of a double precision version of MODFLOW has repercussions for MODFLOW post-processing. In particular, all binary files written by the double precision MODFLOW record 8-byte versions of all real numbers. Post-processing software must be capable of reading these files using the same protocol.

A number of alterations were made to the USGS MODFLOW96 source code prior to recompilation. These are as follows.

1. The size of the LENX array was increased to accommodate the storage demands of the ETPV GAM.
2. Code was added to allow writing of the stress period and time step number to the screen as the simulation progresses.
3. MODFLOW is prevented from ceasing execution if solution convergence failure occurs during a transient run (this is essential for unsupervised running of MODFLOW by PEST).
4. Unformatted file storage is declared as “binary” rather than “unformatted” as this affords greater compatibility with commonly used MODFLOW graphical user interfaces (and with post-processing software of the Groundwater Data Utilities suite).

3.3.6 Transferring data between MODFLOW model components

As discussed in greater detail in later sections, the batch file used by PEST as “the model” instructs MODFLOW to run the steady-state model first, and then the transient model. Heads calculated by the steady state model are used as initial heads by the transient model. The MANY2ONE program supplied with the Groundwater Data Utilities extracts heads arrays from the binary *heads_ss.dat* file written by the steady state model, and rewrites these to ASCII files *sh1.ref*, *sh2.ref*, and *sh3.ref* for use of the transient model.

TWDB Report ##: Final – Application of PEST to Re-Calibrate
the Groundwater Availability Model for the Edwards-Trinity (Plateau) and Pecos Valley Aquifers

LIST 3 output.dat

BAS	1	bas.dat
BCF	11	bcf.dat
OC	22	oc.dat
WEL	12	wel.dat
DRN	13	drn.dat
GHB	17	ghb.dat
RCH	18	rch.dat
STR	28	str1.dat
PCG	23	pcg2.dat
DATA(BINARY)	50	budget.dat
DATA(BINARY)	51	heads.dat
DATA(BINARY)	52	ddown.dat

Figure 3-1. Contents of name file *eddt_p_tr.nam*.

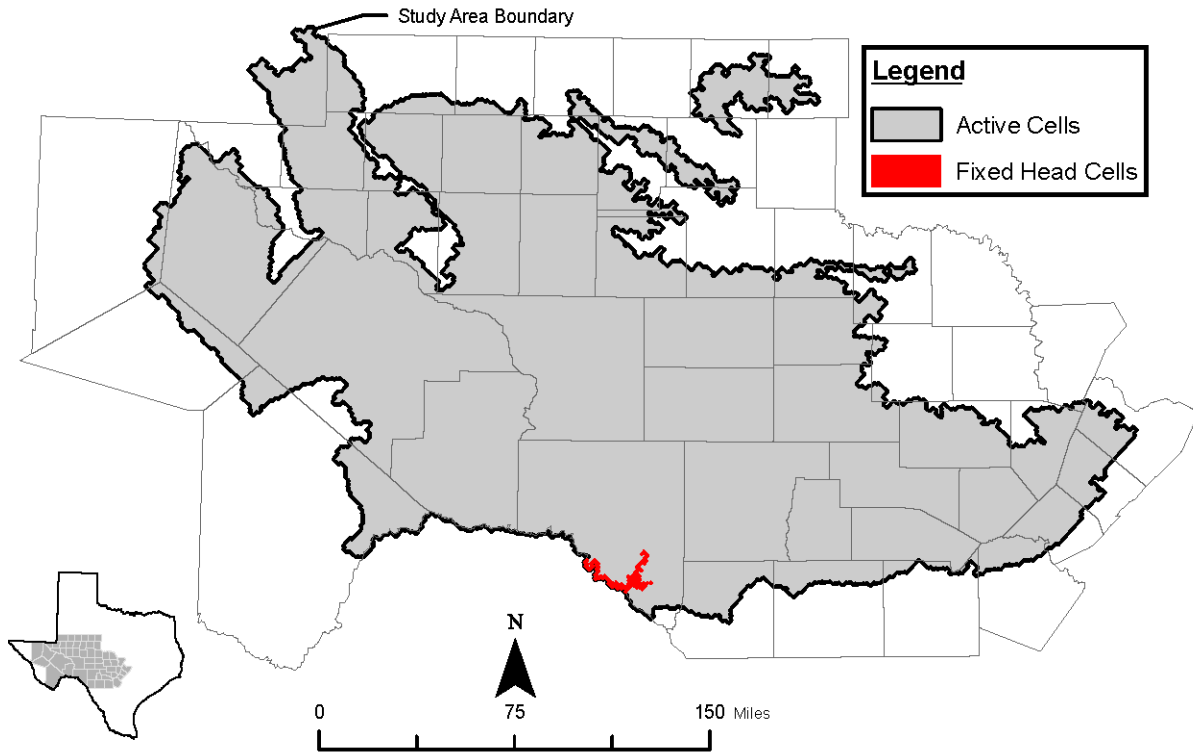


Figure 3-2. Active area of layer 1 of the model domain. Fixed head cells are colored red.

TWDB Report ##: Final – Application of PEST to Re-Calibrate the Groundwater Availability Model for the Edwards-Trinity (Plateau) and Pecos Valley Aquifers

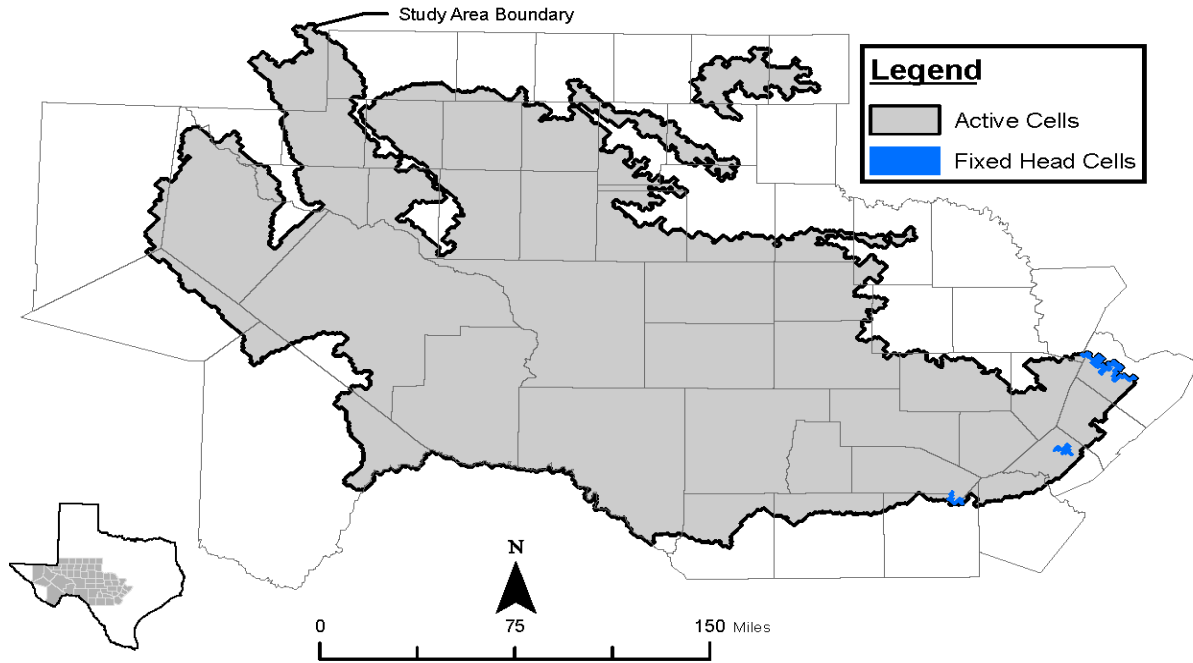


Figure 3-3. Active area of layer 2 of the model domain. Fixed head cells are colored blue.

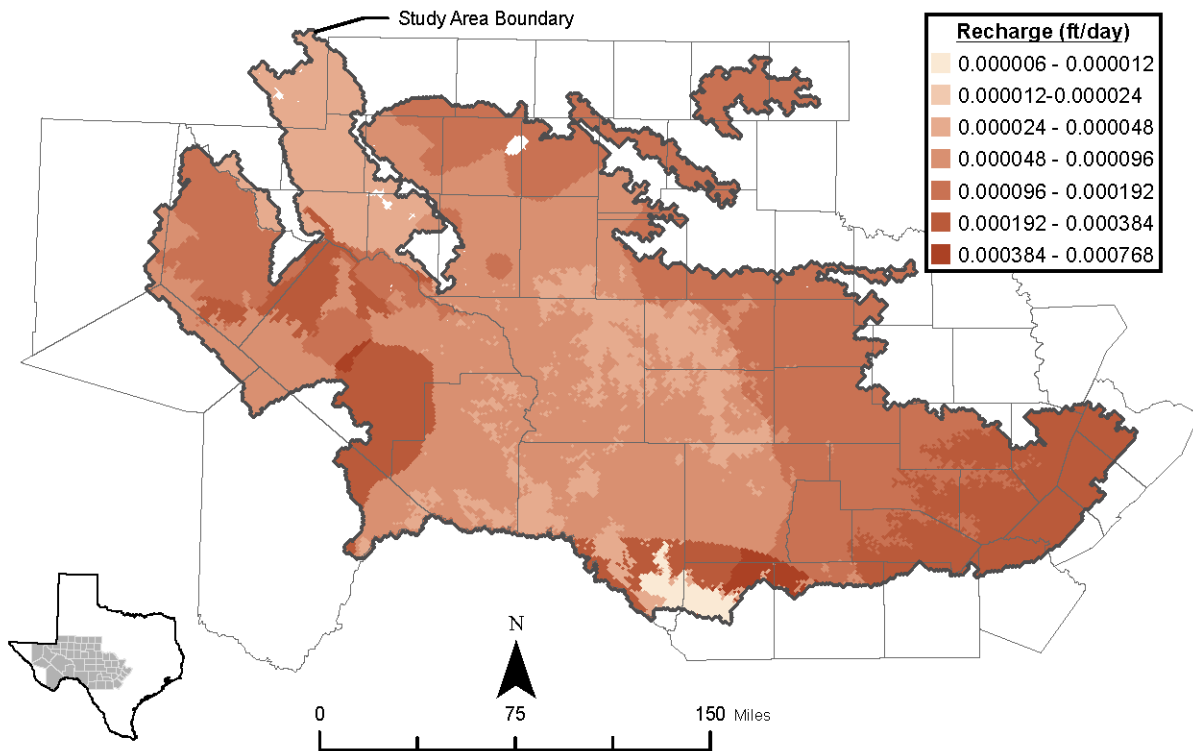


Figure 3-4. Recharge distribution for the steady-state model representing predevelopment conditions .

TWDB Report ##: Final – Application of PEST to Re-Calibrate
the Groundwater Availability Model for the Edwards-Trinity (Plateau) and Pecos Valley Aquifers

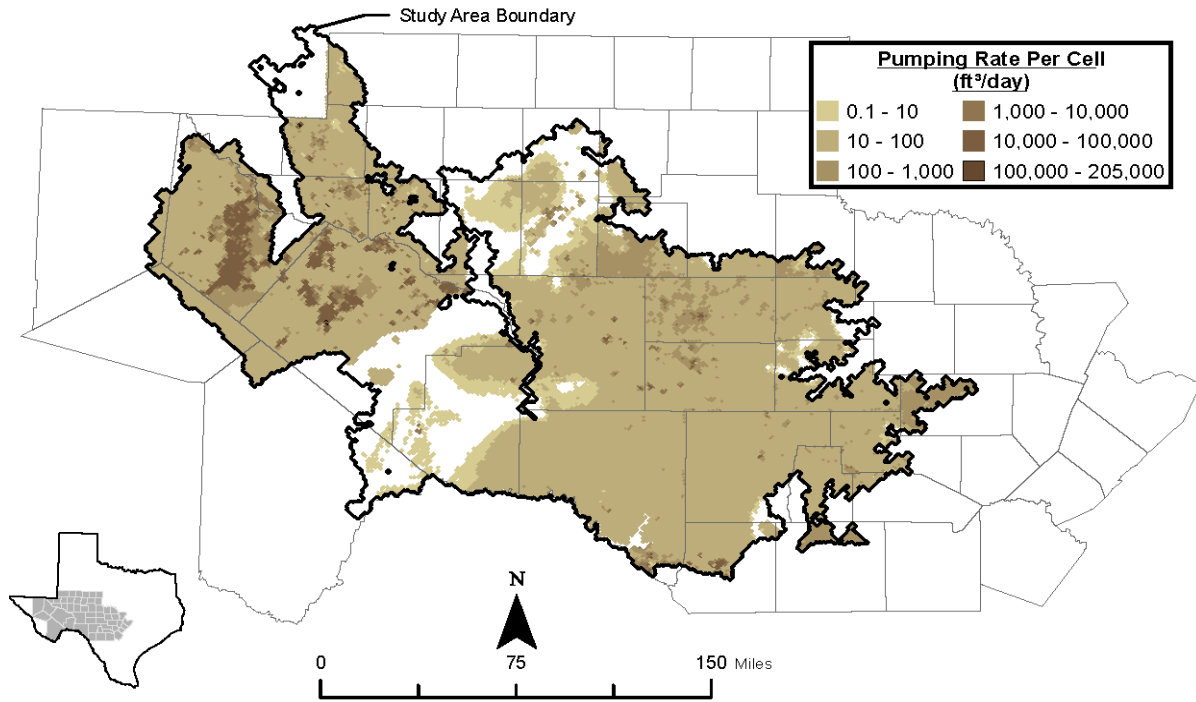


Figure 3-5. Pumping rate in layer 1 in stress period 21.

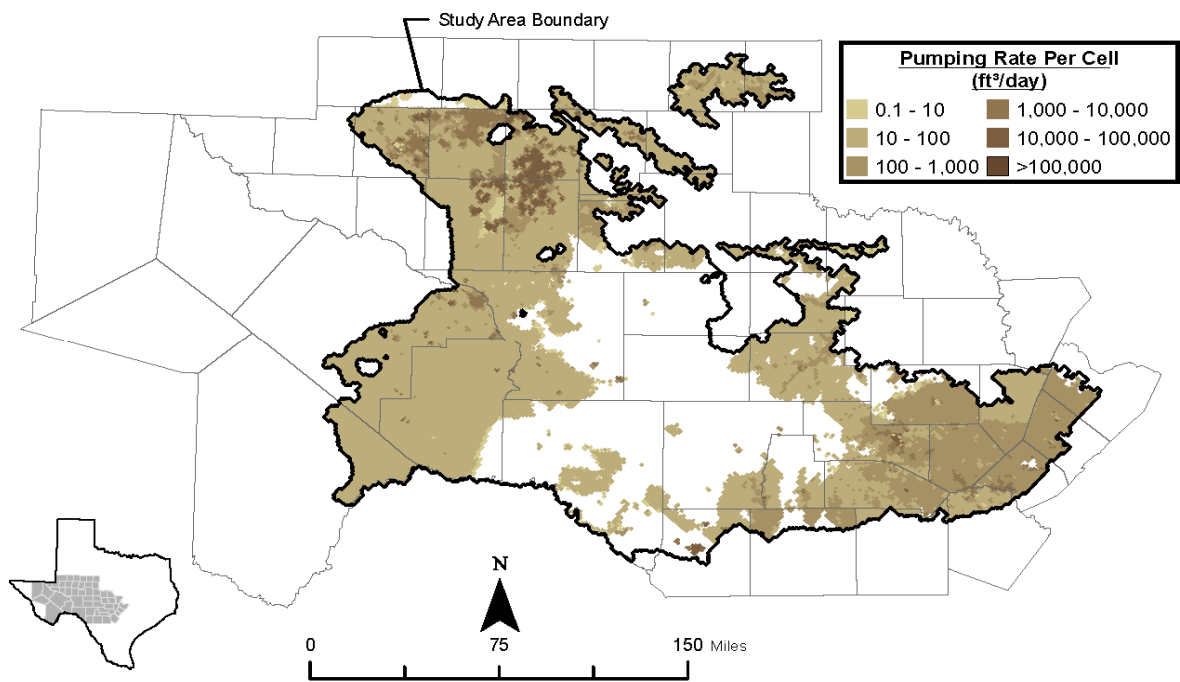


Figure 3-6. Pumping rate in layer 2 in stress period 21.

TWDB Report ##: Final – Application of PEST to Re-Calibrate
the Groundwater Availability Model for the Edwards-Trinity (Plateau) and Pecos Valley Aquifers

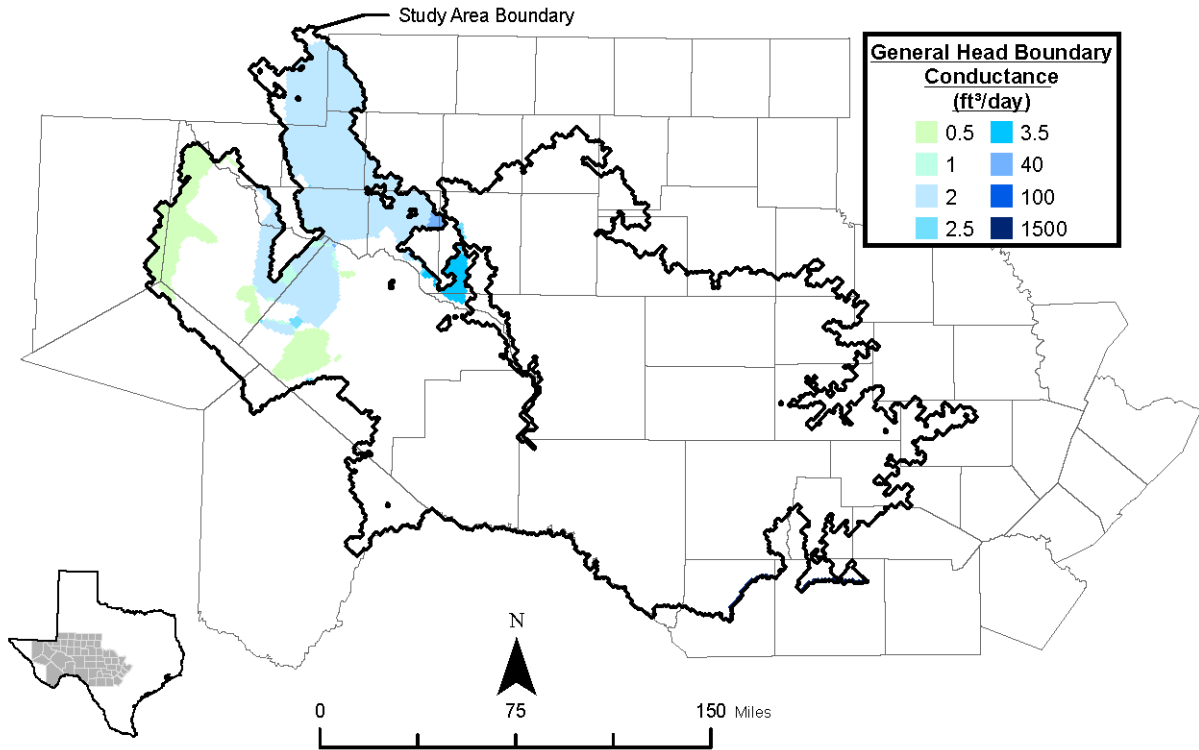


Figure 3-7. General head boundary cells in layer 1 colored by conductance.

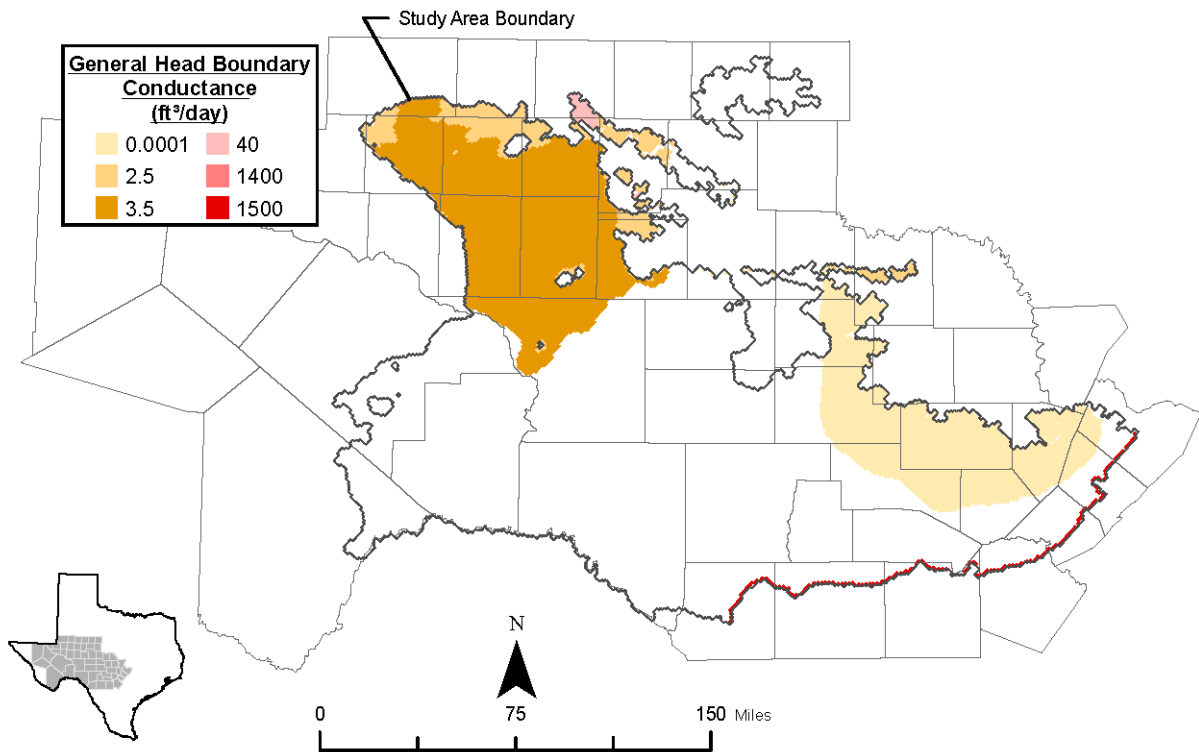


Figure 3-8. General head boundary cells in layer 2 colored by conductance.

TWDB Report ##: Final – Application of PEST to Re-Calibrate
the Groundwater Availability Model for the Edwards-Trinity (Plateau) and Pecos Valley Aquifers

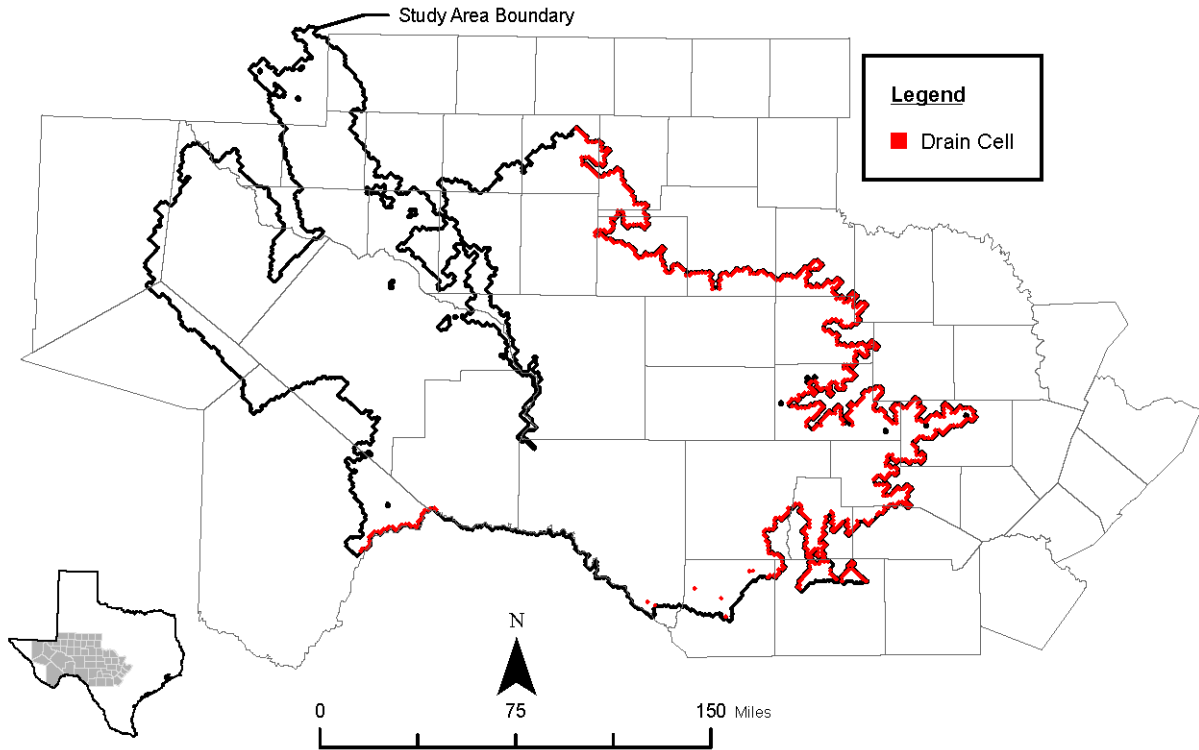


Figure 3-9. *Drain cells in layer 1. Conductance is uniformly 1,000 ft²/day.*

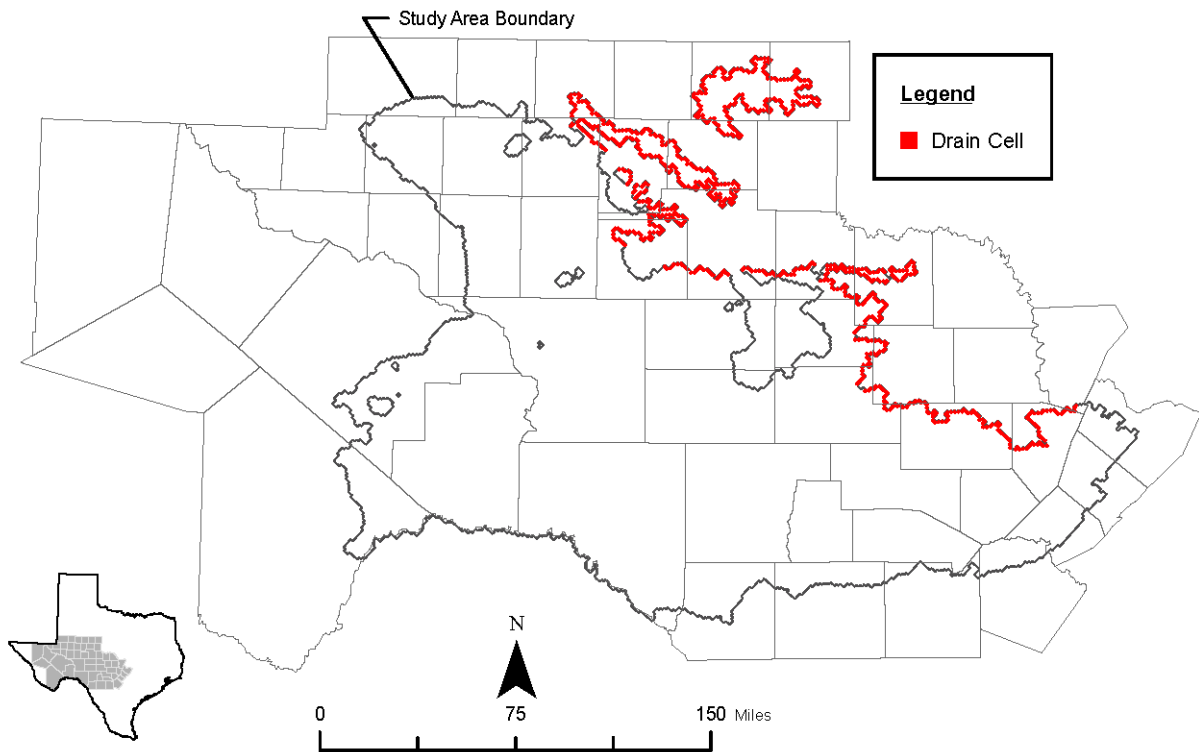


Figure 3-10. *Drain cells in layer 2. Conductance is uniformly 1,000 ft²/day.*

TWDB Report ##: Final – Application of PEST to Re-Calibrate
the Groundwater Availability Model for the Edwards-Trinity (Plateau) and Pecos Valley Aquifers

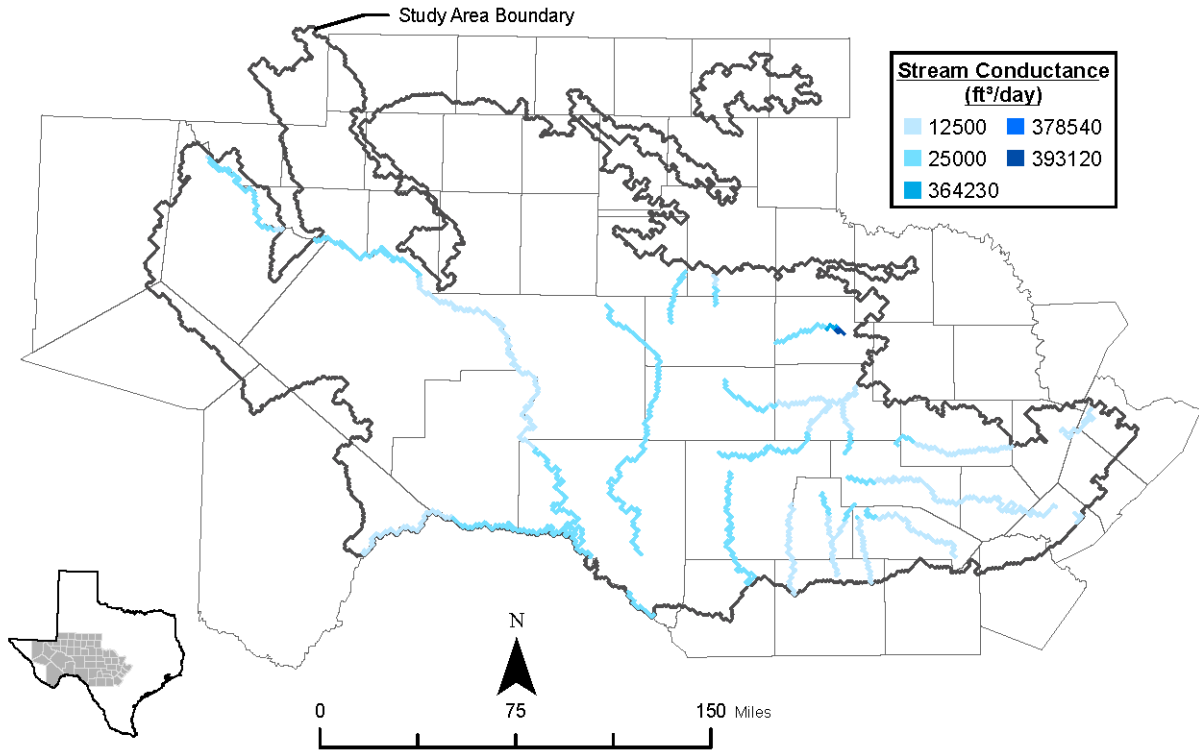


Figure 3-11. Stream cells in both layers, colored by conductance value.

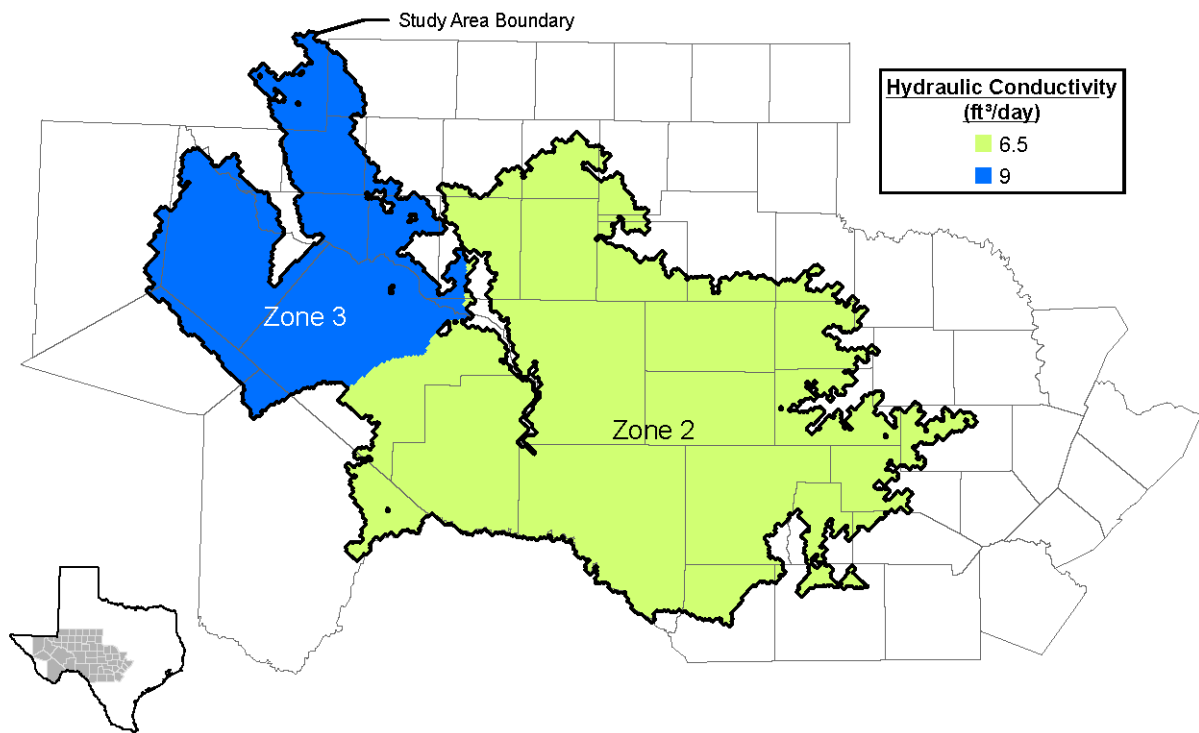


Figure 3-12. Hydraulic conductivity in layer 1. The same zonation as that depicted in this figure is employed in the recalibration process described below; zone numbers are provided.

TWDB Report ##: Final – Application of PEST to Re-Calibrate the Groundwater Availability Model for the Edwards-Trinity (Plateau) and Pecos Valley Aquifers

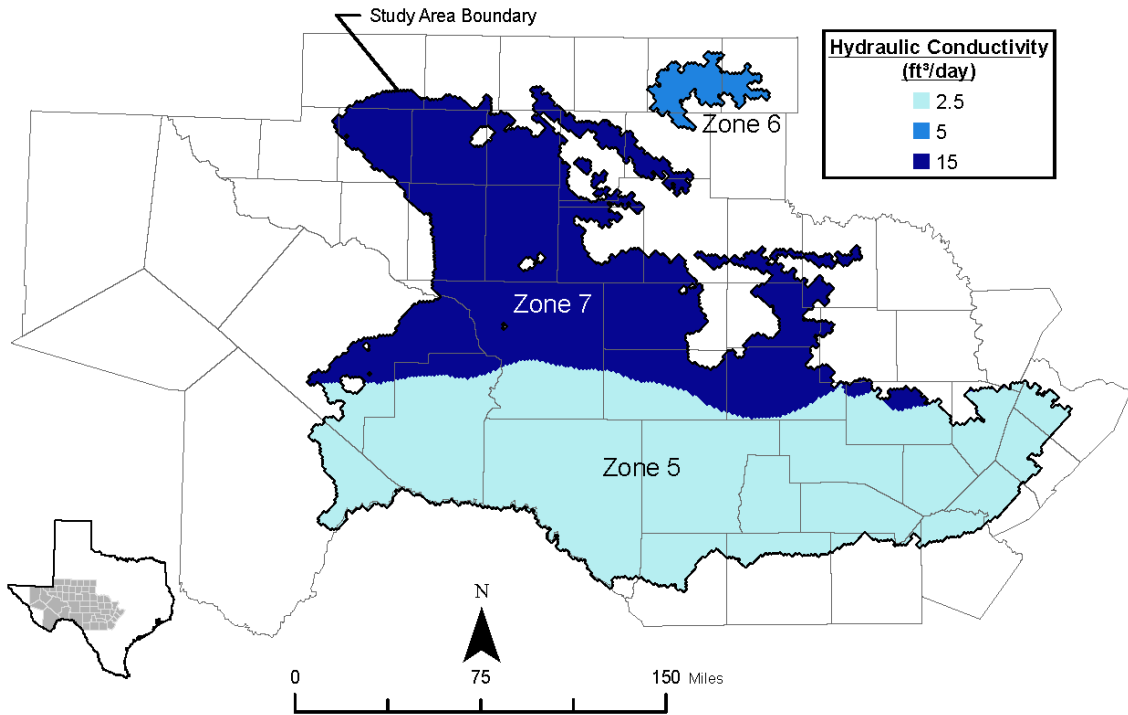


Figure 3-13. Horizontal conductivity in layer 2. The same zonation as that depicted in this figure is employed in the recalibration process described below; zone numbers are provided.

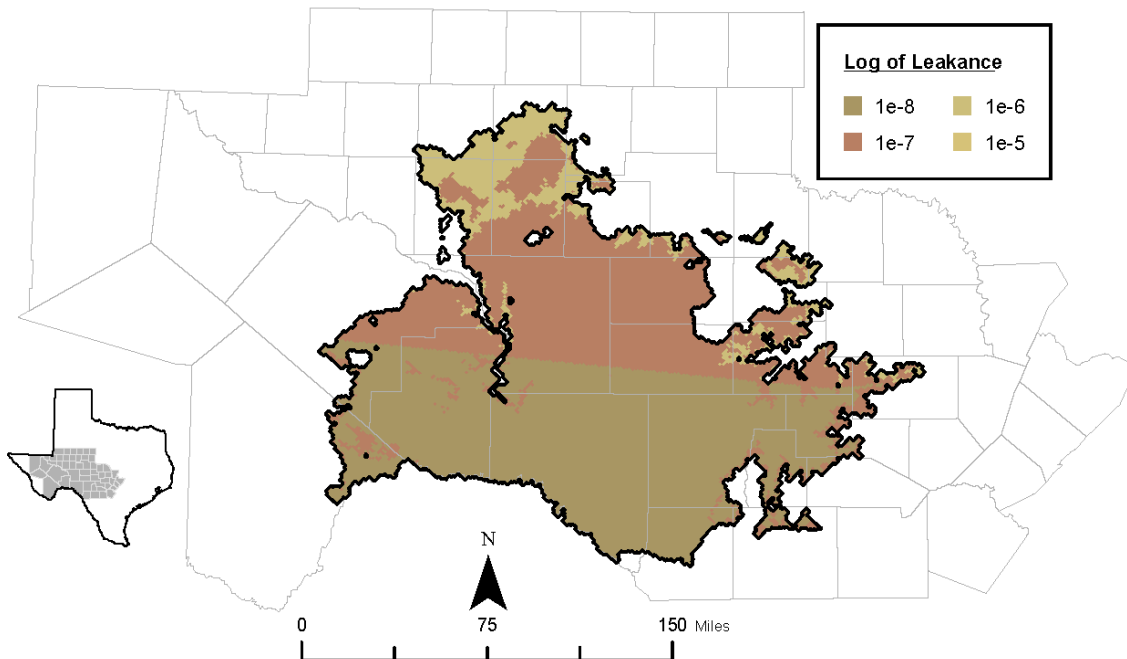


Figure 3-14. Log (to base 10) of vertical conductance between layers 1 and 2; greener indicates higher conductance. Log of conductance ranges between -8 and -4.6.

TWDB Report ##: Final – Application of PEST to Re-Calibrate
the Groundwater Availability Model for the Edwards-Trinity (Plateau) and Pecos Valley Aquifers

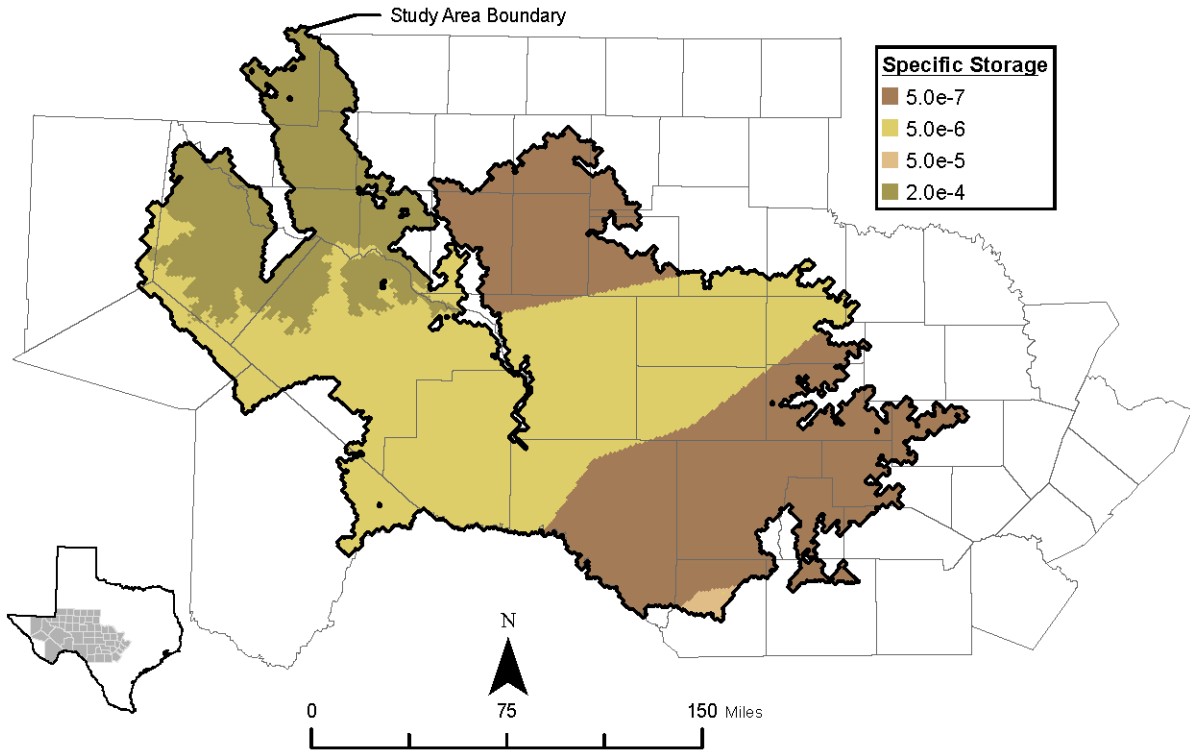


Figure 3-15. Specific storage assigned to layer 1.

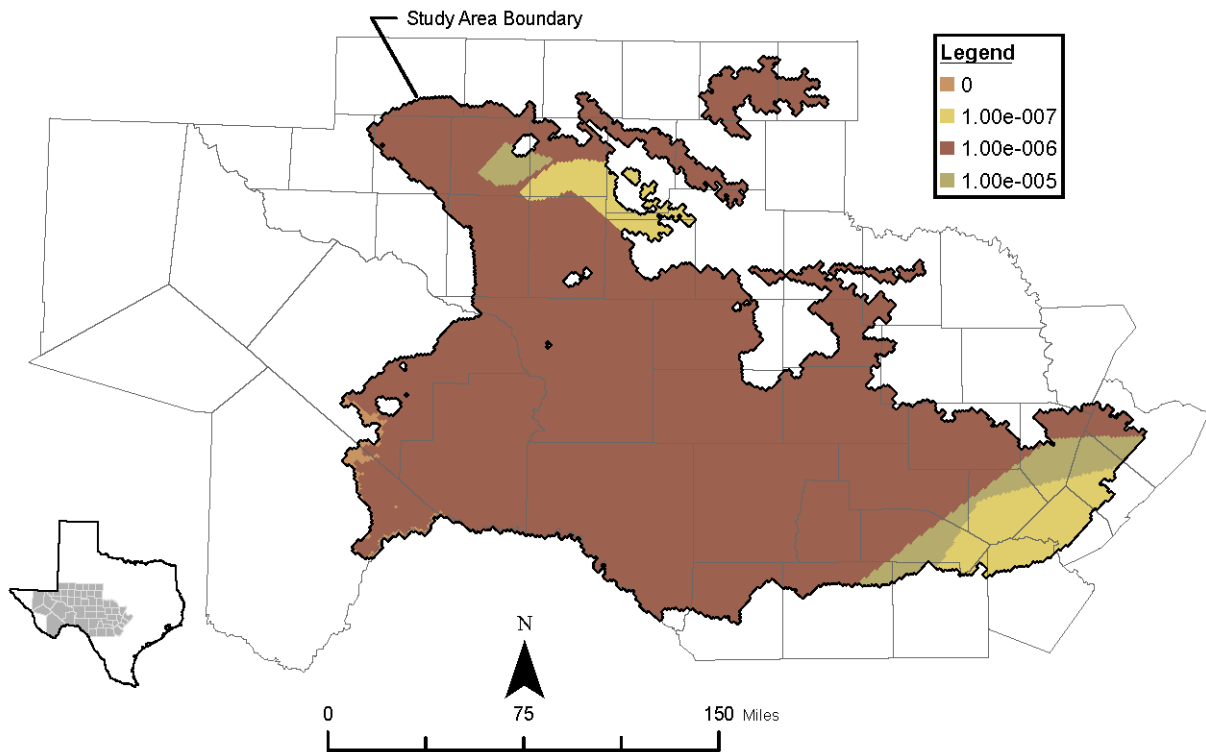


Figure 3-16. Specific storage assigned to layer 2.

TWDB Report ##: Final – Application of PEST to Re-Calibrate
the Groundwater Availability Model for the Edwards-Trinity (Plateau) and Pecos Valley Aquifers

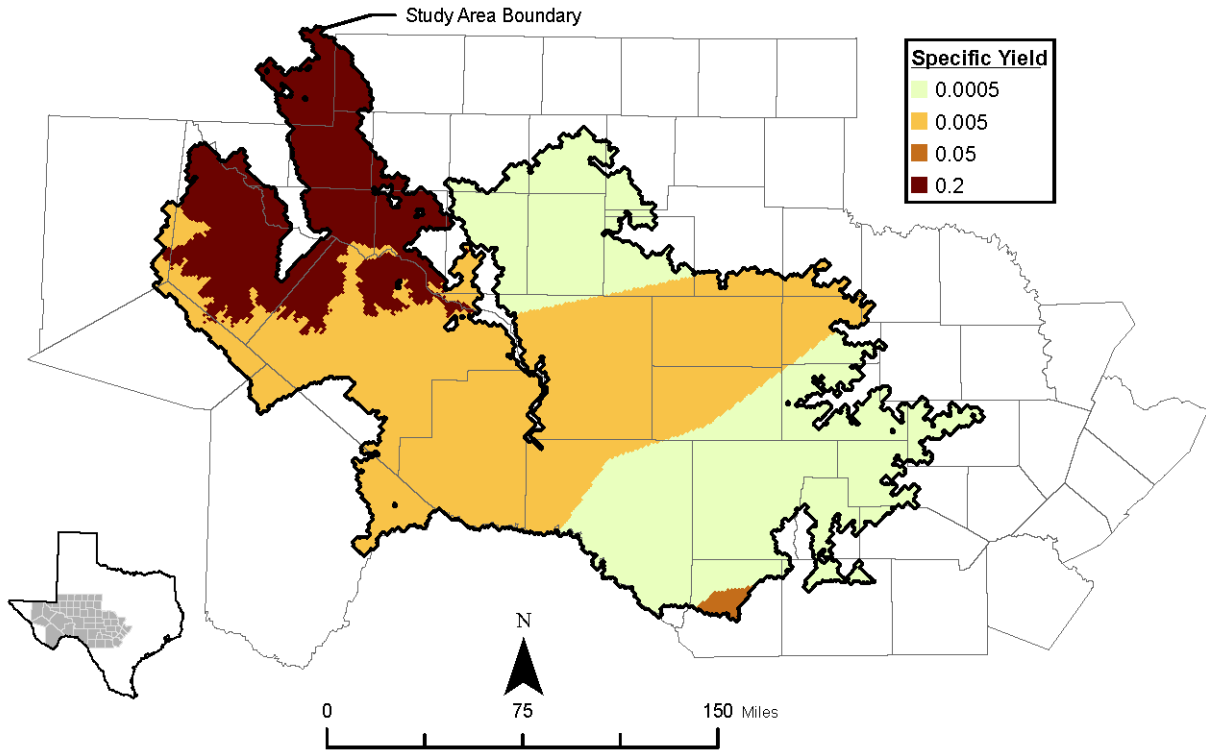


Figure 3-17. Specific yield assigned to layer 1.

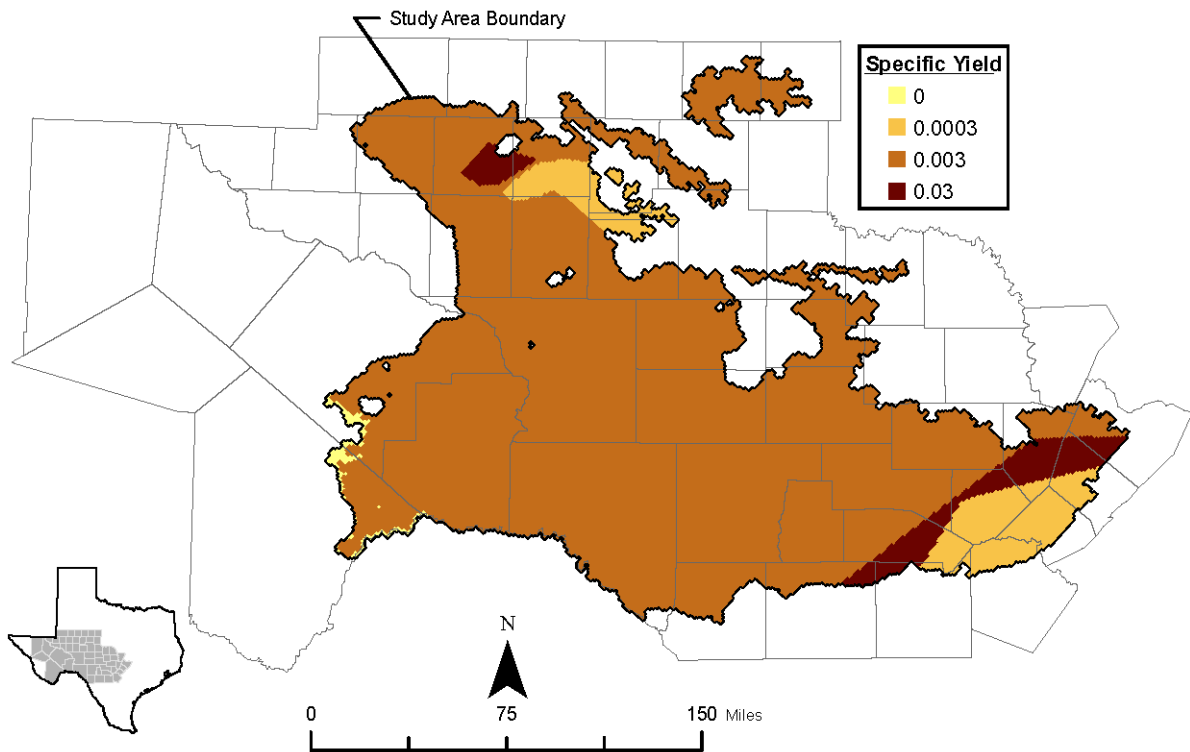


Figure 3-18. Specific yield assigned to layer 2.

TWDB Report ##: Final – Application of PEST to Re-Calibrate
the Groundwater Availability Model for the Edwards-Trinity (Plateau) and Pecos Valley Aquifers

```
LIST      3  output_ss.dat
BAS       1  bas_ss.dat
BCF      11  bcf_ss.dat
OC       22  oc_ss.dat
DRN      13  drn.dat
GHB      17  ghb.dat
RCH      18  rch_ss.dat
STR      28  str1.dat
PCG      23  pcg2_ss.dat
DATA(BINARY)  50  budget_ss.dat
DATA(BINARY)  51  heads_ss.dat
DATA(BINARY)  52  ddown_ss.dat
```

Figure 3-19. Name file *eddt_p_ss.nam* for steady state component of model.

```
LIST      3  output_tr.dat
BAS       1  bas_tr.dat
BCF      11  bcf_tr.dat
OC       22  oc_tr.dat
WEL      12  wel_tr.dat
DRN      13  drn.dat
GHB      17  ghb.dat
RCH      18  rch_tr.dat
STR      28  str1.dat
PCG      23  pcg2_tr.dat
DATA(BINARY)  50  budget_tr.dat
DATA(BINARY)  51  heads_tr.dat
DATA(BINARY)  52  ddown_tr.dat
```

Figure 3-20. Name file *eddt_p_tr.nam* for transient component of model.

4 Observations and parameters for model calibration

4.1 Observation dataset

4.1.1 General

As discussed below, substantial changes were made to the parameterization of the model and the to the calibration metrics prior to recalibration using PEST. These changes included the introduction of many new head observations, model parameters, and adoption of the philosophy of highly parameterized inversion discussed in the preceding section. At the same time, a new head dataset was constructed, this being far more extensive than that used in previous calibration of the ETPV GAM. This calibration dataset is now briefly described.

4.1.2 Head observations

Yearly averaged head observations over the period 1980 to 2000 were used in recalibration of the ETPV GAM. The data used to develop the calibration targets were queried from the TWDB water well database. Only the wells that contained five or more head measurements were used. If fewer measurements than five were available the well was not represented in the calibration dataset for that year. The restriction of five measurements was imposed in case a future modeler wished to examine the hydrographs for outlier or an unrepresentative measurements.

Our query of the TWDB database yielded a total of 574 wells with 4,773 head measurements. 278 of the wells and 2,549 of the measurements were assigned to layer 1. The remaining 296 wells and 2,224 measurements were assigned to layer 2. The well locations for each layer are shown on Figures 4-1 and 4-2. In these figures, the well locations are color-coded based on the number of well measurements for each well over the period from 1980 to 2000. Appendix F provides the state well ID and the coordinates of the wells mapped on Figures 4-1 and 4-2. Table 4-1 summarizes the number of wells and water level measurements associated with each county.

The well information shown in Appendix F is provided to PEST through the file *measure_cnt_gt_6b.crd*. Yearly averaged heads are provided in file *measure_cnt_gt_6b.smp*. The former file uses the “bore coordinates file” format described in documentation of the PEST Groundwater Data Utilities, while the latter file employs “bore sample file” format which is described in the same documentation.

Table 4-1. Well locations used for the model calibration from 1980 to 2000.

County	Number of wells	Number of measurements	County	Number of wells	Number of measurements
All Counties	574	4,773	Loving	2	20
Andrews	0	0	Martin	0	0
Bandera	40	267	Mason	0	0
Bexar	10	77	McCulloch	1	12
Blanco	11	82	Medina	0	0
Brewster	1	6	Menard	13	110

County	Number of wells	Number of measurements	County	Number of wells	Number of measurements
Burnet	0	0	Midland	7	66
Coke	0	0	Mitchell	0	0
Comal	8	34	Nolan	4	35
Concho	4	30	Pecos	31	333
Crane	2	18	Reagan	21	168
Crockett	27	191	Real	9	91
Culberson	0	0	Reeves	55	555
Ector	17	117	Schleicher	15	126
Edwards	10	65	Sterling	2	19
Gillespie	46	317	Sutton	17	159
Glasscock	23	249	Taylor	0	0
Hays	15	72	Terrell	5	22
Howard	0	0	Tom_Green	0	0
Irion	7	66	Travis	7	62
Jeff_Davis	0	0	Upton	6	72
Kendall	31	246	Uvalde	0	0
Kerr	31	213	Val_Verde	0	0
Kimble	13	91	Ward	34	369
Kinney	6	52	Winkler	12	115

4.1.3 Head differences

Use of temporal head differences (at the same time as heads themselves) in the calibration process of a transient model often endows the calibration process with a greater capacity to estimate storage parameters than would otherwise be the case, particularly if these differences are weighted such that they are as “visible” in the overall calibration objective function as the heads themselves. Temporal differencing of head measurements can be accomplished using the SMPDIFF utility from the Groundwater Data Utilities suite. File *measure_cnt_gt_6b_diff.smp* (written by SMPDIFF through processing of data contained in file *measure_cnt_gt_6b.smp*) contains differences between the head pertaining to each measurement date cited therein, and the head pertaining to the first cited measurement for each well. These head differences are also cited in the PEST control file used in the calibration process. For the first several calibration runs using PEST, these differences were used as part of the objective function. However, for the last series of PEST applications, the head differences were assigned a weight of zero and therefore not used as part of the model calibration as it was discovered that the level of noise associated with these differences precluded their effective use in the calibration process.

4.1.4 Streamflow observations

Although not eventually used as a basis for parameter adjustment, streamflow observations were included in the calibration dataset. These were compared with flows calculated by the MODFLOW *streamflow routing* package at cells depicted in Figure 4-3.

As for heads in observation wells, yearly averaged flows for each of years 1980 to 1999 (where available) were compared with corresponding model-calculated quantities. Observed yearly average flows are provided in file *observ_flows.smp*. These also appear in the PEST control file used for model calibration but are awarded weights of zero in this file.

The stream flows were not used in the final PEST calibration for two reasons. One of the reasons is a concern that the vertical and the areal discretization in the ETPV GAM may not be adequate to support the prediction of stream flows. Young, et al. (2009) show that similar grid spacing in a Gulf Coast Aquifer GAM is inadequate to provide for a situation where the model could be calibrated to stream flows without using unrealistic aquifer parameters or adversely affecting the model's ability to predict drawdowns caused by regional pumping. As discussed by Young, et al. (2009), the more refined grid is needed to give the groundwater model the capability to adequately represent vertical head gradients and elevations of stream channel bottoms. Proper representation of vertical head gradients can be important where there are head differences between the shallow groundwater flow system that interacts with the streams and the deeper groundwater flow system associated with the regional pumping. Proper representation of stream channel bottoms can be important where there is appreciable differences in elevation along the reach of a stream channel that is contained within a single grid cell.

The second reason for not including streamflow observation in the final model calibration is that a PEST calibration that does not include streamflow as a model calibration parameter was found to nevertheless provide a relatively good match to the measured stream flow based on the criterion used by Anaya and Jones (2009). The criterion is a visual comparison of the simulated groundwater discharge to the streams at the bottoms of rises and falls of the stream hydrograph curves.

4.2 Parameterization of aquifer properties

4.2.1 Horizontal hydraulic conductivity

Because the ETPV GAM employs type 2 MODFLOW layers, MODFLOW's input data requirements are such that it requires arrays of cell-by-cell transmissivity rather than of hydraulic conductivity. Nevertheless, the recalibration process is designed in such a way as to allow direct estimation of hydraulic conductivity. This approach is used so that the development of the transmissivity field could be guided and constrained based on the set of hydraulic conductivity values provided to the authors by the TWDB and used by Anaya and Jones (2009) to calibrate the ETPV GAM (Anaya and Jones, 2009). Transmissivity is computed from hydraulic conductivity automatically prior to the running of MODFLOW as part of the composite model run by PEST encapsulated in the batch file *model_gt6b.bat*.

Parameterization of hydraulic conductivity takes place at two levels. Values are assigned to the five zones represented in Figures 3-12 and 3-13 and used by the ETPV GAM to delineate the major aquifers. Within each layer these zonal values are then multiplied by a layer-specific multiplier field parameterized using pilot points. Note that this multiplier field allows continuity of interpolation across zone boundaries, thus allowing some "blurring" of these boundaries to take place to accommodate the fact that the locations of corresponding geological boundaries are not exactly known. During the parameter estimation process, upper and lower bounds were placed on zonal hydraulic conductivity values and on final post-multiplier-field hydraulic conductivity values. Preferred values (enforced by Tikhonov constraints) for the initial PEST simulations were supplied for zonal parameters, these being computed through geometric

averaging of pumping-test-estimated hydraulic conductivities within each zone. A Tikhonov-constrained preferred value of 1.0 was provided for all pilot-point-multiplier parameters.

The preferred values and bounds described above were developed using a set of hydraulic conductivity values provided to us by the TWDB at the start of his project. Anaya and Jones (2009) discuss how these conductivity values were assembled from different information sources and types of hydraulic tests. In total, the data include 190 conductivity values for the Edwards hydrostratigraphic unit (zone 2), 655 conductivity values for the Trinity hydrostratigraphic unit (zones 5, 6, and 7) and 56 conductivity values for the Pecos Valley Aquifer (zone 3). Figures 4-4 and 4-5 show that some of the wells associated with 901 conductivity values lie outside the model coverage. Figure 4-6 shows the cumulative distribution for the conductivity values in each zone. Using these cumulative distributions, the values in Table 4-2 were computed to help develop input parameters for PEST.

Table 4-2. Selected statistical values for the hydraulic conductivity estimates in each aquifer zone based on the cumulative distributions shown on Figure 4-6 (all values in ft/day).

	Zone			
	2	3	5	7
	Edwards	Pecos Valley	Trinity South	Trinity North
Geometric average	3.41	7.81	1.90	3.91
Median (50 th percentile)	4.72	7.54	1.99	3.74
Standard deviation	18.28	2.72	6.89	4.66
Minimum	<0.01	0.89	0.01	0.05
10 th percentile	0.07	2.96	0.14	0.56
25 th percentile	0.37	3.73	0.48	1.87
40 th percentile	1.54	4.62	1.24	2.85
60 th percentile	11.16	9.34	3.40	4.81
75 th percentile	22.4	11.6	7.3	7.74
90 th percentile	185.7	39.7	21.6	29.6
Maximum	3,467	95	741	221
Number of measurements	105	55	504	74

Our analysis of the conductivity data was primarily limited to statistical and geostatistical analyses. Our analysis did not provide any conclusive evidence for using a preferred orientation or trend to constrain the spatial variability of conductivity in each aquifer zone. Moreover, we did not investigate the possible stratigraphic and lithologic controls on the conductivity field within each aquifer zone; nor did we investigate possible biases in the conductivity values. The values for the latter were evident based on the results of conductivity values for the Chicot and Evangeline aquifers (Young and Kelley, 2006, who identified several sources of bias in their review). Among the findings of their work is that in the development of regional groundwater models such as the ETPV, significant biases can exist in conductivity values as a result of the associated screen interval of the tested well and the type of hydraulic test performed.

To constrain the PEST simulation, we used the 40th percentile and the 60th percentile in Table 4-2 as the lower and upper bounds, respectively, for the preferred average for each aquifer zone. The preferred zonal and bounds averages were changed over the course of the project based on the results of the previous PEST calibration runs. Table 4-3 shows the set of conductivity

constraints used for the final PEST simulations. Because of the lack of conductivity data for zone 6, the bounds for zone 6 were assumed equal to those of zone 5.

Table 4-3. Preferred values and bounds on horizontal hydraulic conductivity imposed during calibration (all values in ft/day).

Zone number	Preferred zonal average	Zonal lower bound	Zonal upper bound
2	4.72	1.54	11.16
3	7.54	4.72	9.34
5	1.99	1.24	3.4
6	3.74	2.85	4.8
7	3.74	2.85	4.8

The pilot point locations for the layer 1 and 2 multiplier fields are depicted in Figures 4-7 and 4-8. Interpolation from pilot points to the model grid employs ordinary kriging. Our attempts to construct a useful variogram based on the hydraulic conductivity data were not successful. As a result, we calculated kriging factors based of an exponential variogram with a distance exponent of 120,000 ft. As is explained by Doherty et al. (2009c), it is not necessary that the interpolation variogram be based on what is perceived to be the true variogram of hydraulic properties occurring within a model domain. Use of a relatively short range bestows on the interpolation process responsiveness to local heterogeneity, should this be present; in its absence, smooth parameter fields are promulgated through use of Tikhonov regularization.

Kriging factors were computed using the PPK2FAC1 utility (a member of the Groundwater Data Utility suite). Layer 1 kriging factors are housed in file *factors_lay1.dat* while layer 2 kriging factors are housed in file *factors_lay2.dat*.

Hydraulic conductivity zonal dispositions are provided in integer array files *kzone1.inf* and *kzone2.inf* (“*.inf*” is the recommended extension for formatted integer array files when using Groundwater Data Utility suite programs). Real array files containing zone-based hydraulic conductivities are housed in files *hk_klay1_zone.ref* and *hk_lay2_zone.ref*. The latter are built from the former using the INT2REAL utility supplied with the Groundwater Data Utilities. INT2REAL is run as part of the composite model employed by PEST; “integer-real-correspondence” files named *hk_lay1.irc* and *hk_lay2.irc* list hydraulic property values assigned to zones. (These files are written by PEST on the basis of corresponding template files during the parameter estimation process.) The TWOARRAY utility (another member of the Groundwater Data Utility suite) is used to multiply these zone-based parameter fields by the pilot-point-based multiplier fields contained in files *hk_kay1_mul.ref* and *hk_lay2_mul.ref*. The latter are built through interpolation from pilot point multipliers contained in pilot point files *hk_lay1_mul.dat* and *hk_lay2_mul.dat* by the FAC2REAL utility (another member of the Groundwater Data Utility suite). The resulting real array files are named *hk_lay1.ref* and *hk_lay2.ref*.

To restrict the amount of spatial variability in conductivity values, zone-specific limits are recorded in files *hk_lay1.lim* and *hk_lay2.lim*. These limits are imposed onto the conductivity distribution by the CLIPARRAY1 utility; this utility was written for the present project and is documented in Appendix G. Table 4-4 lists the limits used for the final PEST simulations. The

limits were selected to minimize the range in the conductivity values while achieving an average that lies between the lower and upper bounds for the preferred conductivity average for each zone. In addition, the lower boundary for Zone 2 was a select value closer to the 60th percentile instead of the 50th percentile to promote higher recharge rates based on results from previous PEST runs.

Table 4-4. Cut-off limits imposed on the hydraulic conductivity values assigned to each grid cell (all values in ft/day).

Limit on Hydraulic Conductivity	Aquifer Zone				
	2	3	5	6	7
Lower	8	5	1	1	1
Upper	13	15	7	7	6

The transmissivity real array files that are actually used by MODFLOW (i.e., *trans1.ref* and *trans2.ref*) are computed by the TWOARRAY utility which multiplies hydraulic conductivity by layer thickness (the latter housed in files *thick1.ref* and *thick2.ref*) on a cell-by-cell basis. All of INT2REAL, FAC2REAL, TWOARRAY (twice) and CLIPARRAY1 are run for each layer within the batch file *model_gt6b.bat* employed by PEST as the model for the purpose of model calibration.

See Appendix E.3 for a listing of all files (including those mentioned above) used in the calibration process.

4.2.2 Aquifer specific storage and specific yield

For aquifer storage, a parameterization process similar to developing the conductivity field is used. The parameterization is performed on the basis of specific storage, which is multiplied by layer thickness to determine storage coefficient. Zone-based parameterization of specific storage values is subjected to adjustment through multiplication by pilot-point-based multiplier parameters. Zonation is based on the same integer arrays as are used for definition of zonal horizontal hydraulic conductivity parameters. Multipliers are based on the same set of pilot points, and the same kriging factors. Appendix E lists the computer files used in this process. Table 4-5 lists (Tikhonov enforced) preferred values, as well as upper and lower bounds placed on specific storage during the parameter estimation process.

Table 4-5. Preferred values and bounds on specific storage imposed during calibration (all values in ft⁻¹).

Zone number	Preferred value	Zonal lower bound	Zonal upper bound
2	1e-5	1.e-6	5e-5
3	5e-5	1e-5	1e-4
5	1e-5	1e-6	5e-5
6	1e-5	1e-6	5e-5
7	1e-5	1e-6	5e-5

Construction of specific yield arrays for the use of MODFLOW is similar; however, no multiplication by layer thickness is required. Appendix E lists the computer files used in this process. Table 4-6 lists preferred values, and upper and lower bounds placed on specific storage during the parameter estimation process.

Table 4-6. Preferred values and bounds on specific yield imposed during calibration (all values are dimensionless).

Zone number	Preferred value	Zonal lower bound	Zonal upper bound
2	0.01	0.008	0.03
3	0.10	0.09	0.15
5	0.08	0.07	0.125
6	0.08	0.07	0.125
7	0.08	0.07	0.125

Prior to developing the values in Tables 4-5 and 4-6, the authors reviewed 22 storativity values used by Anaya and Jones (2009) to help calibrate the ETPV GAM. Of these 22 values, there are less than three values in all of the aquifer zones except for the Zone 7, the northern Trinity. As a result, the authors primarily used professional judgment for developing the calibration constrains for the preferred averages and the limits for clipping for specific storage and specific yield used by PEST. Table 4-7 lists the limits used to clip the storage parameters – these values were selected to minimize the range in the storage values while achieving an average between the lower and upper bounds for the preferred storage value average for each zone.

Table 4-7. Cut-off limits imposed on storage parameters assigned to each grid cell.

Storage Parameter		Aquifer Zone				
		2	3	5	6	7
Specific Storage (ft ⁻¹)	Lower	5e-06	1e-05	1e-06	1e-06	1e-06
	Upper	5e-05	1e-04	1e-05	1e-05	1e-05
Specific Yield (Dimensionless)	Lower	0.008	0.100	0.050	0.050	0.050
	Upper	0.030	0.150	0.100	0.100	0.100

4.2.3 Vertical conductance

The assignment of vertical conductance is a two-step process involving zonal parameterization followed by domain-wide multiplication by a pilot point-based multiplier field.

Zone-based vertical conductance between layers 1 and 2 is computed from zone-based horizontal conductivity ascribed to layers 1 and 2 through the process described above. Vertical conductance calculations are undertaken by a utility program named VCONTCALC. This was written as part of the current project and is documented in Appendix G. VCONTCALC obtains zonal horizontal conductivities from files *hk_lay1_zone.ref* and *hk_lay2_zone.ref*; these are multiplied by a global vertical anisotropy factor (which is estimated by PEST). VCONTCALC obtains thicknesses of layers 1 and 2 from files *thick1.ref* and *thick2.ref*. Its calculated vertical conductance is stored in a file named *vcont_lay1_zone.ref*.

PEST-estimated vertical conductance pilot point multiplier parameters are stored in file *vcont_lay1_mul.dat*. Kriging factors stored in file *factors_lay1.dat* are used (by the FAC2REAL utility) to create the actual multiplier array, this being stored in file *vcont_lay1_mul.ref*. Multiplication of the array stored in file *vcont_lay1_zone.ref* by that stored in file *vcont_lay1_mul.ref* to produce the vertical conductance array *vcont1.ref* used by MODFLOW is undertaken by the TWOARRAY utility. Appendix E describes the files used in vertical conductance parameterization. The preferred value, lower bound, and upper bound of the global vertical anisotropy used by PEST on its final run were set to 2.0e-5, 1e-7 and 2.5e-5, respectively.

4.3 Parameterization of recharge

For the purpose of recharge assignment, the model domain was subdivided into the same 11 zones used by Anaya and Jones (2009) and shown on Figure 4-9. The zonation of these 11 areas is supplied in file *rechzones.inf*. Recharge rates across each zone are calculated using a similar procedure to that used by Anaya and Jones (2009). This procedure includes multiplying the annual rainfall assigned to each grid by a recharge factor. A recharge factor of 0.01 means that 1% of the precipitation become recharge. Table 4-8 shows the preferred value and the upper and lower bounds for the average recharge factor for each zone used for the final PEST calibration. The preferred recharge factors in Table 4-8 are based on the recharge rates developed by Anaya and Jones (2009) for each zone. The lower and upper bounds are within 2% of the recharge factors developed by Anaya and Jones (2009).

Table 4-8. Preferred recharge factors and bounds for zones illustrated on Figure 4-9.

Recharge Zone		Preferred recharge factor	Lower bound	Upper bound
Number	Name			
1	Cenozoic Pecos – Alluvium North	0.01	0.0009	0.020
2	Buda Limestone or Del Rio Formation	0.01	0.0009	0.020
3	Edwards Group	0.02	0.016	0.04
4	Ogallala Sediments	0.03	0.02	0.06
5	Hill County Trinity Group	0.047	0.037	0.057
6	Edwards – Devil’s River Formation	0.05	0.04	0.07
7	Edwards-Trans-Pecos Basin and Range	0.06	0.045	0.08
8	Edwards – Stockton Plateau	0.08	0.06	0.10
9	Edwards – Maverick Basin	0.109	0.08	0.12
10	Cenozoic Pecos Alluvium South	0.05	0.04	0.07
11	No Recharge	0.0	NA	NA

The rainfall rate for each stress period is supplied on a cell-by-cell basis in arrays *rainfall_*.ref* where * is replaced by “ss” (for steady state), “av_1930s”, “av_1940s”, “av_1950s”, “av_1960s”, “av_1970s” (for use during the first five stress periods of the transient model), or “19[80-99]” (for use during each of the last 20 stress periods of the transient model). The INT2REAL utility is employed to create a real array of corresponding recharge factors based on the “integer-real correspondence file” *rf.irc*; a template of this file named *rf.tpl* allows PEST to provide values for zonal recharge factors during the parameter estimation process.

To allow introduction of intra-zonal variability to recharge, zone-based recharge factors are multiplied by a pilot-point-based multiplier field. Values assigned to pilot points (these being contained in file *rechmul.dat*) are adjustable by PEST (using a template of this file); the preferred value of all pilot point multipliers (enforced by Tikhonov constraints) is 1.0. Pilot points used in parameterization of this multiplier field are depicted in Figure 4-10; kriging factors (pre-calculated by PPK2FAC1) are housed in file *factors_all.dat*; the multiplier real array (recalculated by FAC2REAL on each occasion that the model is run) is written to file *rechmul.ref*. The TWOARRAY utility then multiplies this by zone-based recharge multipliers to obtain the total recharge factor array, this being named *rechmul_total.ref*. The TWOARRAY utility is used again to multiply pertinent rainfall arrays by these cell-based recharge factors to obtain a recharge array for each stress period. These are named *rech_*.ref* where “*” is variously “*ss*”, “*av_1930s*”, “*av_1940s*”, “*av_1950s*”, “*av_1960s*”, “*av_1970s*”, or “*19[80-99]*”.

4.4 Boundary condition parameterization

4.4.1 General head boundary conductance

The disposition of *general head boundaries* within the model domain is depicted in Figures 3-7 and 3-8. For the purpose of model calibration, the areal distribution of conductance of these cells was represented using pilot points. Thus conductance values were assigned by PEST to pilot points, and these values were spatially interpolated to *general head boundary* cells using kriging. Pilot points for layers 1 and 2 are housed in files *ghb_lay1.pts* and *ghb_lay2.pts* for which template files named *ghb_lay1.tpl* and *ghb_lay2.tpl* were produced so that PEST could provide these points with conductance values prior to each model run. Kriging factors (pre-calculated using the PPK2FAC1 utility) are housed in files *factors_ghb_lay1.dat* and *factors_ghb_lay2.dat*.

Pilot points were subdivided into groups for the purpose of assigning conductances to different groups of *general head boundary* cells within the model domain. Spatial interpolation from points belonging to one group was restricted to *general head boundary* cells belonging to the same group. Pilot points and *general head boundary* cells belonging to matching groups are assigned the same colors in Figures 4-11 and 4-12. In these figures, the same color is used to represent different groups of general head boundaries in different layers.

The FAC2REAL utility creates real arrays named *ghb_lay1.ref* and *ghb_lay2.ref* of *general head boundary* conductance in each of layers 1 and 2, respectively. The REAL2COND utility (written for the current project and documented in Appendix G) writes a MODFLOW *general head boundary* input file using these conductances, replacing those found in the *general head boundary* file pertaining to the original TWDB version of the ETPV GAM, this being stored in file *ghb.dat.keep*. Note that both the steady state and transient model use this same input file.

Tikhonov-enforced preferred values for *general head boundary* conductances are provided in Figures 4-11 and 4-12. These were obtained from a preliminary calibration exercise in which all parameters were varied on a zonal basis. Bounds on *general head boundary* conductances ranged between 1e-10 and 5,000 ft²/day.

4.4.2 Stream conductance

Pilot points were used for parameterization of conductances employed by the MODFLOW *streamflow routing* package. The locations of these points are provided in Figure 4-13. These are housed in file *strcond.pts*, for which a template file named *strcond.tpl* was created for the purpose of PEST-based parameter adjustment.

On the basis of PPK2FAC1-generated kriging factors contained in file *factors_str1.dat*, the FAC2REAL utility writes a real array file named *strcond.ref* containing *stream* conductances on each occasion that the model is run. A utility program named REALSTR2CND then builds the *streamflow routing* package input file for both the steady state and transient components of the ETPV GAM, this being named *str1.dat*. In doing this, it modifies conductances provided in the *streamflow routing* package input file for the original ETPV GAM, this having been renamed to *str1.dat.keep*.

As described in Appendix G, REALSTR2CND requires input from the keyboard. One of its required inputs is the roughness to apply to all stream cells. This parameter is adjustable by PEST. Prior to each model run PEST writes a REALSTR2CND keyboard input file (named *realstr2cnd.in*) on the basis of a template of this file (named *realstr2cnd.tpl*), in which the appropriate roughness parameter for that model run is recorded.

The preferred value for all *stream* conductance pilot-point parameters was set to 20,000 ft²/day. The lower bound was set very low to 1.0e-10 ft²/day. The upper bound was set to 30,000 ft²/day. A preferred roughness parameter value of 3.895e-7 ft^{-1/3}/day was provided; the bounding interval was set very wide.

4.4.3 Drain conductance

For representation of spatial variation of *drain* conductance, the spatial zonation pattern illustrated in Figures 4-14 and 4-15 is employed. Ten zones are employed for representation of layer 1 conductances while nine zones are employed for representation of layer 2 conductances. Zonation is defined in integer arrays *drain_lay1.inf* and *drain_lay2.inf*. These, together with integer-real-correspondence files *draincond_lay1.irc* and *draincond_lay2.irc* serve as input files to the INT2REAL utility, on which basis conductance real arrays *draincond_lay1.ref* and *draincond_lay2.ref* are written. PEST-based adjustment of conductance within each zone is based on template files of the integer-real correspondence file. The MODFLOW *drain* package input file *drn.dat* (which serves as an input file for both the steady state and transient model components) is written by the REAL2CND utility which replaces conductances in file *drn.dat.keep* (the *drain* package input file for the original ETPV GAM) with conductances cited in the INT2REAL-produced conductance real array files *draincond_lay1.ref* and *draincond_lay2.ref*. Appendix E provides further file details.

A preferred value of 1000 ft²/day was assigned to all *drain* conductances. This conductance is based on the values in the ETPV GAM. Upper and lower bounds were set to 1e-10 ft²/day and 5000 ft²/day, respectively.

4.5 Parameterization of pumping rates

As described in a previous section, the original ETPV GAM is comprised of a steady-state model for 1980 and a transient model for 1980 to 2000. Pumping for all of the model is supplied in a standard MODFLOW *well* package input file. The revised ETPV GAM is comprised of two sub-models. The first of these is steady-state for 1930. No pumping is provided to this model as it is intended to simulate pre-development conditions. The second is a revised transient model for 1930 to 1980. During the stress periods spanning the years 1980 to 1999, the pumping well package for the revised ETPV GAM is identical to that provided in the original model for the same stress periods.

The first five 10-year stress periods of the revised transient model are intended to simulate a transition period between undeveloped conditions and those which prevailed at the start of 1980. Over this period a “standard” set of pumping rates is supplied; these are the same as those applied for the 1980 steady-state ETPV GAM, and are depicted in Figures 4-16 and 4-17. However, a PEST-adjustable “build-up factor” is applied to these pumping rates during each of the five 10-year stress periods spanning from 1930 to 1980.

4.5.1 *Estimated pumping rates from 1930 to 1980*

For the purpose of this study, we assume that pumping from 1930 to 1980 can be reasonably estimated based on irrigated crop acreage and an average application rate per acre. Our approach to develop estimated historical pumping rates includes the following steps:

1. Identify the counties that are most likely to have the greatest pumping from 1930 to 1980 from the TWDB irrigation surveys for 1958, 1964, 1969, and 1974.
2. For counties in Step 1 identified with high irrigation rates, assemble the historical acres for crops based on the agriculture reports prepared by the United States Bureau of the Census for the years 1930, 1935, 1940, 1945, 1950, 1954, 1959, and 1964.
3. To estimate pumping rates from crop acreage in Step 2, estimate a generic irrigation rate per acre for selected counties using the TWDB survey information from 1958 to 1974.
4. For those counties with the high irrigation rates, merge the irrigation values developed in Steps 1 and 3 to create an estimated pumping record from 1930 to 1980.
5. For the purposes of estimating general trends in the total historical pumping rates from 1930 to 1980 divide the counties into 16 groups based on the TWDB pumping rates for each county from 1980 to 2000, our estimated total pumping rates from 1930 to 1980 for the counties identified with high irrigation, the proximity of the counties to each other, and the perceived historical agricultural practices in each county.
6. Subject the best estimate of the average yearly pumping for each of the 16 groups of counties in Step 5 by a multiplier of the total pumping reported by TWDB for the group of counties for the year 1980. Average this multiplier over 10-year periods so that the number of multipliers is reduced from 50 to 5 for each group of counties.
7. Apply PEST to adjust the pumping multipliers for each decade as part of the transient model calibration process.

Step 1 Results: TWDB Pumping Rates from 1958 to 1974

Table 4-9 shows a summary of the irrigation pumping from the earliest TWDB surveys from 1958 to 1974. The table shows the eight counties with the greatest total pumping. These eight counties comprise over 95% of the total estimated pumping within the model domain. For this analysis the following counties were removed from contention, because although they report irrigation pumping, they pump predominantly from other aquifers: Andrews, Bexar, Culberson, Martin, Mason, Medina, Tom Green, and Uvalde.

Table 4-9. Summary of irrigation pumping for the model domain and the eight counties with the highest irrigation based on the TWDB surveys.

County	Year(s)				
	1958	1964	1969	1974	1958-1974
All counties	733,935	887,001	673,094	651,190	294,5220
Reeves	335,168	402,017	327,376	303,162	1,367,723
Pecos	328,015	362,405	192,556	177,455	1,060,429
Glasscock	11,597	24,577	34,185	55,103	125,462
Midland	24,866	14,847	33,429	36,264	109,406
Reagan	4,270	15,334	15,434	14,531	49,569
Ward	5,051	16,700	13,049	12,397	47,197
Kinney	2,427	10,147	12,333	10,820	35,727
Schleicher	4,500	7,623	4,787	1,951	18,861

Step 2 Results: U.S. Agricultural Census Reports 1930 to 1964

Figures 4-18 and 4-19 show total agricultural acreage from 1930 to 1964 based on the US Agricultural Census for seven of the eight counties listed in Table 4-9. The reported acreage represents the sum of sorghum, cotton, and corn acreages. The total acreages do not necessarily agree with results from the Texas Water Development Board. In general, the discrepancies are small with the exception of Pecos County, where the discrepancy is huge. The discrepancy may be attributed to the fact that Census reports simple acreage, while the TWDB reports “irrigated acreage”. We did not include crops like hay because we thought it unlikely to be irrigated.

Step 3 Results: Estimated Application Rates

Table 4-10 shows the application rate from the TWDB surveys from 1958 to 1974 for the eight counties listed in Table 4-8. The table lists the rates in acre-ft/acre for each county, and provides an overall average of that metric. Pecos, Reeves and Ward counties have the highest estimated application rates. This may be due to the type of crop being irrigated, or the general practices in the area. Unfortunately, we have little solid data on earlier fractions of irrigated crops. We do know that cotton was a dominant crop, but we do not know what proportion of it was irrigated, outside of the Pecos/Reeves region.

Table 4-10. Application rates (acre-ft/acre) from the TWDB irrigation surveys.

County	1958	1964	1969	1974	Average
Glasscock	1.1	1.4	1.5	2.0	1.5
Kinney	1.4	1.9	1.9	1.8	1.8
Midland	2.0	1.3	1.2	1.3	1.4
Pecos	3.0	3.1	3.7	3.5	3.3
Reagan	1.6	1.5	0.9	1.3	1.3
Reeves	3.9	3.6	4.1	4.1	3.9
Schleicher	1.9	1.9	1.1	0.8	1.4
Ward	2.4	3.3	3.4	4.1	3.3

Step 4 Results: Estimated Pumping Curves for 1930 to 2000

To generate overall curves, we used the application rates and the acreages from the U.S. Census for the period from 1944 to 1958, then used the TWDB irrigation survey estimates up to 2000. Before 1944, we scaled the “fraction irrigated” estimate linearly from 1930, to account for the greater proportion of dryland farming that would have been likely at that time. The final pumping curves are shown in Figures 4-20 and 4-21.

There is significant uncertainty in these pumping estimates, especially prior to 1958 when the TWDB surveys are not available. In general, if pumping increased in the 1940s and then decreased some time after that, it is considered to be a plausible curve. A comparison of the pumping curve for Reeves County to Figure 9 in Ashworth (1990) indicates that the peak pumping predicted by the current method is about 20% lower, but most of the curve is similar. Whether the pumping in Pecos should have increased sooner (see the difference in acreages estimated by the U.S. Census and the TWDB surveys in the previous section) is unknown. This could make a large difference in Pecos County in the years from approximately 1945 to 1955. There is considerable leeway for increasing pumping in that county over that decade, possibly by as much as 200,000 AFY.

Step 5 Results: Division of the Counties into 16 Groups

Figure 4-22 shows the 16 zones into which we grouped the counties. Counties such as Reeves and Pecos counties that have large estimates of historical pumping were not grouped with other counties. Based primarily on our analysis of the TWDB irrigation surveys, we grouped the counties based on apparent similarities in agricultural practices, total pumping amounts, or historical trends in pumping.

Step 6 Results: Division of the Counties into 16 Groups

Based on analysis of all of the assembled pumping information, we created the listing of preferred build-up factors in Table 4-11 for the 16 county groups.

Table 4-11. Preferred values for pumping build-up factor parameters.

Zone	Time period				
	1930 - 1940	1940 - 1950	1950 - 1960	1960 - 1970	1970 – 1980
1	0.2	0.4	0.6	0.8	1.0
2	0.1	0.87	2.79	5.0	2.93
3	0.1	0.49	5.0	5.0	1.47
4	4.0	18.0	2.3	4.0	3.0
5	0.2	0.4	0.6	0.8	1.0
6	0.2	0.4	0.6	0.8	1.0
7	0.7	3.3	2.0	1.06	1.3
8	0.1	0.4	0.4	0.7	1.1
9	0.2	0.4	0.6	0.8	1.0
10	0.2	0.4	0.6	0.8	1.0
11	0.2	0.4	0.6	0.8	1.0
12	0.2	0.4	0.6	0.8	1.0
13	0.2	0.4	0.6	0.8	1.0
14	0.2	0.4	0.6	0.8	1.0
15	9.1	18.0	18.0	7.8	1.7
16	0.2	0.4	0.6	0.8	1.0

Step 7: PEST Application

To allow further spatial variability, pumping rates are subject to multiplication by pilot-point-based multiplier arrays. Different multiplier arrays are provided for the two different layers. However the same layer-specific, pumping array is employed for all transient model stress periods, both during and after the build-up period.

The MODFLOW *well* package input file for the transient component of the ETPV GAM is written by the WELLBUILD2 utility; this utility was written specifically for the present project. WELLBUILD2 reads the following files:

- Multiplier array files *welmul_lay1.ref* and *welmul_lay2.ref*. These are written by FAC2REAL on the basis of pilot point multiplier values contained in files *welmul_lay1.dat* and *welmul_lay2.dat*. Pilot point locations are the same as those depicted in Figures 4-7a and 4-8b; kriging factors are contained in files *factors_lay1.dat* and *factors_lay2.dat*.
- A “standard” *well* pumping file for use during the first five stress periods. Pumping rates listed in this file are subject to multiplication by zone-based build-up factors and pilot-point-based multiplier arrays. This file is named *standard.wel*.
- A file containing a MODFLOW/MT3D-compatible integer array defining the zones on which build-up factors are based (pictured in Figure 4-22). This file is named *buildup_zones.inf*.
- A file containing build-up factors to apply in each zone during each build-up stress period. This file is named *buildup_factors.dat*.

TWDB Report ##: Final – Application of PEST to Re-Calibrate
the Groundwater Availability Model for the Edwards-Trinity (Plateau) and Pecos Valley Aquifers

- A file containing pumping rates for use during the 20 stress periods comprising the calibration period. This file is named *post-buildup.wel*. Pumping rates listed in this file are subject to multiplication by pilot-point-based multiplier arrays.

After having read and processed all of this data, WELBUILD2 writes the MODFLOW *well* package input file for the transient component of the ETPV GAM. This is named *wel_tr.dat*.

TWDB Report ##: Final – Application of PEST to Re-Calibrate
the Groundwater Availability Model for the Edwards-Trinity (Plateau) and Pecos Valley Aquifers

TWDB Report ##: Final – Application of PEST to Re-Calibrate
the Groundwater Availability Model for the Edwards-Trinity (Plateau) and Pecos Valley Aquifers

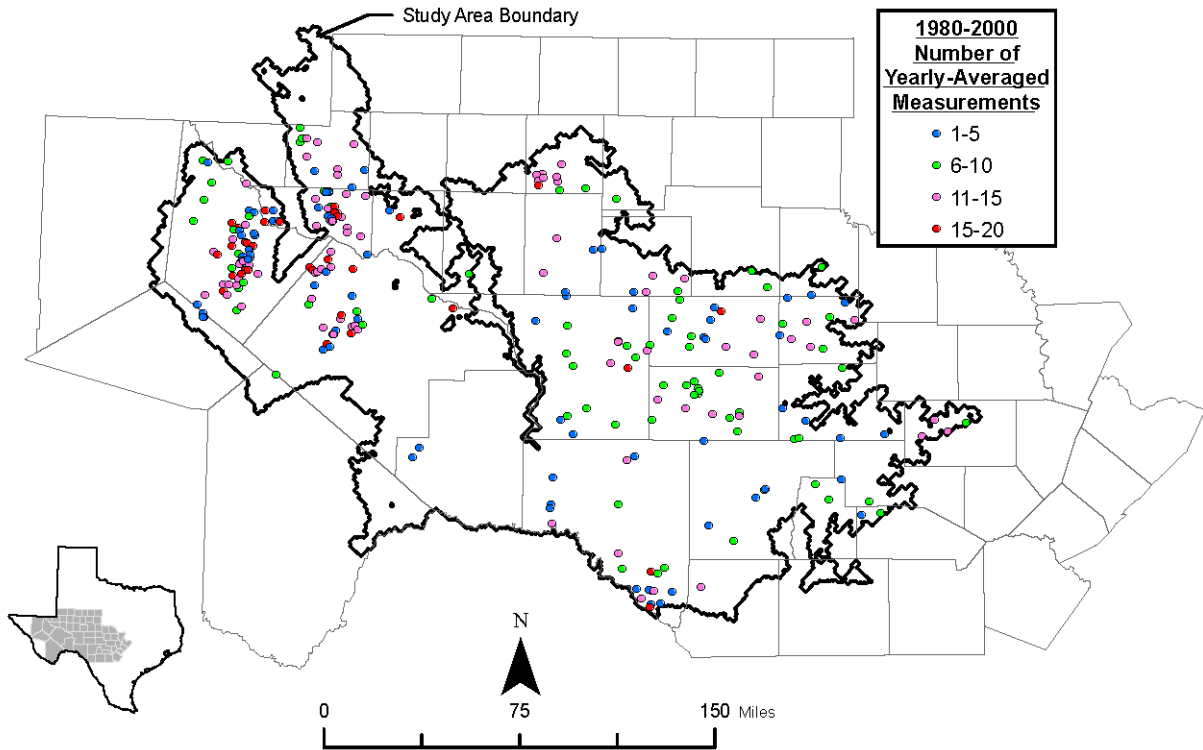


Figure 4-1. Wells in layer 1 for which head observations were employed in the calibration process.

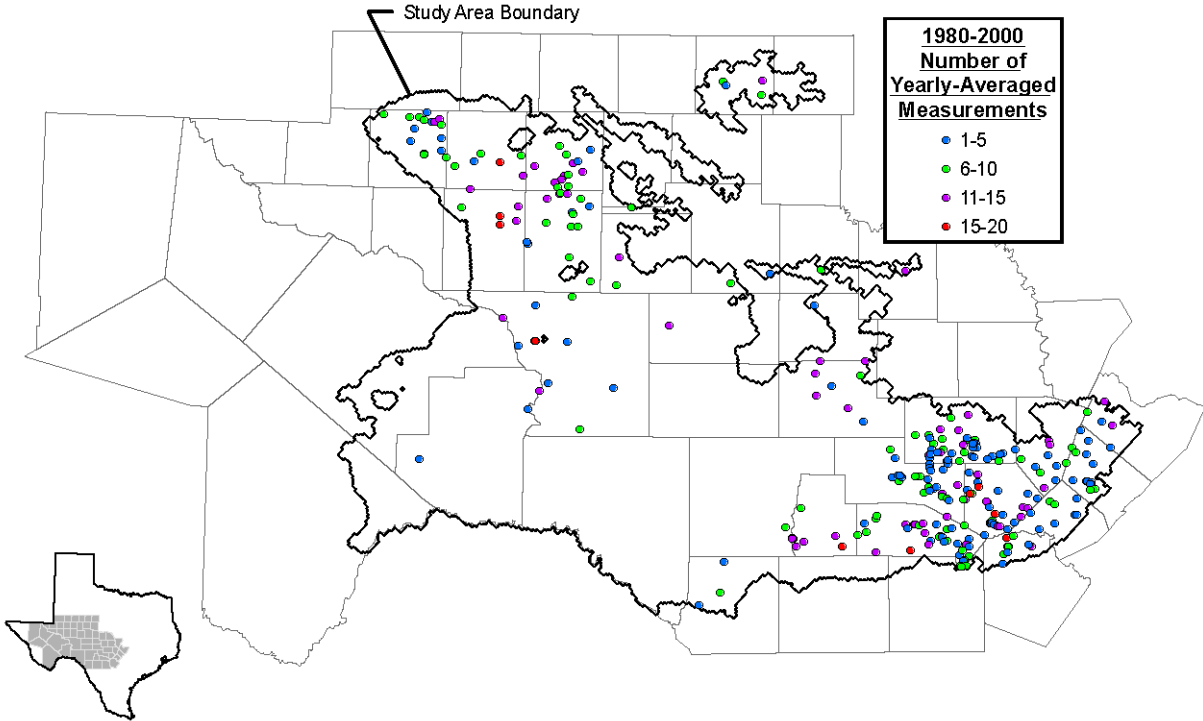


Figure 4-2. Wells in layer 2 for which head observations were employed in the calibration process.

TWDB Report ##: Final – Application of PEST to Re-Calibrate the Groundwater Availability Model for the Edwards-Trinity (Plateau) and Pecos Valley Aquifers

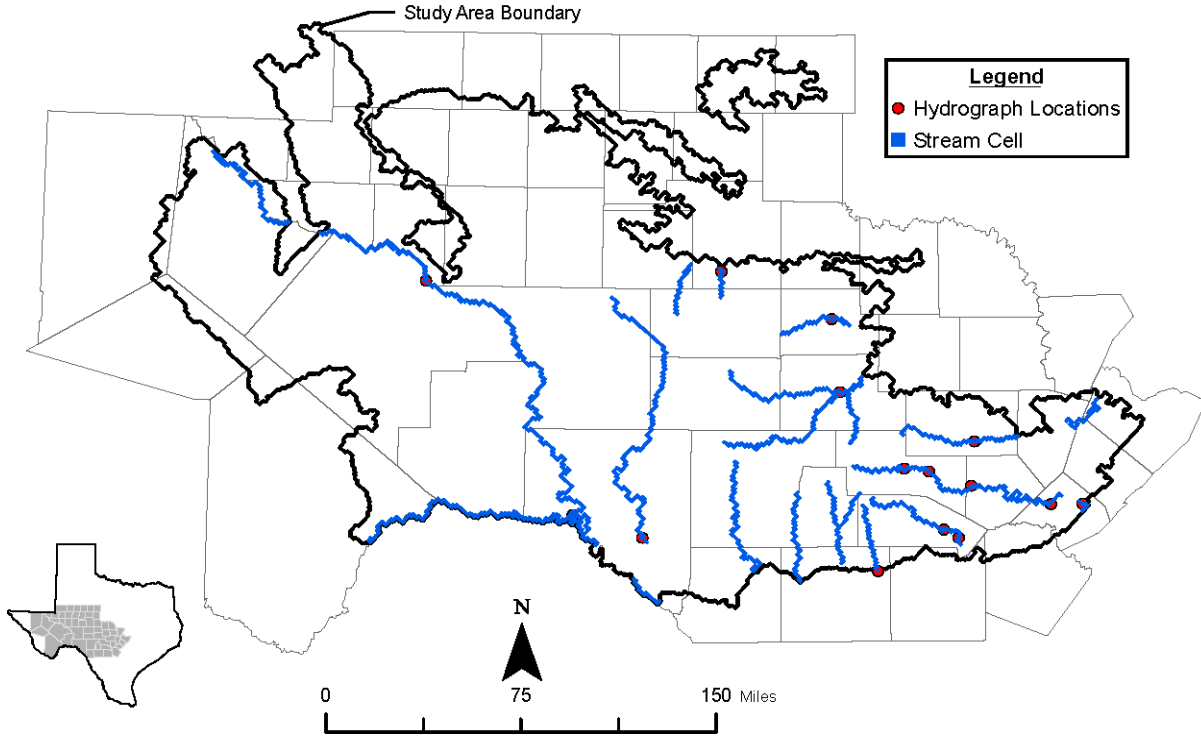


Figure 4-3. Stream reaches (light blue) and streamflow measurement points (red) featured in the calibration dataset.

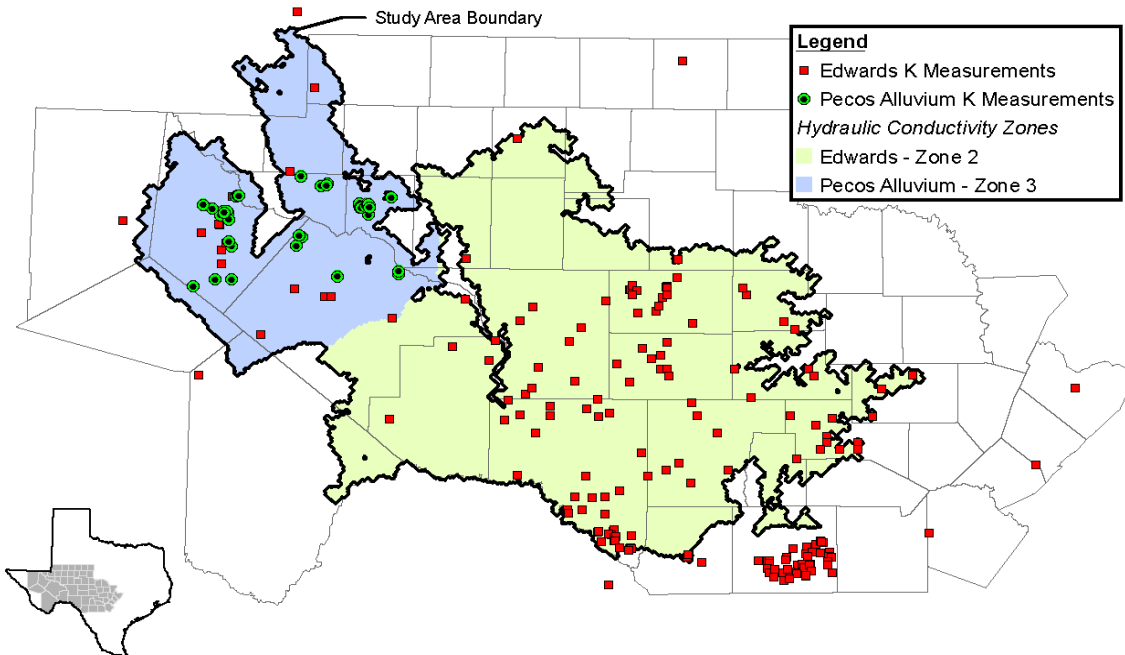


Figure 4-4. Distribution of pumping test locations used in hydraulic property assignment of layer 1.

TWDB Report ##: Final – Application of PEST to Re-Calibrate
the Groundwater Availability Model for the Edwards-Trinity (Plateau) and Pecos Valley Aquifers

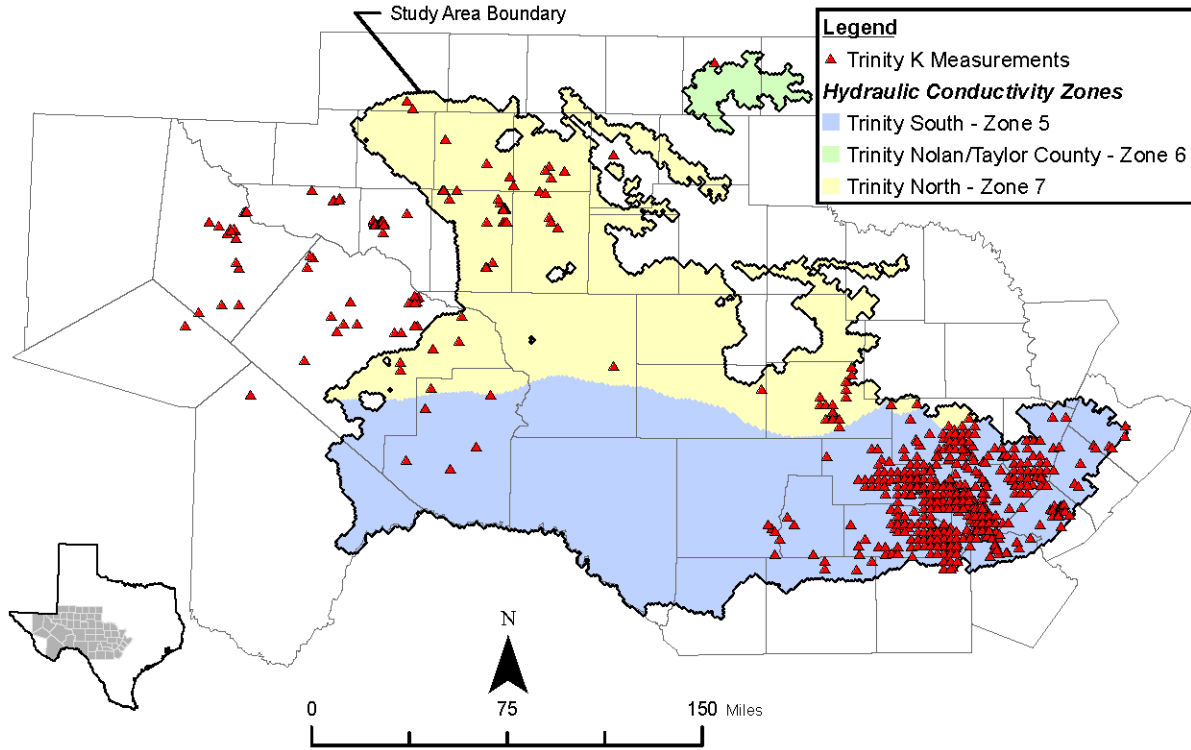


Figure 4-5. Distribution of pumping test locations used in hydraulic property assignment of layer 2.

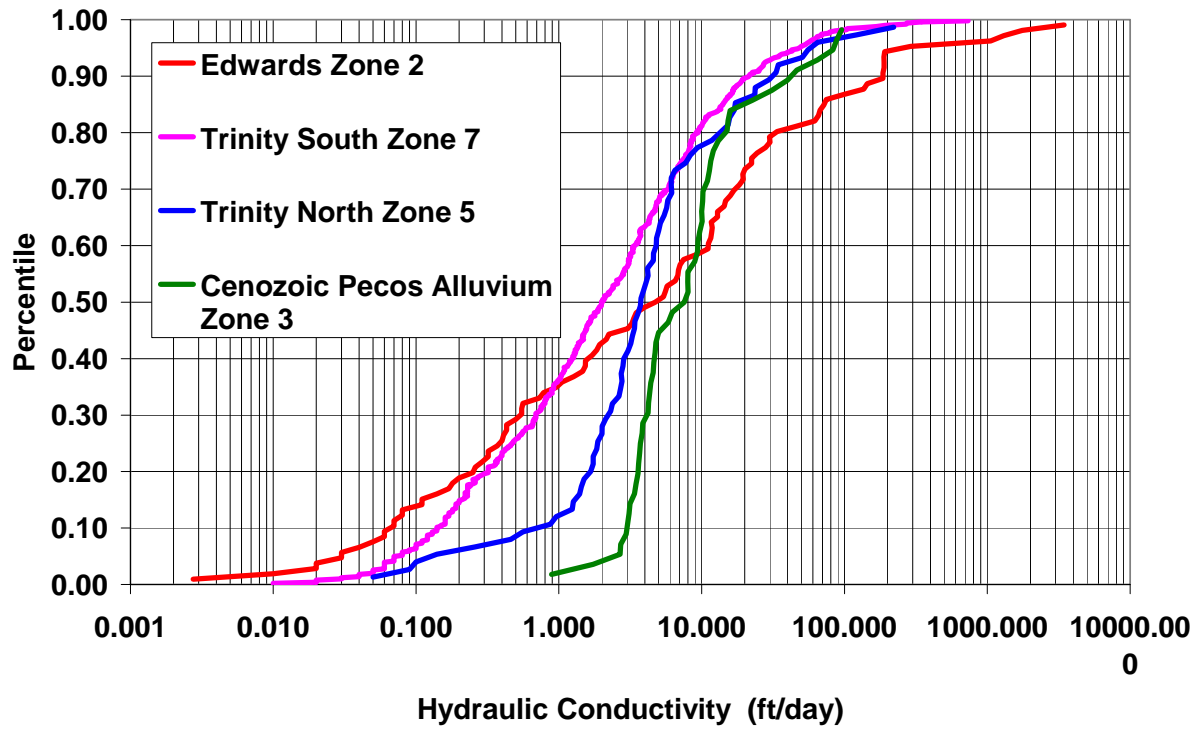


Figure 4-6. Cumulative distribution for conductivity values in each zone.

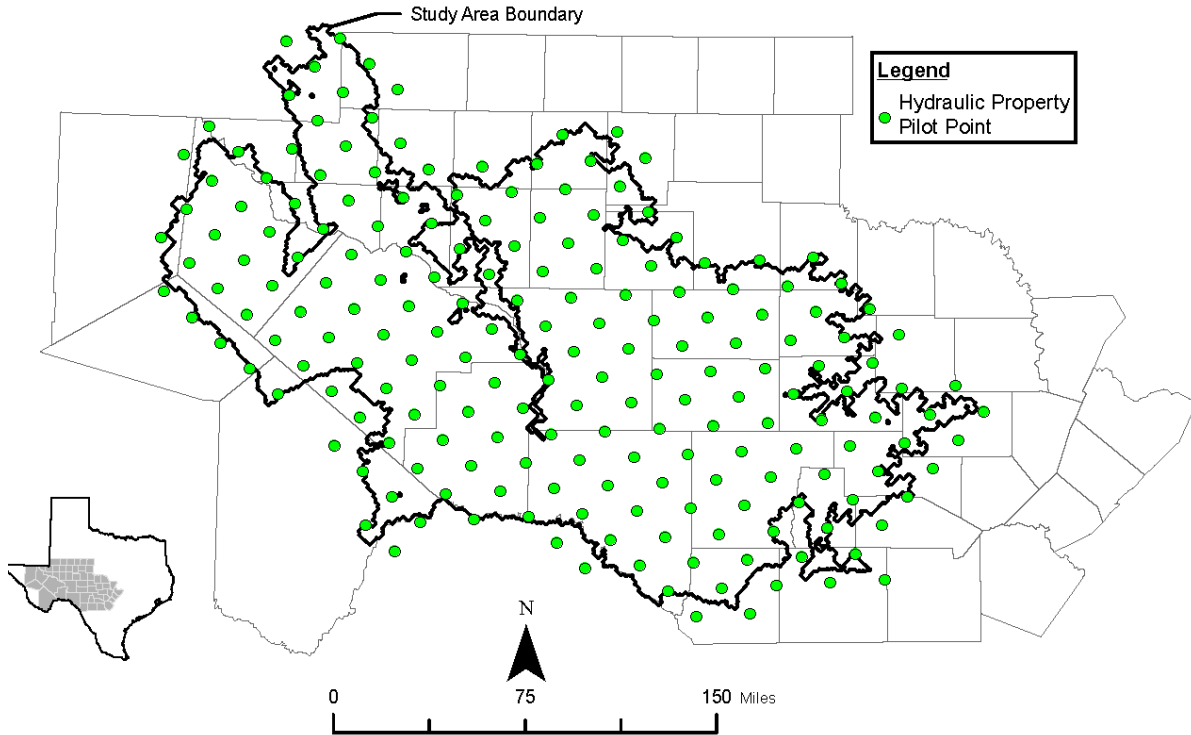


Figure 4-7. Distribution of hydraulic property pilot points used for parameterization of layer 1.

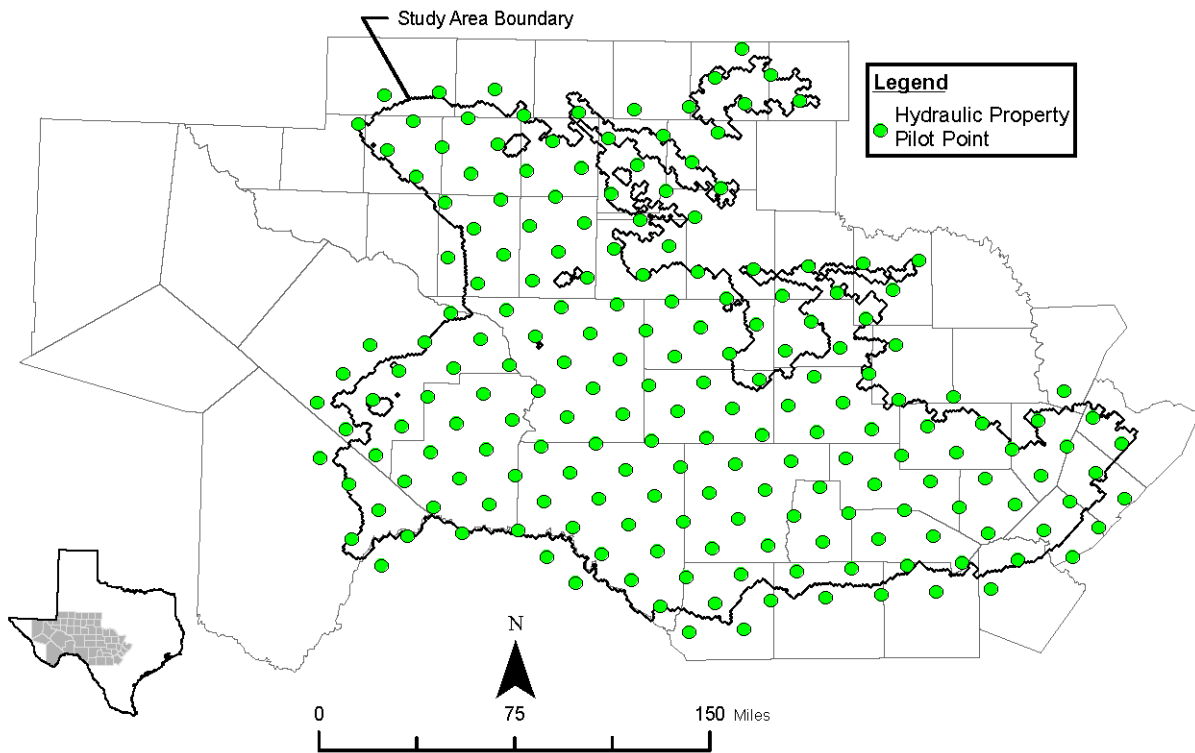


Figure 4-8. Distribution of hydraulic property pilot points used for parameterization of layer 2.

TWDB Report ##: Final – Application of PEST to Re-Calibrate
the Groundwater Availability Model for the Edwards-Trinity (Plateau) and Pecos Valley Aquifers

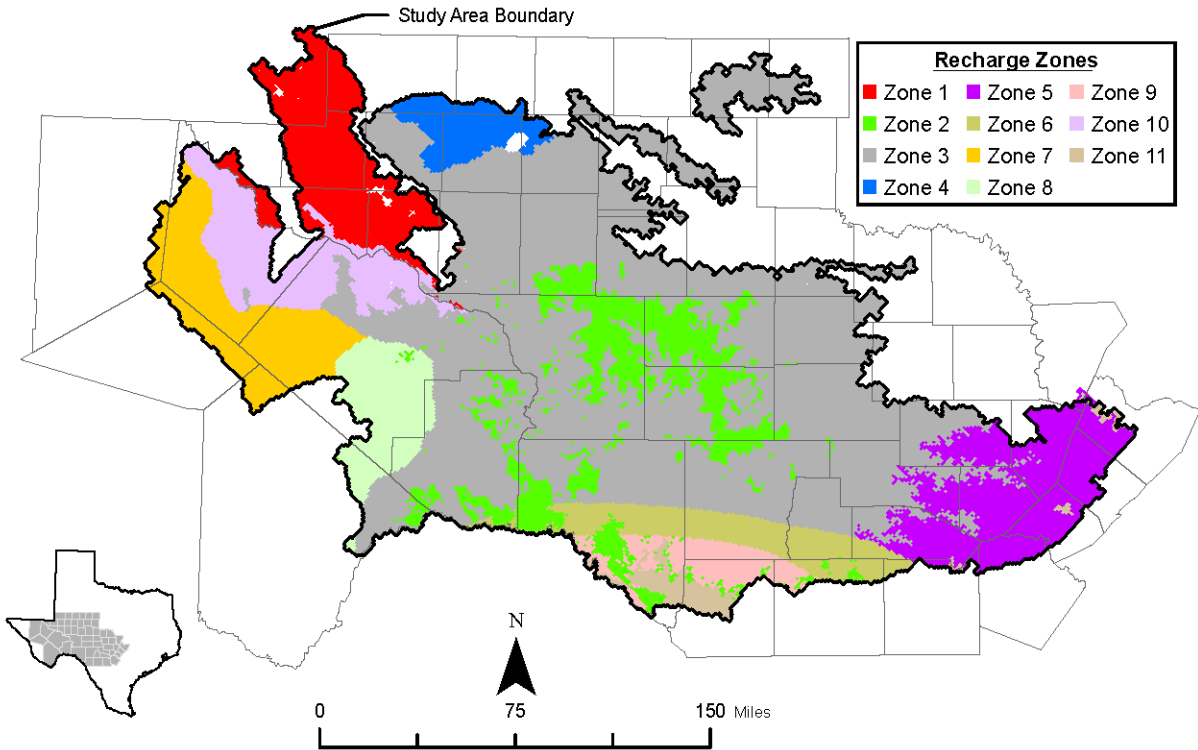


Figure 4-9. Recharge zonation and zone numbering scheme.

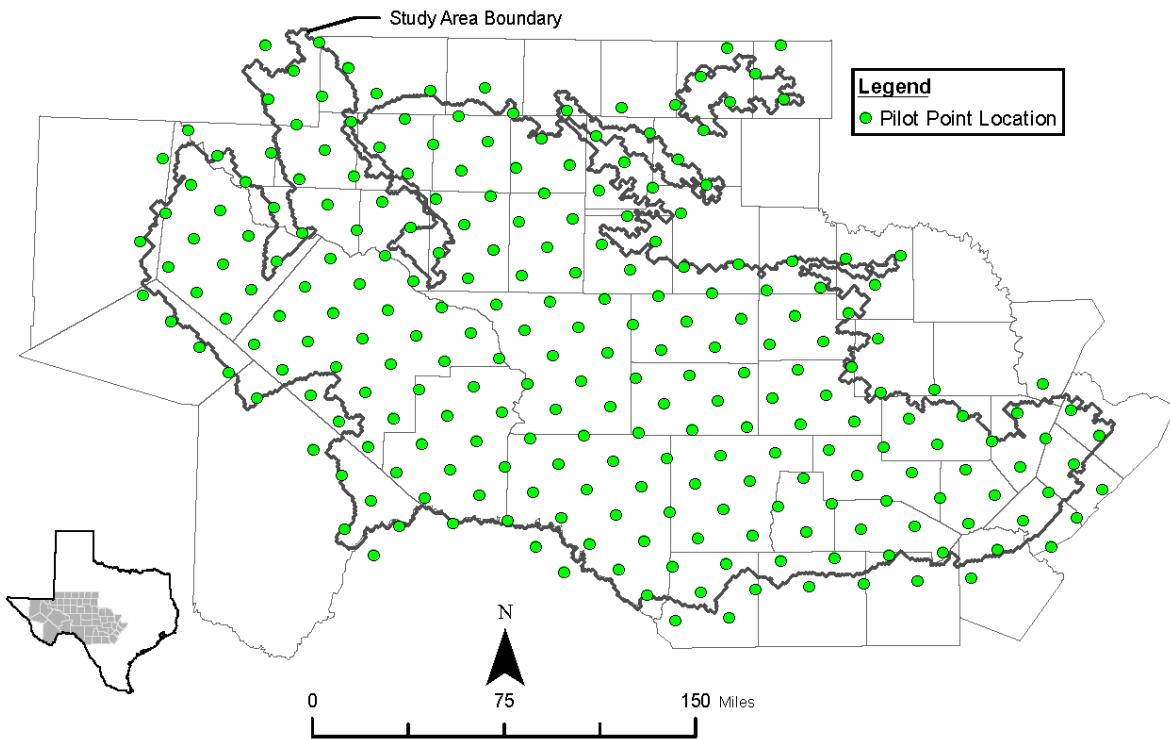


Figure 4-10. Pilot points used in recharge parameterization.

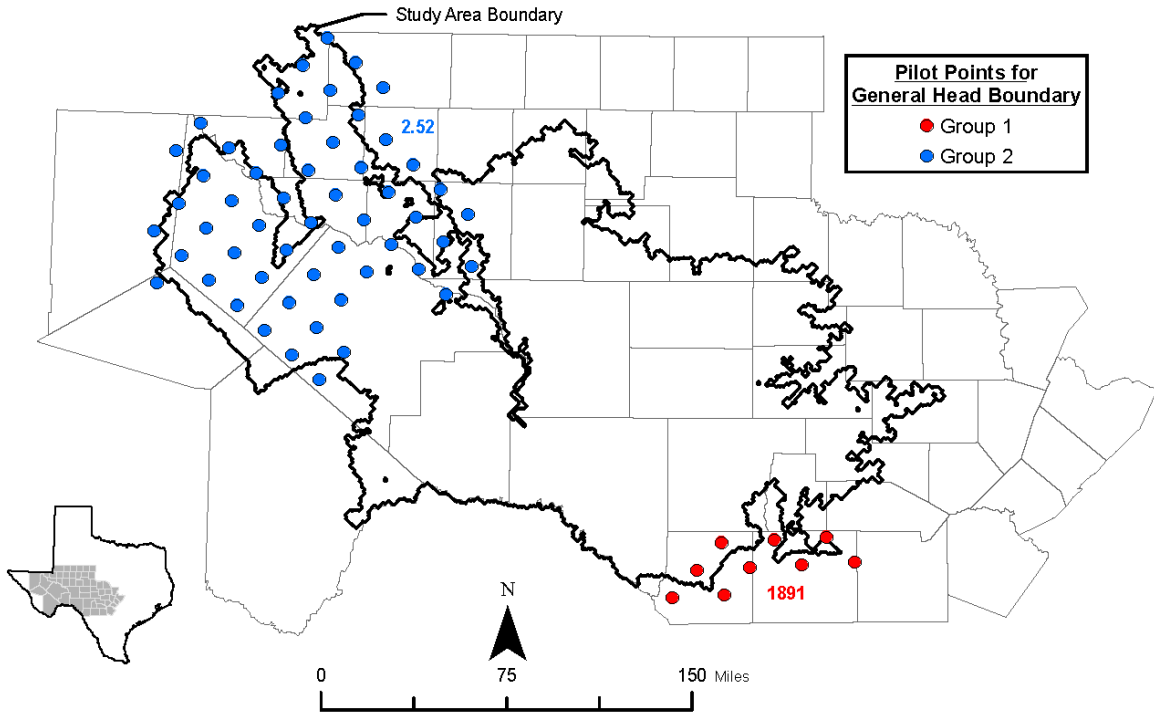


Figure 4-11. General head boundary pilot points in layer 1. Preferred conductances in ft^2/day are also depicted.

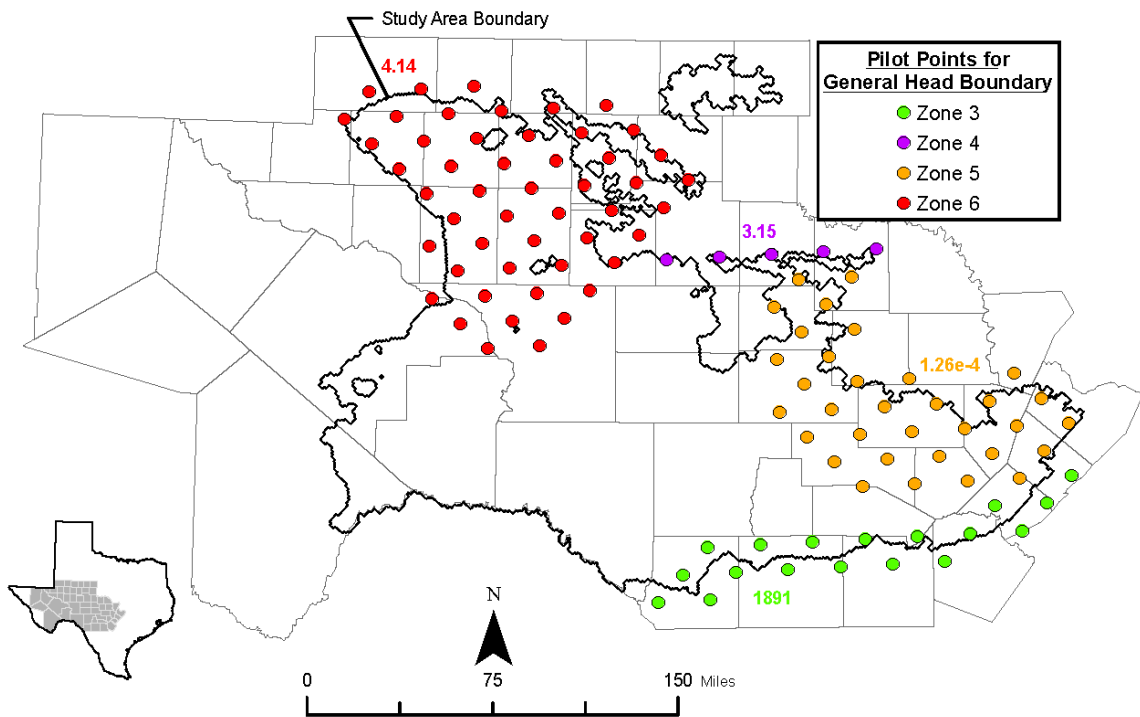


Figure 4-12. General head boundary pilot points in layer 2. Preferred conductances in ft^2/day are also depicted.

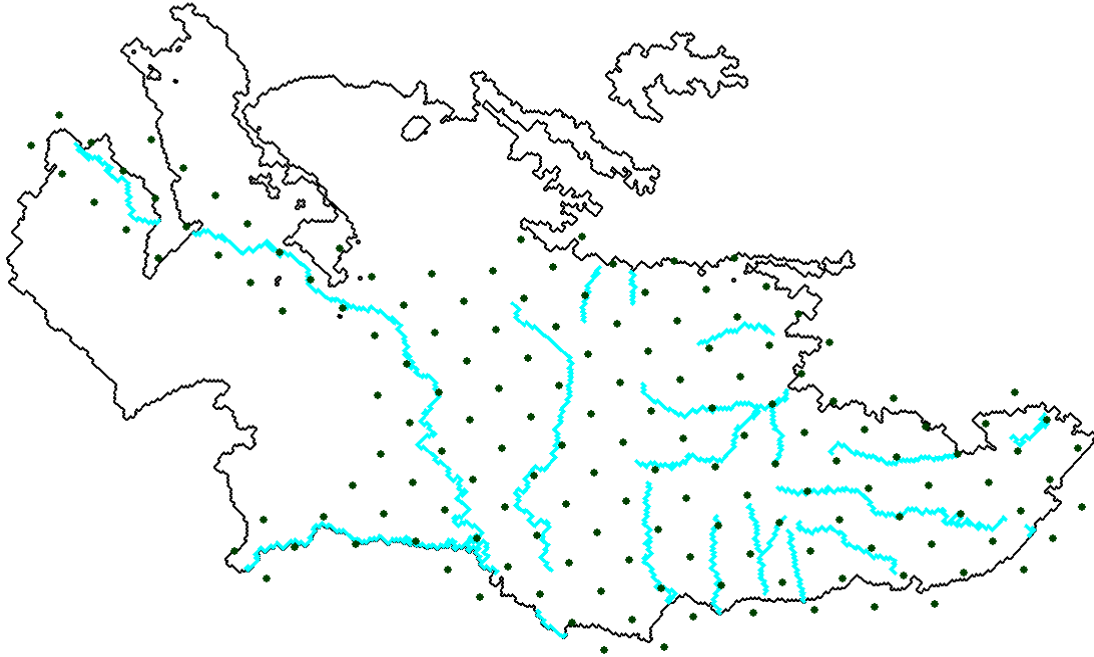


Figure 4-13. Pilot points used in stream conductance parameterization.

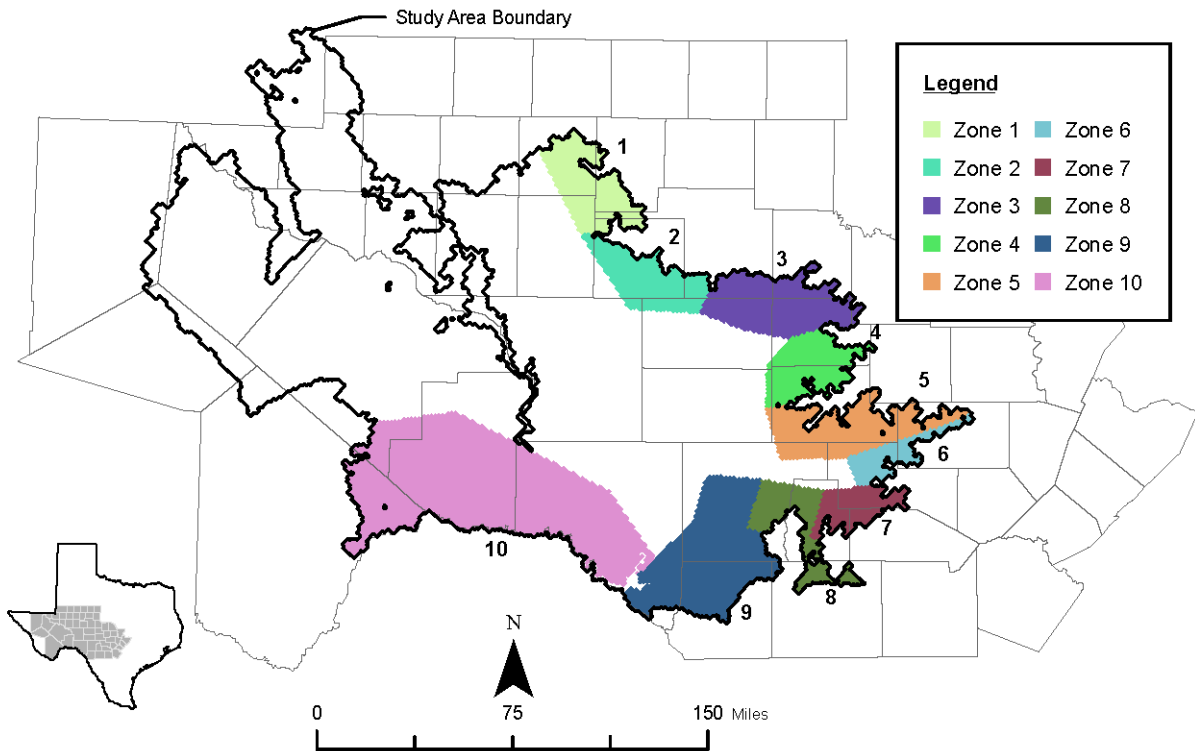


Figure 4-14. Zones and zone numbers used in estimation of drain conductance for layer 1.

TWDB Report ##: Final – Application of PEST to Re-Calibrate the Groundwater Availability Model for the Edwards-Trinity (Plateau) and Pecos Valley Aquifers

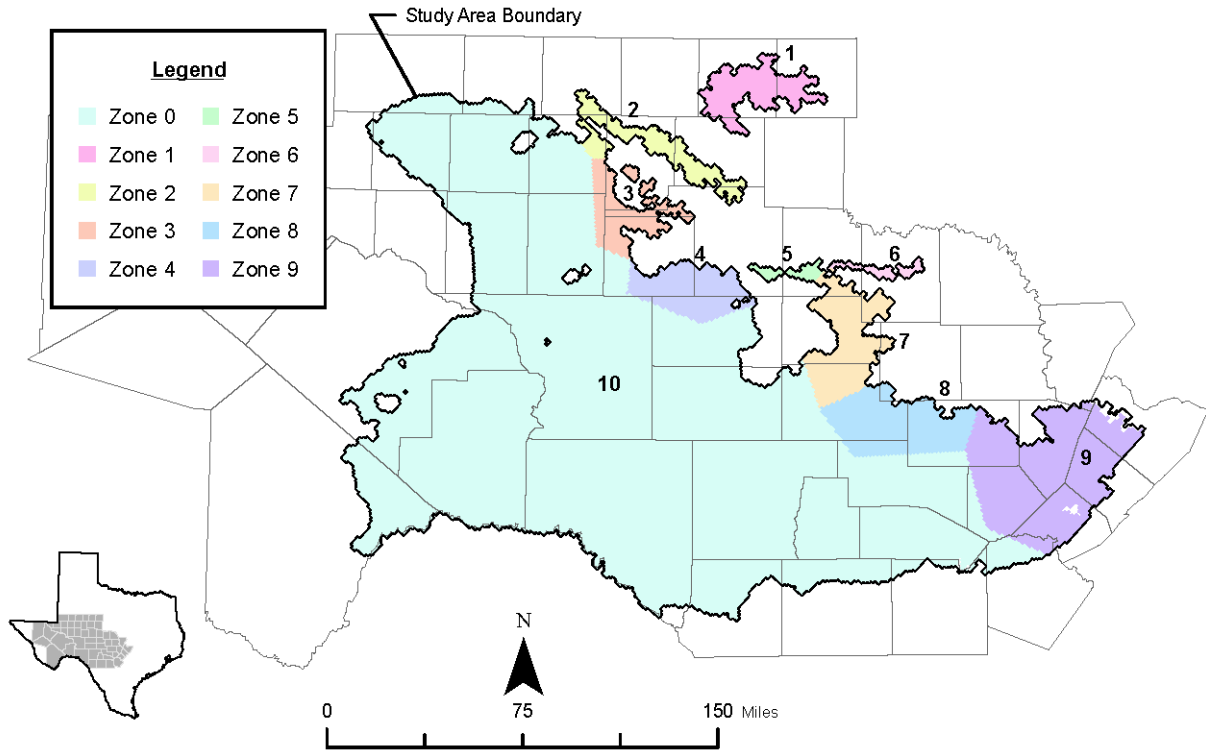


Figure 4-15. Zones and zone numbers used in estimation of drain conductance for layer 2.

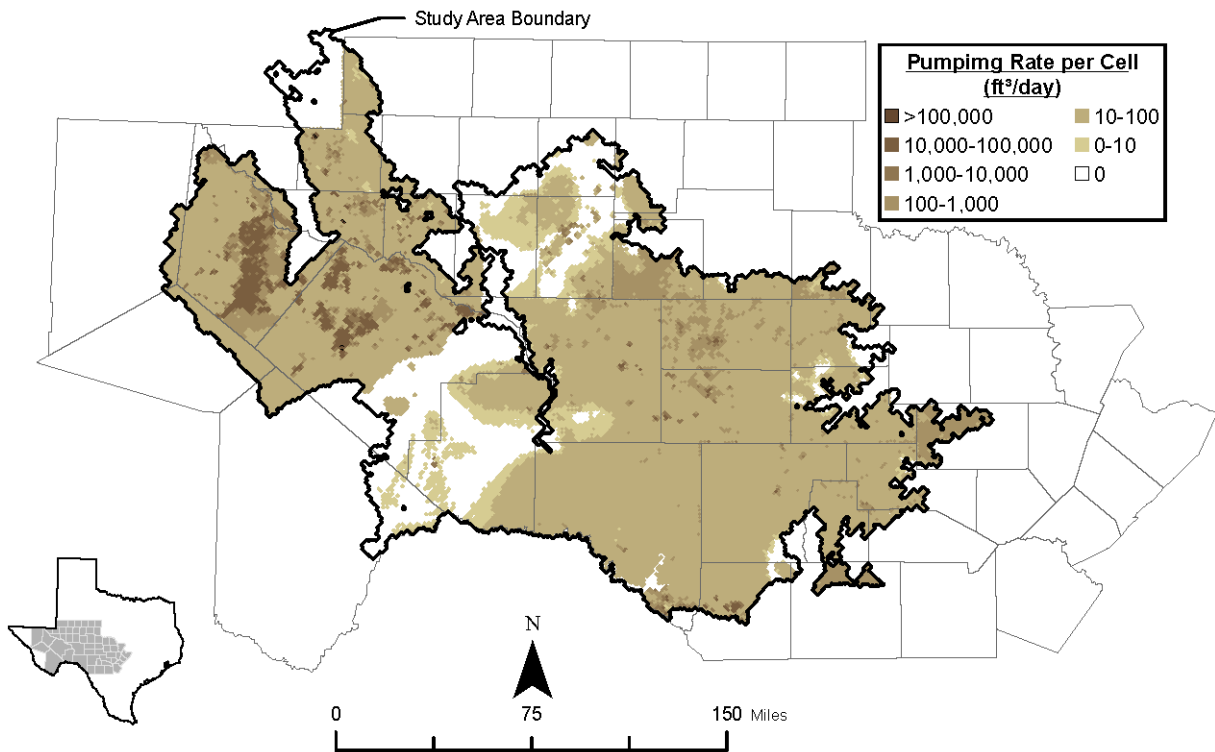


Figure 4-16. “Standard” cell pumping rates in layer 1 used in five precalibration, 10-year stress periods.

TWDB Report ##: Final – Application of PEST to Re-Calibrate the Groundwater Availability Model for the Edwards-Trinity (Plateau) and Pecos Valley Aquifers

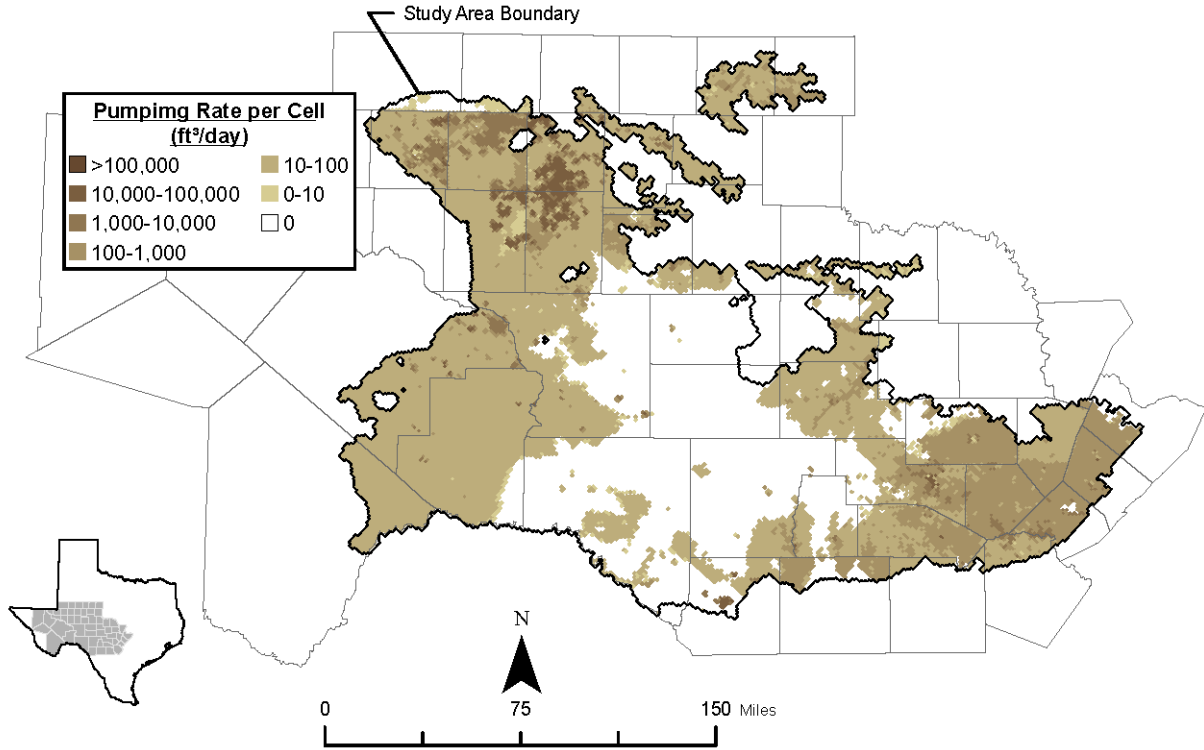


Figure 4-17. “Standard” cell pumping rates in layer 2 used in five precalibration, 10-year stress periods.

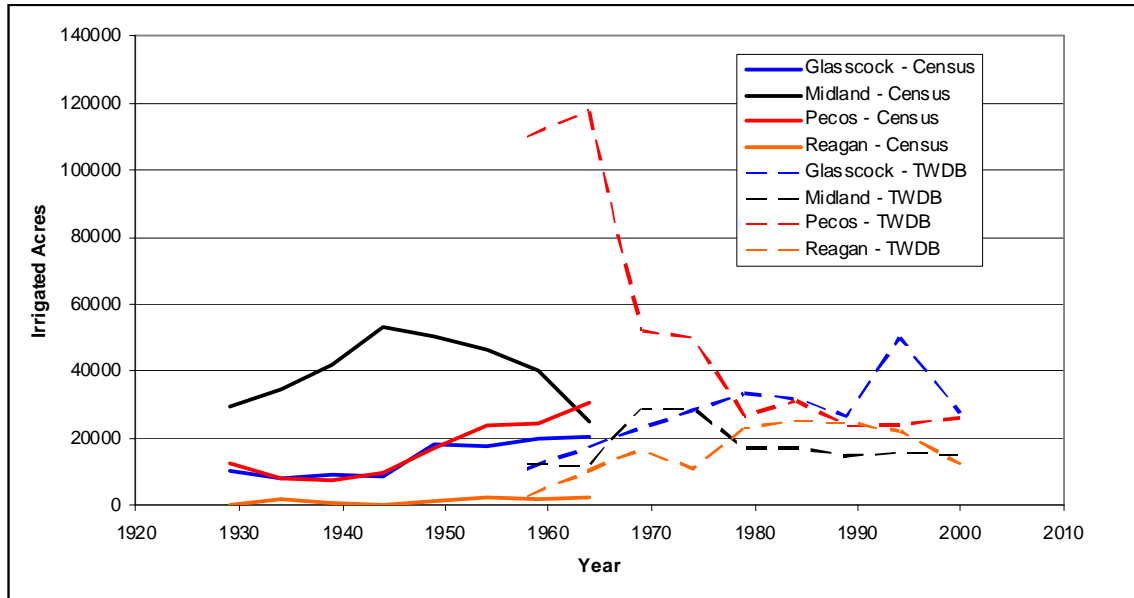


Figure 4-18. Irrigated acres reported by the U.S. Agricultural Census from 1930 to 1964 and by TWDB from 1958 to 2000 for Glascock, Midland, Pecos, and Reagan Counties.

TWDB Report ##: Final – Application of PEST to Re-Calibrate the Groundwater Availability Model for the Edwards-Trinity (Plateau) and Pecos Valley Aquifers

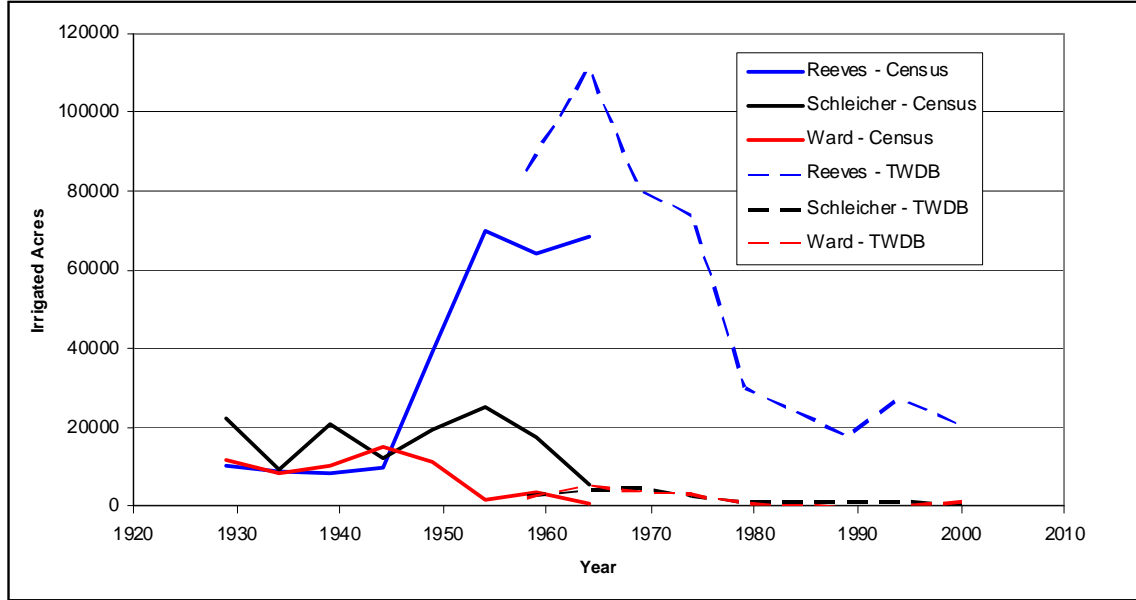


Figure 4-19. Irrigated acres reported by the U.S. Agricultural Census from 1930 to 1964 and by TWDB from 1958 to 2000 for Reeves, Schleicher, and Ward Counties.

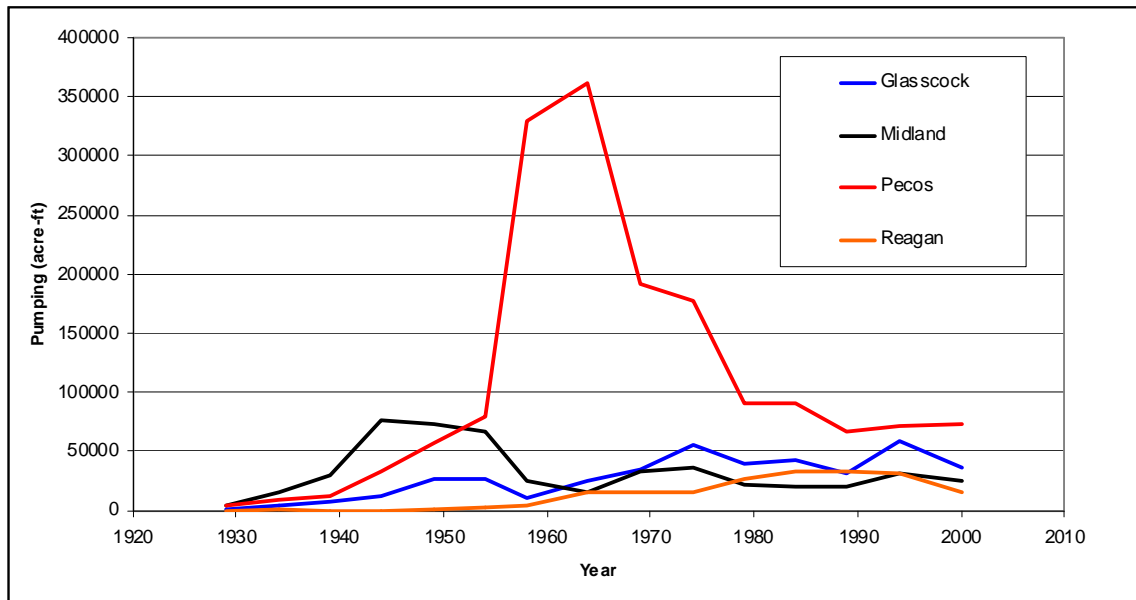


Figure 4-20. Estimated pumping from 1930 to 2000 for Glasscock, Midland, Pecos, and Reagan Counties based on a merger of U.S. Agricultural survey data and TWDB irrigation survey data.

TWDB Report ##: Final – Application of PEST to Re-Calibrate the Groundwater Availability Model for the Edwards-Trinity (Plateau) and Pecos Valley Aquifers

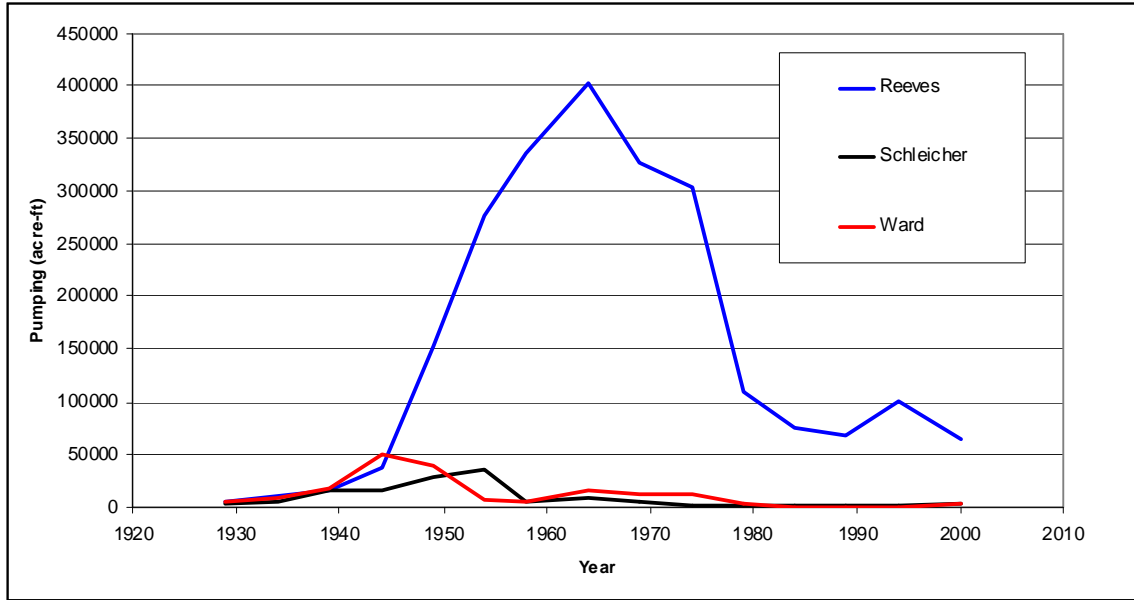


Figure 4-21. Estimated pumping from 1930 to 2000 for Reeves, Schleicher, and Ward Counties based on a merger of U.S. Agricultural survey data and TWDB irrigation survey data.

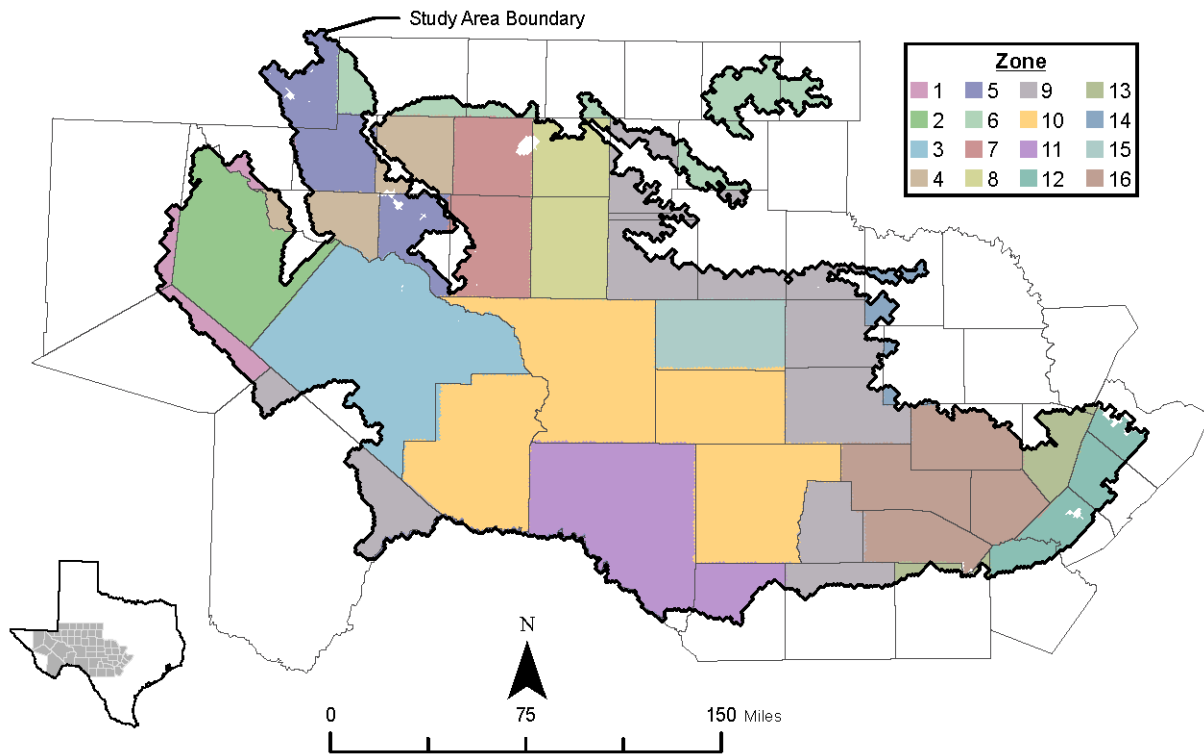


Figure 4-22. Zone definition for pumping rate build-up factor parameters.

TWDB Report ##: Final – Application of PEST to Re-Calibrate
the Groundwater Availability Model for the Edwards-Trinity (Plateau) and Pecos Valley Aquifers

5 Calibration specifications

5.1 Introduction

The previous section describes the parameterization scheme employed in the calibration process, and the observations on which basis parameters comprising this scheme are estimated. This section summarizes how the parameterization scheme and observations are provided to PEST through its input files.

5.2 PEST control file

The primary file that instructs and controls the information flow into and out of PEST is the PEST control file. Many of the data items in the PEST control file are used to “tune” PEST’s operation to the case in hand; such items include parameter change limits, changes in observations and other calibration targets, parameter transformation types, the amount and type of regularization, termination criteria, and solution methodology. Table 5-1 lists the information sections contained in a PEST control file. Most of the mechanics and details associated with each of these sections and the considerations associated with the construction of the PEST control file are beyond the scope of this report. For readers interested in such details, the PEST manual and PEST addendum provide over one-hundred pages of text that discuss the construction of a PEST control file. The PEST control file pertaining to the calibration exercise described here is named *gt6b.pst*. This file includes the sections mark in bold in Table 5-1. File *gt6b.pst* includes approximately 15,000 lines of data. Because of its length, only portions of the file are listed in Appendix H.

Table 5-1. Names of sections that may be present within a PEST control file.
Those cited in the PEST control file used in calibration of the ETPV GAM are shown in bold.

Information section	Issues addressed
Control Data	Array dimensions, input/output options, optimization mode, Marquardt-Levenberg parameters, termination criteria
Automatic User Intervention	Variables that control temporary freezing of insensitive parameters in problematic calibration exercises
Singular Value Decomposition (SVD)	Options of implementing singular value decomposition
LSQR	Options for implementing LSQR iterative solution scheme to inverse problem
SVD Assist	Options for using super parameters in solution of inverse problem
Sensitivity Reuse	Options for reducing number of runs required for filling of the Jacobean matrix
Parameter Groups	Names for parameter groups, instructions for calculating finite-difference derivatives
Parameter Data	Listing of every model parameter to be estimated, upper and lower bounds, preferred values, assignment of each parameter to a parameter group
Observation Groups	Names for observation groups
Observation Data	Listing of every model output that is matched to a field observation, target value, weight assigned to residual, assignment of each observation to an observation group
Derivatives Command Line	Command used to run model where model can supply its own derivatives
Model Command Line	Command to run the model

Information section	Issues addressed
Model Input/Output	Listing of template files used to generate model input files, listing of instruction files to extract information from model output files
Predictive Analysis	Options for implementation of nonlinear predictive uncertainty analysis through constrained maximization/minimization
Prior Information	Linear relationships between parameters comprising extra “observations” to use in calibration process
Regularization	Options for implementing Tikhonov regularization

5.3 Parameters

5.3.1 Initial parameter values

When undertaking regularized inversion the initial values provided to parameters should be “preferred values” according to a modeler's precalibration expert judgment. This principle was followed in calibrating the ETPV GAM. Preferred values for all parameters are provided in Section 4. As discussed in that section, some of the preferred values were based on expert knowledge and local measurement datasets, while others were inherited from the TWDB ETPV GAM. All pilot-point multiplier parameters of all types were assigned a preferred (and initial) value of 1.0.

5.3.2 Parameter naming

Zone-based aquifer hydraulic properties

For zone-base hydraulic property parameters, the zone to which a parameter pertains is embedded in its name. Zone numbers are depicted in Figures 3-12 and 3-13. Parameter names are provided in Table 5-2.

Table 5-2. Zone-based aquifer hydraulic property parameters.

Parameter name	Description
hk1_zone2	horizontal hydraulic conductivity in zone 2 (layer 1)
hk1_zone3	horizontal hydraulic conductivity in zone 3 (layer 1)
hk2_zone5	horizontal hydraulic conductivity in zone 5 (layer 2)
hk2_zone6	horizontal hydraulic conductivity in zone 6 (layer 2)
hk2_zone7	horizontal hydraulic conductivity in zone 7 (layer 2)
Vertanis	vertical anisotropy for all zone-based hydraulic conductivity parameters
s1_zone2	specific storage in zone 2 (layer 1)
s1_zone3	specific storage in zone 3 (layer 1)
s2_zone5	specific storage in zone 5 (layer 2)
s2_zone6	specific storage in zone 6 (layer 2)
s2_zone7	specific storage in zone 7 (layer 2)
sy1_zone2	specific yield in zone 2 (layer 1)
sy1_zone3	specific yield in zone 3 (layer 1)
sy2_zone5	specific yield in zone 5 (layer 2)
sy2_zone6	specific yield in zone 6 (layer 2)
sy2_zone7	specific yield in zone 7 (layer 2)

Pilot-point-based hydraulic property multipliers

Pilot point locations are shown in the previous section. Each pilot point was placed at the center of a model cell; however, this strategy is by no means fundamental to the use of pilot points.

Pilot point parameters are named according to the convention:

pn_rr_cc

where:

- p* denotes the hydraulic property to which the point pertains;
- n* denotes the layer number to which the point is assigned;
- rr* denotes the row number of the cell at whose center it lies; and
- cc* denotes the column number of the cell at whose center it lies.

Property designators are “*hk*” for horizontal conductivity, “*s*” for storage coefficient, “*sy*” for specific yield, and “*v*” for vertical conductance.

Pilot-point-based conductance parameters

The same naming convention as that described above is adopted for pilot points used for representation of *general head boundary* and *streambed* conductance. The parameter codes are “*gc*” for *general head boundary* conductance and “*strc*” for *streambed* conductance.

Drain conductance

The zonation used for *drain* conductance parameterization is shown in Figure 4-14. Drain conductance parameters are named according to the convention *drncncN-Z* where *N* is the layer number in which the zone is situated, and *Z* is the zone number within that layer.

Recharge

Recharge zonation is depicted in Figure 4-9. Recharge factor parameters are named according to the convention *rf_zoneZ* where *Z* is the zone to which a factor pertains.

Pilot-point-based recharge multipliers are named according to the convention *rmrr_cc* where *rr* and *cc* are the row and column numbers of the cell center at which the pertinent pilot point lies.

Pumping parameters

Pumping build-up factor parameters are named according to the convention *welfaczz_pp* where *zz* denotes the zone to which a build-up factor pertains (see Figure 4-22 for zone numbers) and *pp* denotes the stress period to which it pertains (these are 1 to 5).

Pilot-point-based multiplier factors for layers 1 and 2 are named according to the convention *wmN_rr_cc* where *N* is the layer number (1 or 2), and *rr* and *cc* denote the row and column

numbers of the cell in which each pilot point lies. (Pilot points were placed at the centers of cells.)

Other parameters

The single stream roughness parameter is named *str_rough*.

5.3.3 Parameter bounds

Parameter bounds are discussed in the previous section; see also the fifth and sixth columns of the “parameter data” section of the PEST control file.

For many parameters, lower and upper bounds were set to very low and very high values respectively so that pertinent parameters were free to move wherever PEST felt it necessary to move them. In other cases, bounds were set more tightly.

Imposition of tight parameter bounds is a form of regularization. However it is not a preferred form of regularization. In most calibration contexts, other, more sophisticated forms of regularization hopefully will prevent parameters from assuming unrealistic values; the use of parameter bounds constitutes a “last line of defense” against this occurrence. In the course of a parameter estimation process, too many parameters can be prevented from hitting their bounds by setting the Tikhonov target measurement objective function to a higher value, thereby forcing tighter implementation of Tikhonov constraints, and thus limiting the movement of parameters from their preferred values. On the contrary, regularization through imposition of bounds is, of necessity, implemented on a parameter-by-parameter basis as individual parameters (normally relatively insensitive parameters) encounter their bounds.

5.3.4 Tikhonov regularization

General

Some of the principles underlying regularized inversion as a means of calibration of complex models are provided in a previous section and in Appendix D. There it was stated that when undertaking regularized inversion, a modeler is not required to formulate a well-posed inverse problem prior to initiating the calibration process. In fact the opposite is the case. In recognition of the fact that the process of model parameterization is necessarily uncertain, the modeler is able to respect his/her lack of certainty by allowing parameters whose values are unknown the freedom of being adjusted through the calibration process. He/she is not thus consigned to resting the calibration process on a litany of assumptions that will never be tested, but that will nevertheless be reflected in values assigned to model parameters, and to predictions calculated on the basis of those parameters. Instead, by introducing comprehensive regularization to the inverse problem, thus reflecting a modeler's knowledge, the calibration process is free to estimate a set of parameters that can be considered to approach that of minimum error variance. Estimation of the size of this variance can then take place as an adjunct to the calibration process using either linear or nonlinear methods that either implicitly or explicitly allow these parameters to vary within ranges that the modeler determined to be realistic.

Two types of regularization were used to calibrate the ETPV GAM: Tikhonov and subspace. Subspace regularization was implemented through use of the LSQR solution scheme (this providing similar solution characteristics to that provided by truncated singular value decomposition). Tikhonov regularization, as implemented in calibration of the ETPV GAM is described below.

Preferred value regularization

In implementing Tikhonov regularization for the ETPV GAM, each parameter was assigned a preferred value equal to its initial value. This was done through provision of a prior information equation for all parameters implementing this condition. In the PEST control file, these prior information equations are assigned to different observation groups. Note that an observation group that implements regularization is assigned a name that begins with the string “regul”, thus becoming a “regularization group” rather than an observation group.

PEST has license to apply a global weight multiplier to observations and prior information equations belonging to regularization groups. This is the factor β^2 featured in Eq. D.21. However, as well as implementing global weight factor adjustment, PEST is able to adjust weights pertaining to collections of regularization observations and/or regularization of prior information equations, if these collections are assigned to different regularization groups, and if the IREGADJ variable in the “regularization” section of the PEST control file is set to a number greater than 1. This allows PEST to enforce regularization constraints more tightly on parameters that are less informed by the data, thus providing some measure of protection against numerical instability that can sometimes hamper the operation of Tikhonov regularization, especially in its final stages. In calibration of the ETPV GAM, IREGADJ was set to 4, thus implementing “subspace enhanced Tikhonov regularization”. See the PEST documentation for further details. As an inspection of the PEST control file reveals, prior information equations that express preferred values for parameters of different types are indeed assigned to different regularization groups, thereby allowing PEST to weight these equations differentially in implementing Tikhonov constraints.

The ADDREG1 utility supplied with PEST was used to automate the addition of prior information equations implementing regularization constraints to a PEST control file which did not include these constraints. Incorporation of preferred value regularization in the calibration process was thus a simple matter.

Supplementary regularization

As was described in the previous section, the horizontal hydraulic conductivity value assigned to any cell is an outcome of a two-stage parameterization process, the first being the provision of zone-based parameter values, and the second being multiplication by pilot-point-based multiplier arrays. Both zonal and pilot-point multiplier parameterization are layer-specific.

To respect hydraulic conductivity estimates obtained from pumping test interpretation, an extra Tikhonov regularization measure was implemented for horizontal hydraulic conductivity. Implementation details follow.

1. Hydraulic conductivity arrays built on the basis of zone-based parameter assignment followed by spatial pilot-point-based multiplication are divided by arrays that express median hydraulic conductivity within each zone based on analyses of pumping tests undertaken historically within those zones. These median values are the same as initial values (and preferred values) employed for respective zonal parameters. Division is undertaken using the TWOARRAY utility.
2. The logs are taken of these ratios on a cell-by-cell basis; this is done with the LOGARRAY utility.
3. The array of log ratio values is space-interpolated to pilot point locations within each zone; spatial interpolation is undertaken using the ARR2BORE utility.
4. This log is assigned a preferred value of zero.

Log ratios are read by PEST after each model run. They are read from the ARR2BORE-generated files *log_ratio_median_hk_lay1.dat* and *log_ratio_median_hk_lay1.dat* using instructions contained in files *log_ratio_median_hk_lay1.ins* and *log_ratio_median_hk_lay1.ins*. They are assigned to an observation group whose name begins with “regul” and hence are treated as regularization observations and are thus applied as Tikhonov constraints.

Files employed in supplementary horizontal-hydraulic conductivity regularization are listed in Appendix D.

5.4 Observations

5.4.1 Introduction of observations to PEST input dataset

As discussed in the previous section, three types of observations are cited in the PEST input dataset: heads, temporal head differences, and fluxes. However, for reasons described in Section 5, only heads were employed in the ETPV GAM recalibration.

Building a PEST input dataset that incorporates a number of observation suites can be largely automated using programs from the Groundwater Data Utility suite. A brief description of the procedure is provided below. For details, see documentation of the PEST Groundwater Data Utility suite.

Observed heads

Observed heads should be stored in “bore sample file” format; see documentation of the Groundwater Data Utilities for details of this format. The MOD2OBS utility, run as a MODFLOW post-processor, can then undertake spatial and temporal interpolation of model-calculated heads at discrete MODFLOW output times (the ends of time steps and/or stress periods) to the sites and times at which measurements were taken. MOD2OBS reads a binary MODFLOW-generated heads file and records the outcomes of its interpolation process to a bore

sample file, thus becoming the model-generated equivalent of the observation bore sample file. (Note that in the present case a double precision version of MOD2OBS was employed because of the fact that a double precision version of MODFLOW was employed.)

The PESTPREP1 utility can be used to write an instruction file for the MOD2OBS-generated bore sample file, and to build a PEST control file featuring observed heads. Alternatively, the PESTPREP2 utility can be used to add observed heads to an existing PEST control file.

Head differences

Temporal head differences of both observed and model-generated heads can be calculated using the SMPDIFF utility. SMPDIFF should be run as a model postprocessor. SMPDIFF reads a bore sample file; its computed head differences are also recorded in bore sample file format. The PESTPREP2 utility can be used to build an instruction file that reads model-calculated head differences, and to add observed head differences to an existing PEST control file.

Stream flows

In normal MODFLOW operation, the BUD2SMP utility can be employed to read flows accumulated over user-defined zones from a binary cell-by-cell flow term file, and record these flows in bore sample file format. In the present study, however, a utility named STRFLBUD utility is employed. This is because BUD2SMP assumes compact budget storage; however, MODFLOW *streamflow* budget data are stored in expanded form, regardless of user-supplied instructions to do otherwise. Like BUD2SMP, STRFLBUD records MODFLOW-calculated flows in bore sample file format. The SMP2SMP utility can be used to undertake temporal interpolation from model-generated data stored in this format to observed data recorded in the same format, thereby producing the model-generated equivalent of the observed bore sample file.

Once modeled and observed flows are in bore sample file format, instruction files for the former can be built using the PESTPREP2 utility; the same utility also adds observed flows to an existing PEST control file.

5.4.2 Observation weights

Where multiple observation types are employed in a calibration process, or where observations of the same type from different layers and/or locations are employed, these should be placed into different observation groups when building the PEST control file. The contribution made to the overall objective function by these different groups can thereby be monitored; if necessary, weights adjustment can be implemented to ensure that the contribution made by each group to the overall objective function at the beginning of the inversion process is about equal to that made by other groups. This ensures that the information contained in one group neither dominates the objective function, nor is dominated by information contained in other observation groups.

Measured heads pertaining to layers 1 and 2 are assigned to different observation groups in the PEST control file. Within each group, all head measurements are weighted equally. The

PWTADJ1 utility was used to equalize contributions to the initial objective function made by these two observation groups; this initial contribution was arbitrarily set to 1,000 in each case.

As already stated, head difference and streamflow observations are assigned weights of zero in the PEST control file. This was done using the WTFACOR utility.

5.5 PEST input files

5.5.1 Template files

Template files employed by PEST in calibration of the ETPV GAM are listed in Table 5-3, together with the model input file that is associated with each. Note that the same information presented in the first two columns of this table is provided in the “model input/output” section of the PEST control file.

Table 5-3. Details of template files cited in PEST control file used in calibration of the ETPV GAM.

Template file	Model input file	Model program which reads file	Parameter types cited
<i>hk_lay[1-2].tpl</i>	<i>hk_lay[1-2].irc</i>	INT2REAL	zonal hydraulic conductivities
<i>s_lay[1-2].tpl</i>	<i>s_lay[1-2].irc</i>	INT2REAL	zonal storativities
<i>sy_lay[1-2].tpl</i>	<i>sy_lay[1-2].irc</i>	INT2REAL	zonal specific yields
<i>vcontcalc.tpl</i>	<i>vcontcalc.in</i>	VCONTCALC	global vertical anisotropy through which vertical conductance is computed from zonal horizontal conductivity
<i>hk_lay[1-2]_mul.tpl</i>	<i>hk_lay[1-2]_mul.dat</i>	FAC2REAL	pilot-point-based multipliers for horizontal hydraulic conductivity
<i>s_lay[1-2]_mul.tpl</i>	<i>s_lay[1-2]_mul.dat</i>	FAC2REAL	pilot-point-based multipliers for storativity
<i>sy_lay1_mul.tpl</i>	<i>sy_lay1_mul.dat</i>	FAC2REAL	pilot-point-based multipliers for specific yield
<i>vcont_lay1_mul.tpl</i>	<i>vcont_lay1_mul.dat</i>	FAC2REAL	pilot-point-based multipliers for vertical conductance
<i>rf.tpl</i>	<i>rf.irc</i>	INT2REAL	zone-based factors by which rainfall is multiplied to obtain recharge
<i>rechmul.tpl</i>	<i>rechmul.dat</i>	FAC2REAL	pilot-point-based multipliers for recharge
<i>ghb_lay[1-2].tpl</i>	<i>Ghb_lay[1-2].dat</i>	FAC2REAL	pilot-point-based <i>general head boundary</i> conductance
<i>strcond.tpl</i>	<i>strcond.dat</i>	FAC2REAL	pilot-point-based <i>stream</i> conductance
<i>realstr2cnd.tpl</i>	<i>realstr2cnd.in</i>	REALSTR2CND	global streambed roughness
<i>draincond_lay[1-2].tpl</i>	<i>draincond_lay[1-2].irc</i>	INT2REAL	zone-based <i>drain</i> conductances
<i>buildup_factors.tpl</i>	<i>buildup_factors.dat</i>	WELBUILD2	pumping rate build-up factors
<i>welmul_lay[1-2].tpl</i>	<i>welmul_lay[1-2].dat</i>	FAC2REAL	pilot-point-based pumping multipliers

5.5.2 Instruction files

Instruction files employed by PEST are listed in Table 5-4, together with the model output file that is read by each. Note that the same information is presented in the first two columns of this table that is in the “model input/output” section of the PEST control file. Note also that the last row of Table 5-4 pertains to regularization rather than field observations. However, because

computations performed in implementing these regularization constraints are carried out as part of the overall batch-file model that is run by PEST, the outcomes of these calculations must be read using instruction files just like any other model outputs used by PEST.

Table 5-4. Details of instruction files cited in PEST control file used in calibration of ETPV GAM.

Instruction file	Model output file	Model program which writes file	Observation types cited
<i>model_cnt_gt_6b.ins</i>	<i>model_cnt_gt_5.smp</i>	MOD2OBS	heads in layers 1 and 2
<i>model_cnt_gt_6b_diff.ins</i>	<i>model_cnt_gt_5_diff.smp</i>	SMPDIFF	head differences in layers 1 and 2
<i>mo_flows.ins</i>	<i>mo_flows.smp</i>	STRFLBUD	stream flows
<i>log_ratio_median_hk_lay [1-2].ins</i>	<i>log_ratio_median_hk_lay [1-2].dat</i>	ARR2BORE	Pilot-point-based ratios of hydraulic conductivities to median hydraulic conductivity in each zone

5.5.3 The model batch file

In calibrating the ETPV GAM, the batch file that is run repeatedly by PEST is *model_gt6b.bat*. This file is listed in Appendix I. Some aspects of this file are briefly discussed below.

Comments

In a batch file, comments are preceded by the REM (for “remark”) string.

File deletion

As explained in the PEST documentation, all files that are written by one model component and read by another model component should be deleted prior to running the first component. If the first component then fails to run, subsequent components will not read old data files, mistaking them for new ones. In the first part of *model_gt6b.bat*, all such intermediate files are deleted. MODFLOW text output files *output_ss.dat* and *output_tr.dat* are also deleted. The user can thereby be sure that if the model fails to run, these files (if present) will pertain to the latest attempt on PEST’s part to run the model.

Keyboard input files

Where a program requires keyboard input, that input can be placed within a text file. When the program is run, its attention can be directed to that text file rather than to the keyboard by the “<” character. The program can thus be run without user intervention or supervision.

Elimination of screen output

When a batch file is run repeatedly by PEST, it is sometimes convenient to eliminate screen output from that program within that batch file so that screen output from these programs does not interfere with that of PEST. This strategy is implemented through taking the following steps.

1. The “@echo off” command should be placed at the head of the batch file.
2. Output from all system and program commands should be directed to the null file through placement of the string “> nul” at the end of each command.

This strategy was not implemented in the model batch file used for calibration of the ETPV GAM as calibration was undertaken by Parallel PEST. Because PEST runs in a different window (or even on a different machine) from that used by any model incidence, model screen output does not interfere with that of PEST.

5.5.4 Additional settings in the PEST control file

The PEST control file *gt6b.pst* used in calibrating the ETPV GAM is not reproduced in this report because it is too long. A few comments on that file are provided in this section. Variables occurring within this file are fully described in the PEST documentation; therefore, the names of these variables are used liberally in the following discussion.

Mode of operation

The PESTMODE variable is set to “regularization”; hence, PEST implements Tikhonov regularization. As already discussed, regularization is implemented through a series of prior information equations, as well as through a series of observations; all are assigned to an observation group whose name begins with “regul”.

A “regularization” section is present within the PEST control file. The target measurement objective function (PHIMLIM) is set to 300. This value was selected on the basis of previous calibration exercises, and is somewhat above the minimum objective function value that PEST is capable of achieving. On previous PEST runs, PHIMLIM was set to 1.0. However, the FRACPHIM variable maintained visibility of Tikhonov regularization by providing a temporary setting for PHIMLIM equal to 0.1 times the current value of the measurement objective function.

In the final line of the PEST control file, the IREGADJ regularization control variable is set to 4. This gives PEST license to provide differential adjustment of weights applied to individual regularization constraints with stronger weights applied to constraints on parameters for which the information content of the data is weakest.

Solution methodology

A “singular value decomposition” section is provided in the *gt6b.pst*. In that section, the EIGHTHRESH variable is set at 5E-7, thus providing a singular value truncation threshold that will, in most contexts, provide a suitable safeguard against numerical instability. However, because the SVDMODE variable in this section is set to zero, singular value decomposition is not actually used for solution of the inverse problem. While it could have been used, the LSQR method was used instead as this provided greater numerical solution speed, given the number of parameters that required estimation, and the number of observations on which estimation of these parameters was based.

In the “LSQR” section of the PEST control file, the LSQRMODE control variable is set to 1. The LSQR_ATOL, LSQR_BTOL, LSQR_CONLIM and LSQR_ITNLIM variables are set to 1.0e-10, 1.0e-10, 5.0e3 and 50000 respectively. The critical variable is LSQR_CONLIM. This sets a limit on the condition number of the inverse problem. As such, it performs a similar role to the EIGHTHRESH variable. A setting of 5E3 ensures that the condition number will not rise above this level. The equivalent EIGHTHRESH value is the inverse of this number squared, or 4E-8. This is perhaps a little low (and LSQR_ITNLIM was therefore set perhaps a little high). However in the present case, Tikhonov regularization contributes to the numerical stability of the inverse problem.

Parameter transformation

In the “parameter data” section of the PEST control file, all parameters are designated as log transformed. Thus PEST, in a manner that is invisible to the user, estimates the log of each parameter rather than its actual value. This can bring certain advantages to a parameter estimation process, including the following.

1. It tends to equalize sensitivities among different parameter types. This, in turn, often leads to greater numerical stability of the inversion process.
2. The dependence of model outputs on the logs of parameters is often more linear than their dependence on actual parameters.
3. The innate variability of parameters (as would be recorded in a $C(\mathbf{k})$ matrix for those parameters – see the theory presented in earlier sections of this report) is often approximately proportional to the numerical size of the numbers used to represent them. Because differentiation with respect to the log of a parameter is equivalent to differentiation with respect to a relative parameter value, log transformation of parameters provides approximate normalization of parameters by their innate variability. The promise of singular value decomposition (and indeed of LSQR) for a minimum norm solution of the inverse problem thus becomes a promise for the maximum parameter likelihood solution of the inverse problem.

Computation of derivatives

Parameters of different types are assigned to different parameter groups. Settings for finite-difference derivatives computation are applied to these groups, rather than to individual parameters, to save duplication. It is apparent from the “parameter groups” section of the PEST control file that the following strategy is employed for derivatives computation.

1. A relative increment of 1.5 percent of each parameter’s current value is employed for finite-differencing purposes. However for some parameters a groups-specific lower increment bound is designated to prevent this increment from becoming too small. (See the INCTYP, DERINC and DERINCLB variables).
2. A switch is made in the use of central derivatives with parabolic interpolation when progress in lowering the objective function begins to slow. At this stage, the derivative increment is multiplied by two. (See the FORCEN, DERICMUL and DERMTHD variables.)

Marquardt lambda

An initial Marquardt lambda value (RLAMBDA1) of 10.0 is selected. Assignment of a value of -3 to the RLAMFAC variable provides PEST with the ability to change the Marquardt lambda rapidly as it tests new parameter upgrades. Thus the inversion process can rapidly accommodate the onset of a high degree of nonlinearity and/or numerical instability inspired by incipient parameter nonuniqueness. Because the JACUPDATE variable is set to 999, the Jacobean matrix is improved on each occasion when more information on local parameter variability is made available through the Marquardt lambda testing procedure.

6 Implementation and outcomes

6.1 Implementation

6.1.1 *Parallel PEST*

In the course of optimizing parameters for a model calibration, PEST runs the model many times as part of the process of calculating the Jacobean matrix, i.e., the matrix of derivatives of observations with respect to parameters (unless derivatives are supplied to PEST directly by the model in accordance with PEST's external derivatives functionality). In calculating the Jacobean matrix, PEST needs to run the model at least as many times as there are adjustable parameters (and up to twice this number if derivatives for some of the adjustable parameters are calculated using central differences). In most cases, the bulk of PEST's run time is consumed in running the groundwater model. It follows that any time savings that are made in carrying out these model runs will result in dramatic enhancements to overall PEST performance.

Because the Jacobean matrix includes derivatives that can be calculated independently of each other, the opportunity exists to calculate the derivations simultaneously on different computers. To take full advantage of this opportunity, Parallel PEST was used on a computer network to distribute model runs to different machines or processors as they became available and to process the outcomes of these runs as they were finished.

As well as allowing a user to distribute model runs across a network, Parallel PEST can be used to manage simultaneous model runs on a single machine. This type of approach significantly increases PEST's efficiency when carrying out parameter optimization or predictive analysis on a multi-processor computer by keeping all processors simultaneously busy carrying out model runs.

6.1.2 *Model run parallelization*

The PEST control file *gt6b.pst* includes 2485 model parameters and thus requires 2485 model runs to calculate the necessary derivations to create the Jacobean matrix for the model optimization. Each model simulation requires about 3 minutes of computing time on the Dell 330N desktops used for the project. Thus, the total time to perform a single PEST iteration in which derivatives are calculated for all parameters, and parameters are then updated, is approximately 124 hours (unless central differences are used for derivatives calculation in which case the time is increased to 248 hours).

To reduce the model run time per iteration, an HPC cluster of six computers with duo core processors was used. Appendix J provides a description of the cluster, which was constructed specifically for this project. The cluster offers a total of 12 individual processors available to Parallel PEST. Use of Parallel PEST across this HPC cluster reduced the total run time by a factor of about 12 so that the time for one iteration was reduced from 207 hours to about 17 hours.

Use of Parallel PEST is identical to that of PEST except for the fact that Parallel PEST requires that one extra file be prepared prior to undertaking an optimization run, i.e., a “run management file”. This file informs Parallel PEST of the machines to which it has access, of the names of the model input and output files residing on those machines, and of the name of a subdirectory it can use on each of these machines to communicate with a “slave” that carries out model runs on request.

The PEST “master” program sends messages to “slave” programs residing on the same or other computers. Upon receipt of the appropriate message, the slave program runs the model. Upon completion of each model run, it informs PEST that the run has indeed been completed. PEST then reads model output files in the usual manner.

6.1.3 SVD-Assisted parameter estimation

In many calibration circumstances, especially where many parameters require estimation, the SVD-assist methodology provided with PEST can be implemented to solve the inverse problem of model calibration. As previously discussed, this reduces the computational burden of the parameter estimation process considerably. Only as many model runs must be undertaken per iteration as there are super parameters that require estimation. The number of super parameters should be set equal to, or larger than, the dimensions of the calibration solution space.

SVD-assisted parameter estimation was not required for the ETPV GAM because the runtime of this model was sufficiently short to allow parameter estimation to proceed without the use of super parameters.

6.2 Obtaining calibration outcomes

6.2.1 Objective function

The PEST algorithms are designed to obtain the best match between a wide range of calibration targets and their corresponding values as calculated by the model. The user is provided with unlimited flexibility in choosing these targets. The difference between a calibration target and its respective value as calculated by the model is referred to as a residual. In any PEST run, thousands of residuals may be generated. As part of designing a PEST-based calibration process, the modeler needs to define a global objective function, called phi, that provides a single value that PEST then minimizes by adjusting model parameters.

The objective function, phi, is calculated as the squared sum of weighted residuals. One of the modeler's responsibilities is introducing his knowledge and understanding of the site to the PEST-based calibration process through the selection of different weights for all of the calibration targets. The PEST runs for the ETPV GAM included over 14,000 calibration targets. To track and better manage the outcomes of the optimization process, PEST allows the residuals to be divided into groups so their contributions to phi can be tracked and monitored at each iteration. Table 6-1 lists the 27 observation groups employed in the current study.

In Table 6-1, regularization groups (which contribute to the “regularization objective function”) are distinguished from observation groups (which contribute to the “measurement objective

function” through possessing names which begin with “regul”. As was described above, PEST adjusts weighting applied to members regularization groups in accordance with the operation of its Tikhonov regularization scheme.

Table 6-1. Groups that contribute to the value of the objective function phi.

Observation group name	Basis for the sum of the weighted residuals
flows	Stream flows
heads_lay1	Water levels associated with layer 1
heads_lay2	Water levels associated with layer 2
diffs_lay1	Temporal differences in water levels measured in each well associated with layer 1
diffs_lay2	Temporal differences in water levels measured in each well associated with layer 2
regul_mhk1	Differences between modeled and measured median horizontal hydraulic conductivities in layer 1
regul_mhk2	Differences between modeled and measured median horizontal hydraulic conductivities in layer 2
regul_hkzone	Preferred horizontal hydraulic conductivity values for each of the five lithostratigraphic zones
regul_szone	Preferred specific storage values for each of the five lithostratigraphic zones
regul_syzone	Preferred specific yield values for each of the five lithostratigraphic zones
regul_rfzone	Preferred factors used to calculate recharge as a fraction of total precipitation for each of the 11 recharge zones
regul_vertan	Global estimate of the vertical anisotropy (K_v/K_h) between vertical hydraulic conductivity, K_v , and horizontal hydraulic conductivity, K_h
regul_welfac	Preferred value for the multiplier used to adjust the 1980 pumping rate for pumping in each of the 16 pumping zones over 10-year periods beginning in 1930, 1940, 1950, 1960, and 1970
regul_drain	Preferred value for the <i>drain</i> conductance
regul_ghb	Preferred value for the <i>general head boundary</i> conductance
regul_hkmul1	Preferred values for pilot point horizontal hydraulic conductivity multipliers in layer 1
regul_hkmul2	Preferred values for pilot point horizontal hydraulic conductivity multipliers in layer 2
regul_rechmu	Preferred values for recharge pilot point multipliers
regul_smul1	Preferred values for specific storage pilot point multipliers in layer 1
regul_smul2	Preferred values for specific storage pilot point multipliers in layer 2
regul_symul1	Preferred values for specific yield pilot point multipliers in layer 1
regul_symul2	Preferred values for specific yield pilot point multipliers in layer 2
regul_vcontm	Preferred values for pilot point vertical conductance multipliers
regul_wm1	Preferred values for the pilot points used to interpolate the distribution of pumping from 1930 to 1980 in layer 1
regul_wm2	Preferred values for the pilot points used to interpolate the distribution of pumping from 1930 to 1980 in layer 2
regul_strcon	Preferred values for <i>stream</i> conductance
regul_str_ro	Preferred value for the roughness factor used in the <i>stream</i> package

PEST uses a number of different criteria to determine when to halt its iterative optimization process. Most of these criteria are tied to the behavior of the objective function. One set of criteria focuses on whether or not phi is being reduced, as well as its rate of reduction. Other criteria focus on whether the objective function has fallen to a user-specified value that is deemed to indicate that an appropriate level of fit has been obtained between model outputs and corresponding field measurements. For this study, PEST simulations were primarily terminated when the measurement objective function reached a value below 500; when carried out, it was found that further iterations were rarely able to reduce the objective function below this value. This objective function value was usually achieved after six iterations.

6.2.2 Output files

Upon completion of the parameter estimation process, the PEST run record file can be read to obtain a complete history of that process. The run record possesses the same filename base as the PEST control file, but has an extension of “.rec”. The run record file provide a detailed account of PEST inputs, current parameter values during any optimization iteration, optimization control variables, and the value of the objective function. For each iteration, the run record provides the contribution of each observation group to the overall objective function. At any stage of the optimization process, best parameter values achieved up to that point are recorded in the parameter value file. This also has the same filename base as the PEST control file, but possesses an extension of “.par”. Other PEST-generated files of interest are listed in Table 6-2. Note that some of these are not applicable to the present case as their production (or otherwise) is dependent on the implementation details of the current parameter estimation process.

Table 6-2. Some files generated by PEST; casename is the filename base of the PEST control file.

File	Purpose
<u>casename.rec</u>	Run record file, contains a complete history of the parameter estimation process
<u>casename.par</u>	Parameter value file, contains optimized parameter values; if undertaking SVD-assisted parameter estimation, these are the optimized values of super parameters
<u>casename.bpa</u>	Optimized base parameter values if undertaking SVD-assisted parameter estimation
<u>casename.jco</u>	Jacobian matrix file computed on basis of best (or almost best) parameter values
<u>casename.lsq</u>	Information generated by LSQR solver
<u>casename.svd</u>	Information generated through singular value decomposition solution process
<u>casename.cnd</u>	Contains condition number of inverse problem; not generated if LSQR or SVD solvers are employed
<u>casename.mtt</u>	Contains parameter covariance matrix and information derived from it; not generated if any form of regularized inversion is undertaken
<u>casename.res</u>	Contains model-to-measurement residuals calculated on basis of optimized parameters
<u>casename.rei</u>	Contains model-to-measurement residuals calculated at intermediate stages of the parameter estimation process
<u>casename.rmr</u>	Record of communication between the Parallel PEST master and its slaves

Unless it is undertaking SVD-assisted inversion, PEST undertakes one final model run once it has completed the parameter estimation process, this being based on optimized parameters. However, if PEST is stopped prematurely, this final model run will not be carried out. In that case, the following procedure can be followed to undertake a model run based on optimized parameters, the latter being available at any time during the parameter estimation process in the casename.par file.

1. Create a new PEST control file based on which initial parameter values are optimized parameter values, using the PARREP utility.
2. Set the NOPTMAX variable to zero in this new file (thus instructing PEST to undertake just one model run).
3. Run PEST. All model input files will then cite optimized parameters. Numbers appearing in all model output files will be computed on the basis of optimized parameters.

6.3 Model parameters achieved through model recalibration

6.3.1 Aquifer properties

The aquifer properties that were adjusted during the model calibration process were hydraulic conductivity, vertical conductance, specific storage, and specific yield. Figures 6-2 through 6-6 show the distribution of aquifer properties generated from model recalibration. Table 6-3 shows the arithmetic average, median, minimum, and maximum for three of the aquifer properties. Because the calculation of vertical conductance relies on parameters from more than one aquifer layer and therefore more than one zone, vertical conductance is not included in Table 6-3.

Of the four aquifer properties, hydraulic conductivity has by far the greatest number of field measurement values. One of the purposes of including regularization in the calibration process is to constrain model input parameters by information extracted from field measurements. For this recalibration, one of the regularization objectives was to have the median and averages of the model hydraulic conductivity values lie between the 40th and 60th percentile of the hydraulic conductivity field data from each zone. A comparison of the values in Table 6-3 with those in Table 4-2 shows that calibration accomplished this goal.

The results in Figure 6-4 for vertical conductances support a general trend of lower values in the southern portion than in the northern portion of the model domain. This general trend is partly a result of the fact that the thickness of the Trinity Aquifer in layer 2 increases to the south.

Across the model, the specific storage values range from $2e-6 \text{ ft}^{-1}$ to $1e-4 \text{ ft}^{-1}$. As shown in Figures 6-5 and 6-6, the highest values exist in unconsolidated sediments in the Pecos Valley aquifer in layer 1 and the lowest values exist in the more consolidated deposits of the Trinity aquifer in layer 2. Within Zone 2, it is of interest to note that the location of higher hydraulic conductivity and specific storage values in eastern Reeves County coincides with the location of the Pecos Valley aquifer as identified in Figure 2-2. With additional information and focus, it appears that the differences in the geologic properties of the western portion of the model could be better defined.

Across the model, the specific yield values range from 0.009 to 0.15. As shown in Figures 6-7 and 6-8, the highest values exist in unconsolidated sediments in the Pecos Valley aquifer in layer 1 and the lowest values exist in the more consolidated deposits of the Edwards aquifer in layer 1.

Table 6-3. Summary statistics for aquifer parameters in the recalibrated ETPV GAM.

Aquifer parameter		Zone				
		2	3	5	6	7
		Edwards	Pecos Valley	Trinity South	Trinity North	Trinity North
Horizontal Hydraulic Conductivity (ft/day)	Arithmetic average	8.0	7.6	2.9	3.3	3.9
	Median	8.0	7.1	2.1	3.7	3.7
	Minimum	8.0	4.0	1.0	1.0	1.0
	Maximum	8.9	13.0	7.0	6.9	7.0
Specific Storage (1/ft)	Arithmetic average	1.1E-05	4.0E-05	9.0E-06	1.0E-05	9.1E-06
	Median	1.1E-05	4.1E-05	9.2E-06	1.0E-05	9.4E-06
	Minimum	6.7E-06	1.6E-05	2.9E-06	1.0E-05	5.8E-06
	Maximum	4.7E-05	1.0E-04	9.7E-06	1.0E-05	1.0E-05
Specific Yield (-)	Arithmetic average	0.009	0.109	0.077	0.081	0.081
	Median	0.009	0.100	0.080	0.081	0.080
	Minimum	0.008	0.100	0.051	0.075	0.051
	Maximum	0.020	0.150	0.100	0.086	0.100

6.3.2 Recharge

The recharge distribution for the recalibrated ETPV GAM remains similar to that of the original ETPV GAM. As shown in Table 6-4, the recharge factors for each zone vary by less than 0.005 for all of the zones except for zone 5, which is the Hill County Trinity Group. The majority recharge adjustment was the outcome of pilot-point multiplier adjustment. As will be discussed later, the recalibrated model has about 88% of the total recharge in the original model. Figure 6-9 shows the recharge distribution for 1980. This result can be compared to the 1980 recharge distribution in Figure 3-4 for the original GAM.

Table 6-4. Preferred recharge factors and bounds for zones illustrated on Figure 6-9.

Recharge Zone		Calibrated recharge factor	Preferred recharge factor
Number	Name		
1	Cenozoic Pecos – Alluvium North	0.010	0.01
2	Buda Limestone or Del Rio Formation	0.010	0.01
3	Edwards Group	0.022	0.02
4	Ogallala Sediments	0.030	0.03
5	Hill County Trinity Group	0.039	0.047
6	Edwards – Devil’s River Formation	0.049	0.05
7	Edwards-Trans-Pecos Basin and Range	0.061	0.06
8	Edwards – Stockton Plateau	0.080	0.08
9	Edwards – Maverick Basin	0.102	0.102
10	Cenozoic Pecos Alluvium South	0.050	0.05
11	No Recharge	0.0	0.0

6.3.3 Hydraulic boundary conditions

The hydraulic boundary conditions in the model provide the link between groundwater and surface water and other aquifers. The hydraulic boundaries in the ETPV GAM consist of constant head cells, *general head boundary* cells, *drain* cells, and *stream* cells. The constant

hydraulic head cells simulate groundwater flow to and from the reservoirs. The *general head boundary* cells simulate groundwater flow to and from other aquifers. The *drain* cells simulate groundwater flow to springs. The *stream* cells simulate groundwater flow to and from the rivers.

All of the hydraulic boundaries are characterized by a hydraulic head value, which represents a source of water outside of the aquifer. Except for the constant head boundaries, all of the hydraulic boundaries are also characterized by a conductance, which represents how well connected the water source is to the groundwater. For the model recalibration, the values for the hydraulic head values were not adjusted. Instead all of the hydraulic heads values associated with the hydraulic boundary conditions were set to the same values as they were in the original ETPV GAM. Thus, for all of the hydraulic boundary conditions, the only variable that was adjusted was conductance.

Conductance associated with *general head boundary* cells in the recalibrated model is shown on Figures 6-10 and 6-11. In general, the conductance value at each grid cell remains within a factor of 10 of the value in the original ETPV GAM. Prior to recalibrating the model, 442 *general head boundary* cells were removed from the model. These cells were removed because they were in the same cells as either a *drain* cell or a *stream* cell. These 442 *general head boundary* cells were removed for two reasons. One was to improve the convergence properties of the model's iterative solution, and the other was to streamline the model's water budget.

The locations of *drain* cells are shown in Figure 3-9. The conductance values for the *drains* in the recalibrated model ranged from 317 ft²/day to 1233 ft²/day and averaged 945 ft²/day. The location and values of the conductances for the recalibrated *stream* cells are shown in Figures 6-12 and 6-13.

6.4 Model solutions

6.4.1 1930 steady-state simulation

The simulated water levels for the 1930 steady-state simulation are shown in Figures 6-14 and 6-15. These water levels represent predevelopment conditions. For this average condition, the recharge is approximately 955,000 AFY. As shown Figure 6-15, recharge comprises over 86% of the total inflow to the aquifer; the remaining sources of inflow consist of rivers (95%), reservoirs (4%), and adjacent aquifers (1%). The primary mechanism for groundwater outflow from ETPV aquifers is rivers (52%) with the remaining discharge mechanisms being, in order of magnitude, adjacent aquifers (24%), springs and seeps (17%), and reservoirs (7%). Table 6-5 provides a water balance for the two layers in the model.

Table 6-5. Water balance for steady-state recalibrated ETPV GAM for predevelopment conditions in 1930.

Recalibrated ETPV model steady state	Flow (AFY) in layer 1			Flow (AFY) in layer 2		
	In	Out	Net	In	Out	Net
Reservoir	16,131	50,492	-34,361	24,020	23,960	60
Inter-layer exchange	20,158	112,821	-92,663	112,821	20,158	92,663
Wells	0	0	0	0	0	0
Spring/seeps	0	120,091	-120,091	0	65,741	-65,741
Recharge	622,845	0	622,845	333,441	0	333,441
Adjacent aquifers	575	89,131	-88,556	7,125	174,065	-166,940
Rivers	67,041	354,216	-287,175	27,026	220,509	-193,483
% Difference			0.00			0.00

6.4.2 1930-2000 transient simulation

The transient calibration period for the original and the recalibrated GAMs is from 1980 to 2000. Among the differences between the two models is the manner in which their initial water levels for 1980 were developed. The original ETPV GAM developed the initial water levels by running a steady-state simulation using 1980 pumping rates. The recalibrated ETPV GAM developed the initial water levels by running a transient model from 1930 to 1980. Of most value to the TWDB GAM program are the simulated water levels and water budgets from 1980 to 2000.

Water levels for the recalibrated GAM are shown for 1980 (Figures 6-17 and 6-18) and for 2000 (Figures 6-19 and 6-20). For comparison, Figures 6-21 and 6-22 show the year 2000 water levels simulated by the original ETPV GAM. In the southern region of the model, the original and the recalibration GAM provide similar water levels for the year 2000. However, in the northern region of the model (Reagan, Upton, and Crockett counties) and the western region of the model (Pecos and Reeves counties), the two models have water levels that differ by up to several hundred feet.

Tables 6-6, 6-7, and 6-8 and Figure 6-23 provide the average annual water balance for the original and recalibration GAM. Over the 20-year period, the recalibrated GAM averages about 1.1 million AFY of recharge whereas the original GAM averages about 1.2 million AFY of recharge. In terms of the average total flow through the aquifer, the recalibrated GAM averages about 1.8 million AFY compared to 1.7 million for the original. As shown in Figure 6-23, the sources of groundwater for the recalibrated GAM include recharge (62%), storage (26%), rivers (8%), reservoirs (3%), and adjacent aquifers (1%). The mechanisms for groundwater discharge in the recalibrated GAM include rivers (27%), storage (25%), wells (19%), adjacent aquifers (14%), springs/seeps (11%), and reservoirs (4%).

Table 6-6. Water balance for the original ETPV GAM for transient conditions from 1980 to 2000.

Original ETPV annual average (1980-2000)	Flow (AFY) in layer 1			Flow (AFY) in layer 2		
	In	Out	Net	In	Out	Net
Storage	154,622	77,177	77,445	59,218	41,903	17,315
Reservoir	16,570	36,944	-20,374	10,470	43,484	-33,014
Inter-layer exchange	9,393	151,226	-141,833	151,226	9,393	141,833
Wells	0	204,882	-204,882	0	123,399	-123,399
Spring/seeps	0	126,698	-126,698	0	79,641	-79,641
Recharge	741,097	0	741,097	452,253	0	452,253
Adjacent aquifers	577	41,530	-40,953	29,644	148,226	-118,582
Rivers	80,797	366,570	-285,773	33,195	289,992	-256,797
% Difference			-0.20			0.00

Table 6-7. Water balance for the steady-state recalibrated ETPV GAM for transient conditions from 1980 to 2000.

Recalibrated ETPV model average (1980-2000)	Flow (AFY) in layer 1			Flow (AFY) in layer 2		
	In	Out	Net	In	Out	Net
Storage	305,013	346,765	-41,752	128,828	77,253	51,575
Reservoir	18,722	46,942	-28,220	24,725	22,516	2,209
Inter-layer exchange	21,013	119,025	-98,012	119,025	21,013	98,012
Wells	0	202,274	-202,274	0	123,591	-123,591
Spring/seeps	0	117,491	-117,491	0	66,855	-66,855
Recharge	688,834	0	688,834	362,922	0	362,922
Adjacent aquifers	3,170	78,536	-75,366	8,588	158,427	-149,839
Rivers	108,352	234,075	-125,723	30,548	204,979	-174,431
% Difference			0.00			0.00

Table 6-8. Water balance for layers 1 and 2 for the original and the recalibrated GAM for transient conditions from 1980 to 2000.

ETPV model average (1980-2000)	Flow (AFY) in original GAM			Flow (AFY) in recalibrated GAM		
	In	Out	Net	In	Out	Net
Storage	213,840	119,080	94,760	433,841	424,018	9,823
Reservoir	27,040	80,428	-53,388	43,447	69,458	-26,011
Inter-layer exchange	160,619	160,619	0	140,038	140,038	0
Wells	0	328,281	-328,281	0	325,865	-325,865
Spring/seeps	0	206,339	-206,339	0	184,346	-184,346
Recharge	1,193,350	0	1,193,350	1,051,756	0	1,051,756
Adjacent aquifers	30,221	189,756	-159,535	11,758	236,963	-225,205
Rivers	113,992	656,562	-542,570	138,900	439,054	-300,154
% Difference			-0.12			0.00

6.5 Calibration statistics

PEST performs model calibration by adjusting model parameters to reduce an objective function. As previously discuss the components of the objective function are determined by the modeler. For this report, the primary calibration statistics are those associated with matching historical water level and flow information. A primary calibration goal of the GAM program is to have the root-mean-square error (RMSE), which is defined later, of the simulated water levels to be equal to or less than 10% of the range in the simulated aquifer. A secondary calibration goal in most GAM reports is faithful reproduction of the exchange between groundwater and streams.

6.5.1 Water levels

The difference between an observed and a simulated water level is called a residual and is defined by equation 6-1.

$$r = h_o - h_s \quad (\text{Eq. 6-1})$$

where: r = residual,
 h_m = observed water level, and
 h_s = simulated water level.

Several of the conventional metrics used to evaluate the calibration status of a model to water levels are based on a statistical analysis of the residuals. The RMSE is traditionally the basic measure of calibration for water levels. The RMSE is defined as the square root of the average square of the residuals and is expressed mathematically by Equation 6-2. Although the RMSE is useful for describing model error on an average basis, it does not provide insight into spatial trends in the distribution of the residuals. Information about spatial trends is provided by the mean error and the mean absolute error (MAE). The mean error, which is described in Equation 6-3, is the average of the residuals. The absolute mean error, which is described in Equation 6-4, is the average of the absolute value of the mean error.

$$\text{Root Mean Squared Error} = \sqrt{\frac{1}{n} \sum_{t=1}^n (h_o - h_s)_t^2} \quad (\text{Eq. 6-2})$$

$$\text{Mean Error} = \frac{1}{n} \sum_{t=1}^n (h_o - h_s)_t \quad (\text{Eq. 6-3})$$

$$\text{Absolute Mean Error} = \frac{1}{n} \sum_{t=1}^n |h_o - h_s|_t \quad (\text{Eq. 6-4})$$

where: n = number of observations.

The total number of observed water levels used in the calibration is 4,773. These values are from 574 wells in the TWDB water well database. Out of these measurements, 2,549 are from layer 1 and 2,224 are from layer 2. Figures 6-24 and 6-25 compare the simulated and the observed water levels for the original and the recalibrated ETPV GAM for model layer 1 and 2, respectively. Table 6-9 provides the calibration statistics for layer 1 for the original and recalibrated model. Table 6-10 provides the calibration statistics for layer 2 for the original and recalibrated model. Table 6-11 provides the calibration statistics for layers 1 and 2 for the original and recalibrated model. Figure 6-26 shows the RMSE by county for the two models. Figures 6-27 and 6-28 shows the root-mean-square error for each well location in layers 1 and 2.

Analysis of all residuals indicates that the recalibrated GAM provides a significantly better fit to the historical water level data than does the original ETPV GAM. Across the entire model, the RMSE was reduced by 122 ft, the mean error was reduced by 66 ft, and the absolute error was reduced by 106 feet. By county, the RMSE was reduced by more than a factor of 2 and 3 in 26 and 19 counties, respectively, and by more than 50 ft and 100 ft in 24 and 15 counties, respectively.

At the onset of the project, a specific goal of the project was to improve the calibration metric in Upton and Reagan Counties. The RMSE for Upton and Reagan Counties produced by the original ETPV GAM is 176 ft and 269 feet, respectively. The RMSE for Upton and Reagan Counties produced by the recalibrated ETPV GAM is 9 ft and 23 feet, respectively.

Table 6-9. Calibration statistics for layer 1 for the original and recalibrated GAMs for transient conditions from 1980 to 2000

County	Recalibrated GAM			Original GAM			Number of measurements	Number of wells
	RMSE	Mean error	Absolute error	RMSE	Mean error	Absolute error		
All counties	42.2	-12.3	30.8	189.5	-142.0	152.2	2549	278
Bandera	32.0	21.5	25.8	23.2	-5.2	16.6	9	2
Brewster	9.6	-9.3	9.3	225.9	-225.8	225.8	6	1
Concho	44.2	43.0	43.0	41.5	30.8	41.4	8	2
Crane	7.3	2.2	3.6	25.3	-24.4	24.4	18	2
Crockett	38.8	-11.5	34.3	173.7	-166.8	166.8	113	17
Edwards	33.4	21.0	24.7	45.7	-23.3	29.0	27	6
Gillespie	52.5	39.0	48.7	83.1	74.5	74.5	51	5
Glasscock	40.1	-37.1	37.1	89.8	-88.0	88.0	124	10
Irion	41.2	-36.0	36.0	124.5	-105.9	105.9	45	5
Jeff Davis	38.0	16.4	30.1	162.7	159.3	159.3	23	4
Kerr	22.0	20.5	20.5	21.6	-18.8	18.8	10	1
Kimble	56.6	22.6	33.4	57.2	26.3	33.7	36	7
Kinney	51.1	-18.8	37.3	66.3	-51.6	52.2	40	3
Loving	5.8	-4.4	5.1	19.2	-17.3	17.3	20	2
Menard	21.8	-5.8	19.4	54.6	-19.6	37.9	77	10
Pecos	39.5	-14.6	31.0	230.8	-210.3	210.3	333	31
Reagan	29.8	-17.5	18.4	89.7	-82.1	82.1	45	6
Real	38.4	23.3	27.9	44.3	-23.0	39.5	20	3
Reeves	53.1	-13.9	37.3	301.6	-274.3	278.9	555	55

Table 6-9. Calibration statistics for layer 1 for the original and recalibrated GAMs for transient conditions from 1980 to 2000 (Continued)

County	Recalibrated GAM			Original GAM			Number of measurements	Number of wells
	RMSE	Mean error	Absolute error	RMSE	Mean error	Absolute error		
Schleicher	33.3	-6.7	27.7	158.3	-156.7	156.7	115	14
Sterling	20.4	-19.9	19.9	40.4	-38.5	38.5	9	1
Sutton	54.4	-30.2	44.6	177.8	-167.4	168.3	159	17
Terrell	79.6	-27.0	78.7	292.1	-281.5	281.5	5	2
Tom	24.2	-15.0	16.5	97.5	-85.6	85.6	14	2
Upton	4.2	-0.6	3.8	80.3	-80.1	80.1	6	1
Val	55.0	-23.0	40.7	62.0	-27.8	45.9	197	23
Ward	16.2	-5.5	12.5	111.9	-98.3	98.4	369	34
Winkler	31.5	-12.0	27.8	133.8	-116.0	116.0	115	12

Table 6-10. Calibration statistics for layer 2 for the original and recalibrated GAMs for transient conditions from 1980 to 2000.

County	Recalibrated GAM			Original GAM			Number of measurements	Number of wells
	RMSE	Mean error	Absolute error	RMSE	Mean error	Absolute error		
All counties	76.0	-11.3	46.4	175.1	-3.5	134.6	2224	296
Bandera	123.2	-36.2	80.2	134.1	-59.7	96.7	258	38
Bexar	128.5	-4.6	87.7	172.0	-28.8	129.2	77	10
Blanco	73.9	12.4	54.8	200.1	-179.1	180.6	82	11
Comal	55.4	-28.1	43.1	97.3	-68.2	84.2	34	8
Concho	26.5	-19.1	19.1	28.7	-22.7	22.7	22	2
Crockett	16.5	-4.4	11.0	64.1	-41.3	48.1	78	10
Ector	9.1	2.3	7.2	50.1	-43.4	47.0	117	17
Edwards	163.5	-78.3	116.0	164.0	-81.6	115.7	38	4
Gillespie	74.7	8.8	55.8	129.0	-20.6	115.1	266	41
Glasscock	12.0	1.3	9.6	384.6	380.0	380.0	125	13
Hays	87.0	-28.5	65.6	201.4	-165.8	195.4	72	15
Irion	42.0	-0.9	22.5	70.9	-60.7	69.3	21	2
Kendall	83.8	-10.0	54.3	117.1	-59.0	91.1	246	31
Kerr	68.0	-51.2	56.5	147.2	-132.6	140.1	203	30
Kimble	54.2	-13.4	45.7	163.9	89.7	116.9	55	6
Kinney	55.0	-7.4	40.6	113.1	-66.6	77.5	12	3
Mcculloch	13.7	13.3	13.3	63.9	63.8	63.8	12	1
Menard	9.0	3.9	4.7	181.3	142.4	142.4	33	3
Midland	12.3	-0.7	10.1	135.0	82.1	89.2	66	7
Nolan	40.8	-1.2	34.5	84.7	38.7	74.7	35	4
Reagan	20.6	1.9	14.4	309.5	265.8	274.9	123	15
Real	103.7	17.3	90.1	170.4	-53.2	147.4	71	6
Schleicher	23.8	20.1	20.1	81.8	-80.8	80.8	11	1
Sterling	26.4	12.6	13.5	107.6	103.2	103.2	10	1
Terrell	85.1	2.8	77.4	157.6	19.0	119.6	17	3
Tom	29.5	-19.2	26.9	59.0	-49.9	55.0	12	2
Travis	30.4	-9.1	22.0	149.0	-84.3	103.8	62	7
Upton	9.7	-0.9	7.6	183.0	172.9	172.9	66	5

Table 6-11. Calibration statistics for layers 1 and 2 for the original and recalibrated GAMs for transient conditions from 1980 to 2000.

County	Recalibrated GAM			Original GAM			Number of measurements	Number of wells
	RMSE	Mean error	Absolute error	RMSE	Mean error	Absolute error		
All counties	60.4	-11.8	38.1	182.9	-77.5	144.0	4773	574
Bandera	121.2	-34.2	78.3	131.9	-57.9	94.0	267	40
Bexar	128.5	-4.6	87.7	172.0	-28.8	129.2	77	10
Blanco	73.9	12.4	54.8	200.1	-179.1	180.6	82	11
Brewster	9.6	-9.3	9.3	225.9	-225.8	225.8	6	1
Comal	55.4	-28.1	43.1	97.3	-68.2	84.2	34	8
Concho	32.2	-2.5	25.5	32.6	-8.5	27.7	30	4
Crane	7.3	2.2	3.6	25.3	-24.4	24.4	18	2
Crockett	31.7	-8.6	24.8	139.7	-115.5	118.3	191	27
Ector	9.1	2.3	7.2	50.1	-43.4	47.0	117	17
Edwards	126.9	-37.1	78.1	128.8	-57.4	79.7	65	10
Gillespie	71.6	13.6	54.7	122.8	-5.3	108.6	317	46
Glasscock	29.5	-17.8	23.3	279.8	146.9	234.6	249	23
Hays	87.0	-28.5	65.6	201.4	-165.8	195.4	72	15
Irion	41.4	-24.8	31.7	110.3	-91.5	94.3	66	7
Jeff Davis	38.0	16.4	30.1	162.7	159.3	159.3	23	4
Kendall	83.8	-10.0	54.3	117.1	-59.0	91.1	246	31
Kerr	66.5	-47.8	54.8	143.8	-127.3	134.4	213	31
Kimble	55.1	0.8	40.8	132.4	64.6	84.0	91	13
Kinney	52.0	-16.1	38.1	79.6	-55.1	58.0	52	6
Loving	5.8	-4.4	5.1	19.2	-17.3	17.3	20	2
McCulloch	13.7	13.3	13.3	63.9	63.8	63.8	12	1
Menard	18.9	-2.9	15.0	109.3	29.0	69.2	110	13
Midland	12.3	-0.7	10.1	135.0	82.1	89.2	66	7
Nolan	40.8	-1.2	34.5	84.7	38.7	74.7	35	4
Pecos	39.5	-14.6	31.0	230.8	-210.3	210.3	333	31
Reagan	23.4	-3.3	15.5	268.9	172.6	223.2	168	21
Real	93.4	18.6	76.4	151.9	-46.6	123.7	91	9
Reeves	53.1	-13.9	37.3	301.6	-274.3	278.9	555	55
Schleicher	32.6	-4.4	27.0	153.1	-150.1	150.1	126	15
Sterling	23.7	-2.8	16.5	82.9	36.1	72.5	19	2
Sutton	54.4	-30.2	44.6	177.8	-167.4	168.3	159	17
Terrell	83.9	-4.0	77.7	196.4	-49.3	156.4	22	5
Travis	30.4	-9.1	22.0	149.0	-84.3	103.8	62	7
Upton	9.4	-0.9	7.3	176.8	151.8	165.2	72	6
Val	55.0	-23.0	40.7	62.0	-27.8	45.9	197	23
Ward	16.2	-5.5	12.5	111.9	-98.3	98.4	369	34
Winkler	31.5	-12.0	27.8	133.8	-116.0	116.0	115	12

6.5.2 Groundwater contribution to stream flows

A commonly used metric to evaluate a GAM calibration is a comparison between estimated and simulated contribution by groundwater to surface water. For the original ETPV GAM, Anaya and Jones (2009) perform this evaluation by a visual comparison of the bottoms of the rise and fall of the stream hydrographs to the simulated groundwater discharge to streams. As mentioned previously, the PEST calibration did not include groundwater contribution to stream flows as a

calibration target but these contributions were monitored throughout the PEST simulations and are shown in Figure 6-29.

Figure 6-29 shows the measured river flow and the groundwater contribution to river flow simulated by the original and recalibrated ETPV GAM for the 14 river gauges. At all 14 gauges, the recalibrated GAM results show less groundwater contribution to stream flow than do the original GAM results. This is consistent with the fact that the original ETPV has a bias of over-predicting water levels by about 80 feet, whereas the recalibrated ETPV GAM has a bias of over-predicting the water levels by about 10 feet (first line of Table 6-9).

A comparison of the plots in Figure 6-29 shows that despite providing lower estimates of groundwater discharge to stream flow, the results for the two models are comparable for the majority of the gages. In general, the higher flows provide a better match to the field data except for the gage on the Pecos River near Girvin, for which the lower flows produced by the recalibrated model provide a better match to the field data.

As previously mentioned, the stream flow targets were intentionally omitted from this PEST calibration because of concerns that the model may not be adequately constructed to accurately simulate stream flow targets. Three issues of concerns are that the model lacks sufficient vertical layering, that elevations of the channel bottoms are over estimated, and the areal dimensions of the grid cells are too large.

Young and others, (2009) show that groundwater contributions to streams comes from a shallow flow zone near the river and that this zone should be modeled as a separate layer from the zone where regional pumping occurs. To check to see whether the large grid layers may be a potential problem, the average of the water level measurements within a 3-mile radius of a river gage was compared to the elevation of the stream bottom at the river gauge location. As shown in Table 6-10, out of the 12 river gage locations within the vicinity of water level measurements, 7 of the stream bottom elevations are 30 or more feet above the average nearby water level measurements. Thus, for the ETPV GAM to adequately model both the impacts of pumping on water levels in the regional flow system, which often has water levels below the elevations of nearby rivers, and the contribution of groundwater flow to surface water, which requires higher water levels in the local groundwater flow system than in the streams, the ETPV model should have the local and regional flow systems represented by separate layers and not as a single layer as the ETPV GAM currently does.

Young and others (2009) also demonstrate that over estimates of stream bottoms occur in their model of the Texas Gulf Coast using 30 meters digital elevation models (DEMs). Young and others (2009) therefore support the use of 10-meter DEMs if accurate estimates of groundwater contributions to stream flow are desired. Because the stream bottom elevations in Table 6-12 were developed from 90-meter DEMs, there is a very good likelihood that the majority of the bottom elevations may be considerably too high.

Table 6-12. Water balance for layers 1 and 2 for the original and recalibrated GAMs for transient conditions from 1980 to 2000.

Station number	Station description	Elevation (ft msl)		Difference (ft)
		Stream bottom	Average groundwater elevation*	
08449400	Devils River at Pafford Crossing near Comstock Tx	1191	1137	-53
08167000	Guadalupe River at Comfort Tx	1396	1361	-35
08165500	Guadalupe River at Hunt Tx	1721	1549	-172
08166200	Guadalupe River at Kerrville Tx	1597	1490	-107
08167800	Guadalupe River at Sattler Tx	747	836	89
08167500	Guadalupe River near Spring Branch Tx	947	NC	NC
08150000	Llano River near Junction Tx	1636	NC	NC
08178880	Medina River at Bandera Tx	1196	1041	-155
08179000	Medina River near Pipe Creek Tx	1098	1066	-31
08446500	Pecos River near Girvin Tx	2275	NC	NC
08447410	Pecos River near Langtry Tx	1191	1072	-119
08152900	Pedernales River near Fredericksburg Tx	1557	1585	28
08144500	San Saba River at Menard Tx	1882	1907	25
08128000	South Concho River at Christoval Tx	2003	2034	31

* Based on water levels measured within a 3-mile radius

NC - not calculated because no water level data within a 3-mile radius

Another concern with the ETPV GAM’s limitation to accurately predict stream flow is the size of grid cell relative to the area of capture associated with a stream. As stated by Anaya and Jones (2009) “the actual discharge to streams occurs within small area averaging 50 feet wide, compared to the 1 square mile of model grid cell.”

TWDB Report ##: Final – Application of PEST to Re-Calibrate
the Groundwater Availability Model for the Edwards-Trinity (Plateau) and Pecos Valley Aquifers

TWDB Report ##: Final – Application of PEST to Re-Calibrate
the Groundwater Availability Model for the Edwards-Trinity (Plateau) and Pecos Valley Aquifers

```

12 0 1.0 -5 3
'node1' /pdata/data1/eddt/
'node2' /pdata/data2/eddt/
'node3' /pdata/data3/eddt/
'node4' /pdata/data4/eddt/
'node5' /pdata/data5/eddt/
'node6' /pdata/data6/eddt/
'node7' /pdata/data7/eddt/
'node8' /pdata/data8/eddt/
'node9' /pdata/data9/eddt/
'node10' /pdata/data10/eddt/
'node11' /pdata/data11/eddt/
'node12' /pdata/data12/eddt/
1000 1000 1000 1000 1000 1000 1000 1000 1000 1000 1000 1000

```

Figure 6-1. Parallel PEST run management file.

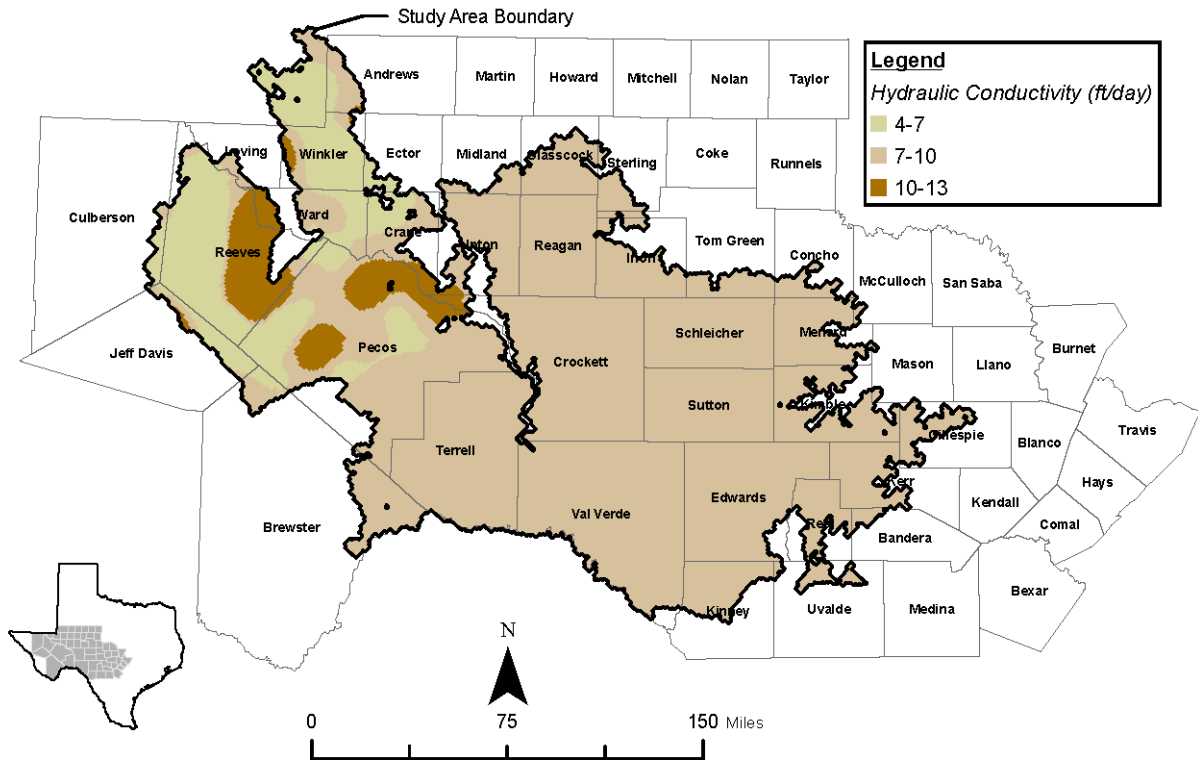


Figure 6-2. Hydraulic conductivity in layer 1 from the PEST recalibration.

TWDB Report #: Final – Application of PEST to Re-Calibrate the Groundwater Availability Model for the Edwards-Trinity (Plateau) and Pecos Valley Aquifers

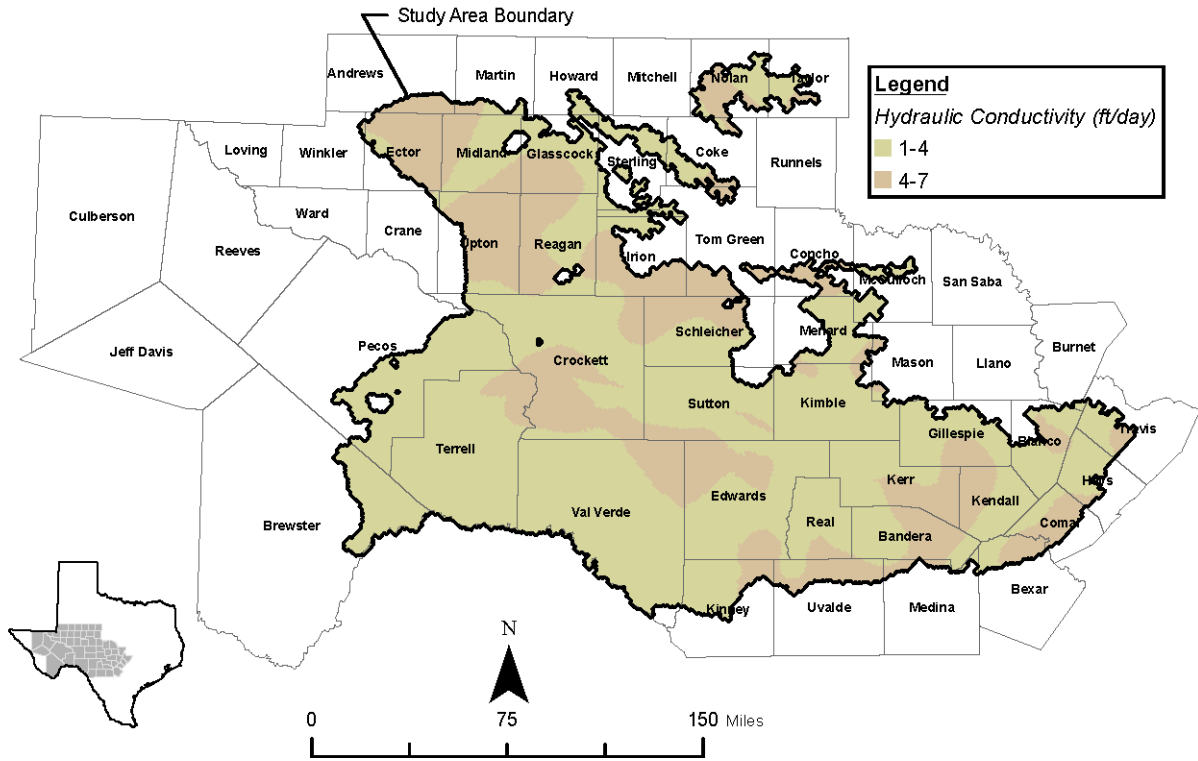


Figure 6-3. Hydraulic conductivity in layer 2 from the PEST recalibration

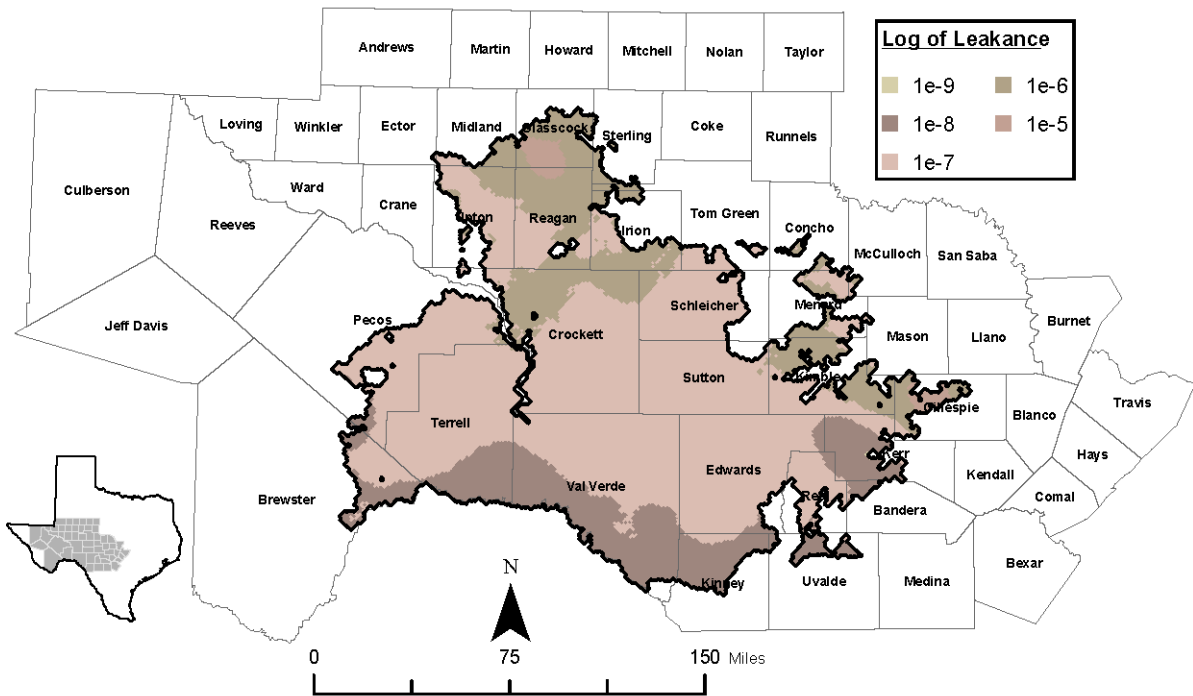


Figure 6-4. Log (to base 10) of vertical conductance between layers 1 and 2 from the PEST recalibration.

TWDB Report ##: Final – Application of PEST to Re-Calibrate
the Groundwater Availability Model for the Edwards-Trinity (Plateau) and Pecos Valley Aquifers

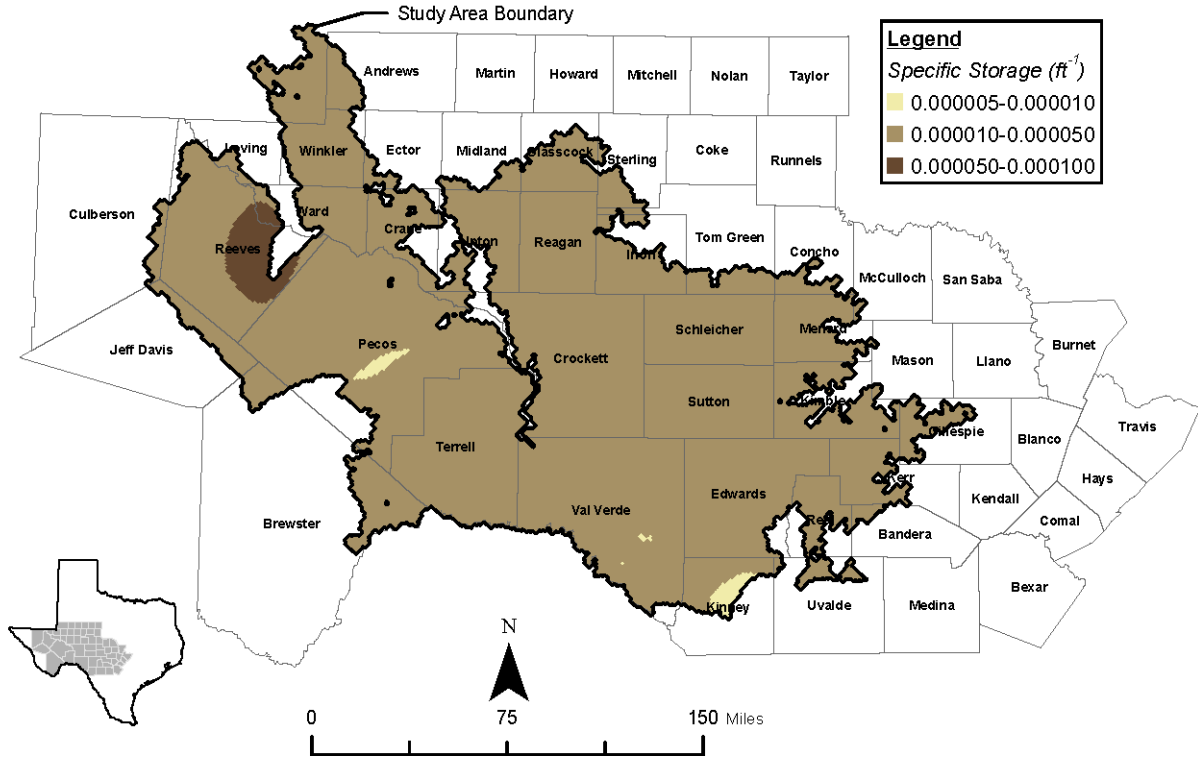


Figure 6-5. Specific storage in layer 1 from the PEST recalibration.

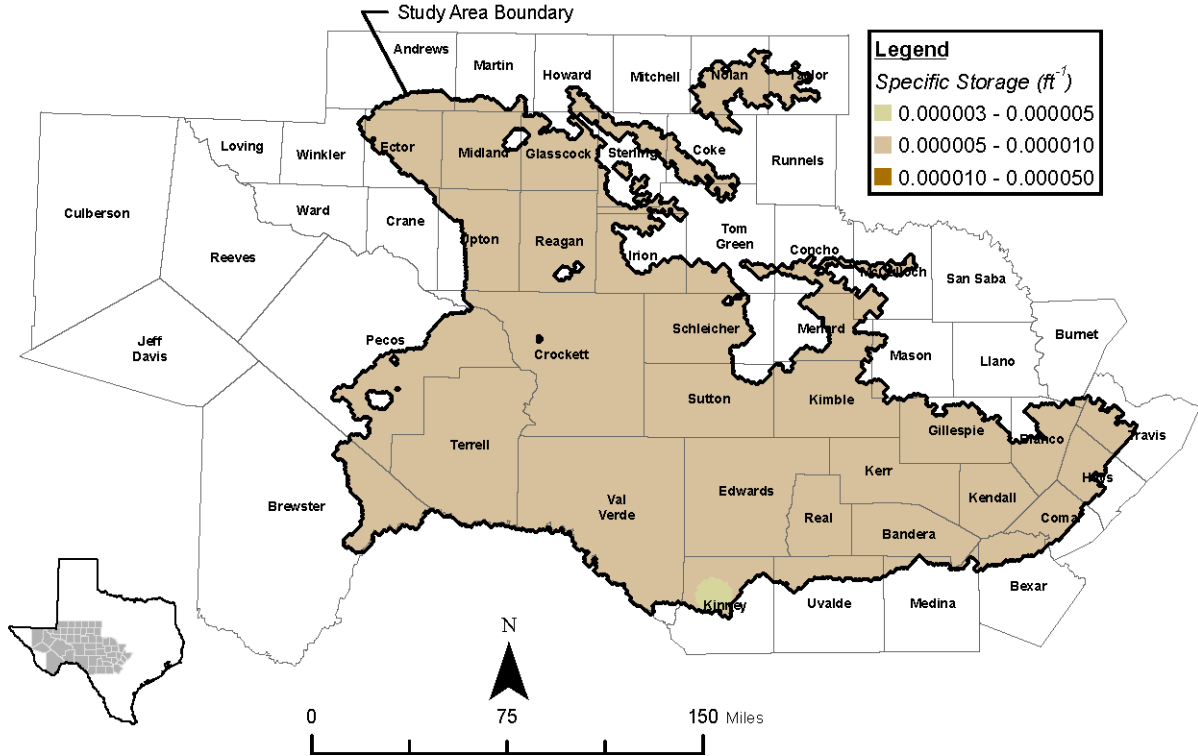


Figure 6-6. Specific storage in layer 2 from the PEST recalibration.

TWDB Report #: Final – Application of PEST to Re-Calibrate the Groundwater Availability Model for the Edwards-Trinity (Plateau) and Pecos Valley Aquifers

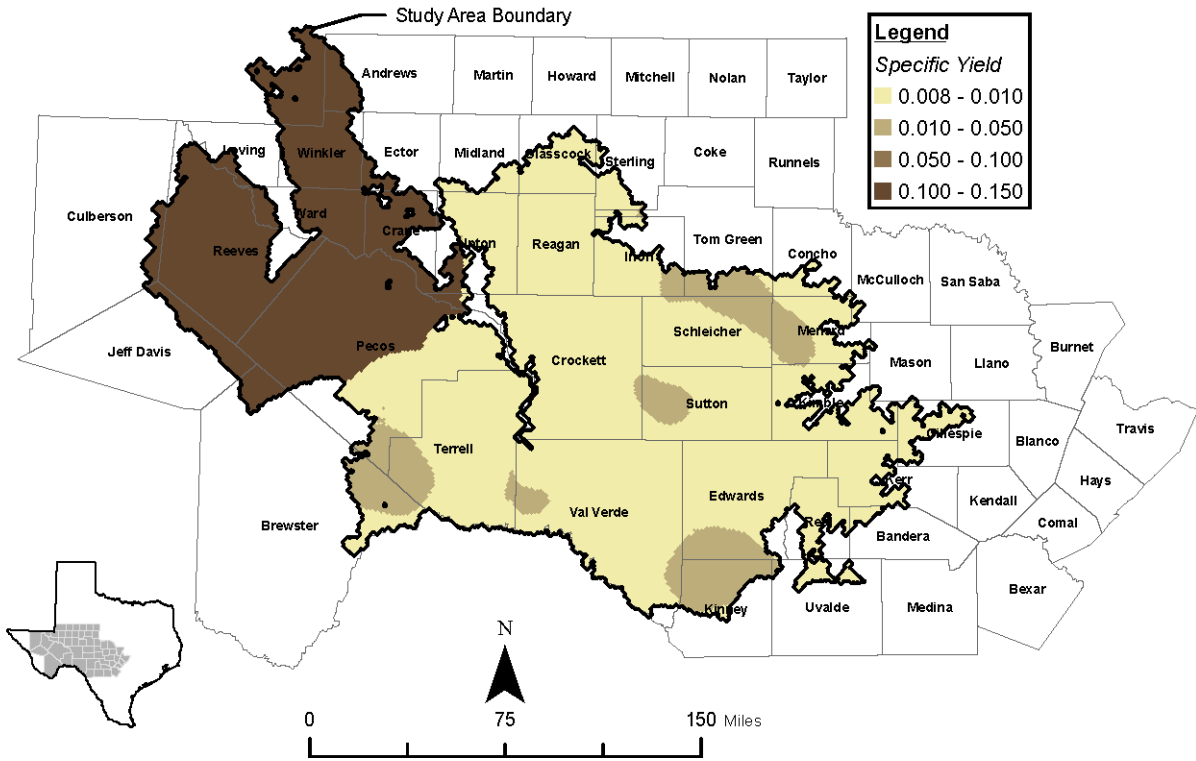


Figure 6-7. Specific yield in layer 1 from the PEST recalibration.

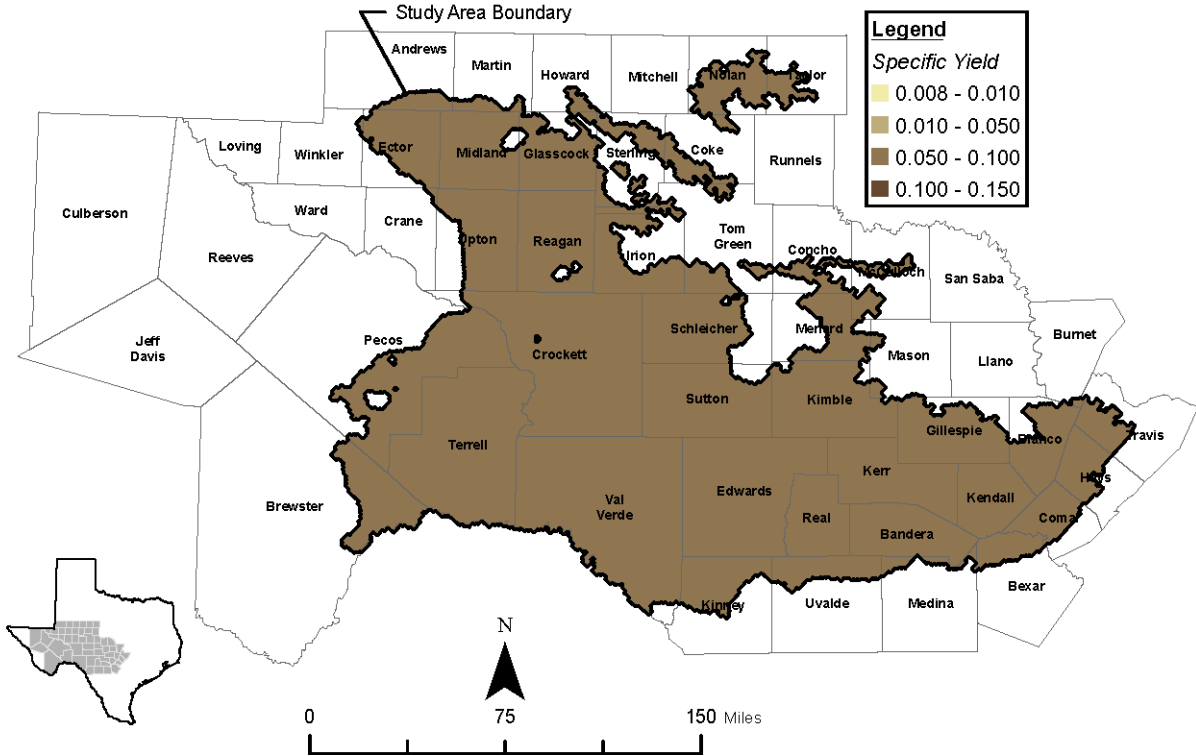


Figure 6-8. Specific yield in layer 2 from the PEST recalibration.

TWDB Report #: Final – Application of PEST to Re-Calibrate the Groundwater Availability Model for the Edwards-Trinity (Plateau) and Pecos Valley Aquifers

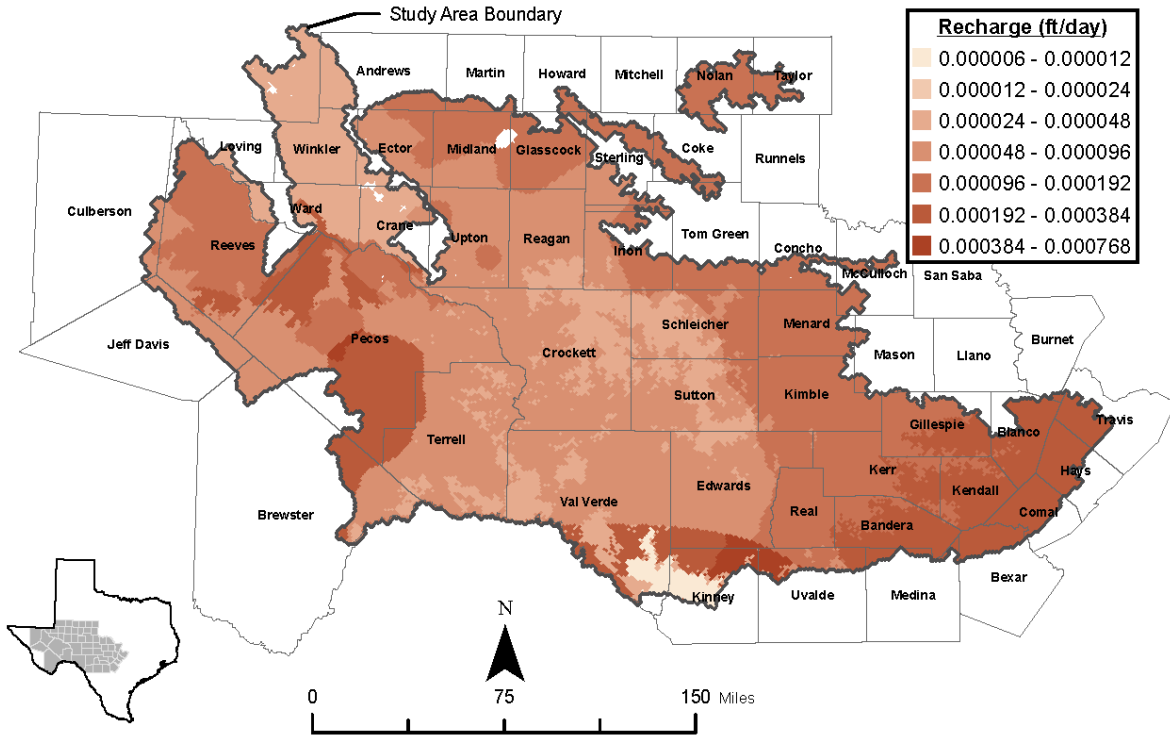


Figure 6-9. Recharge distribution for 1980.

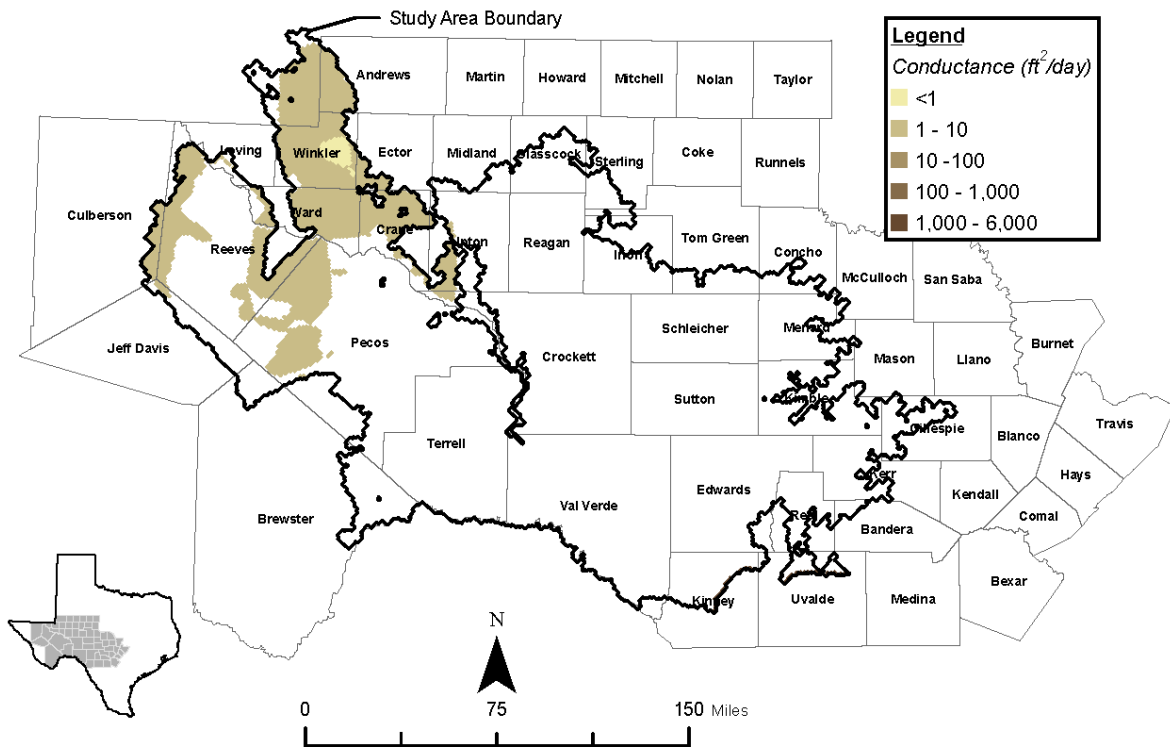


Figure 6-10. Conductance values (ft²/day) for general head boundary cells in layer 1 from the PEST recalibration.

TWDB Report #: Final – Application of PEST to Re-Calibrate the Groundwater Availability Model for the Edwards-Trinity (Plateau) and Pecos Valley Aquifers

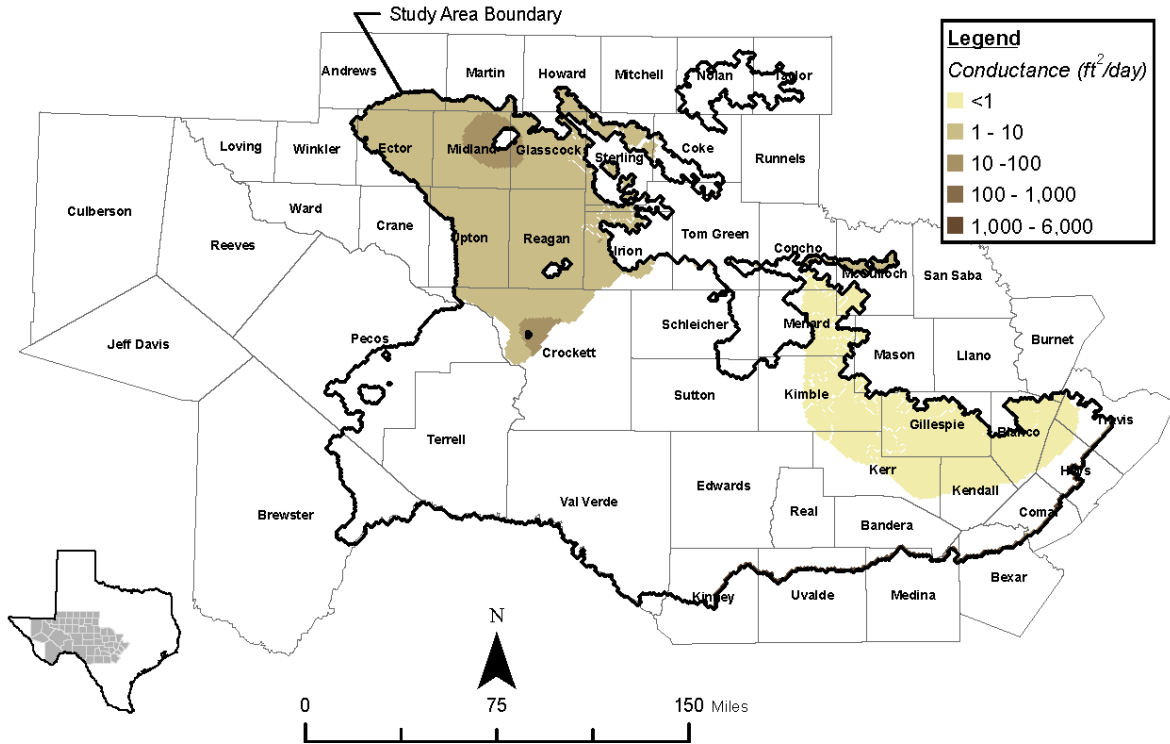


Figure 6-11. Conductance values (ft²/day) for general head boundary cells in layer 2 from the PEST recalibration.

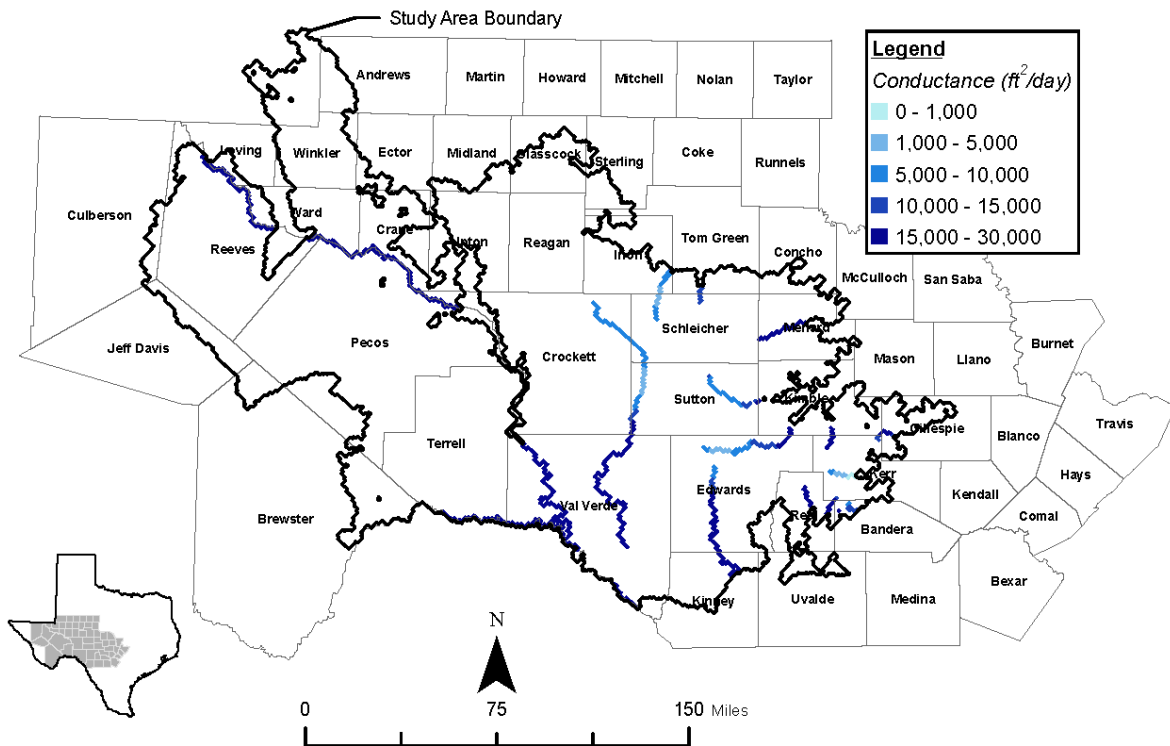


Figure 6-12. The conductance values (ft²/day) for streams in layer 1 from the PEST recalibration.

TWDB Report ##: Final – Application of PEST to Re-Calibrate the Groundwater Availability Model for the Edwards-Trinity (Plateau) and Pecos Valley Aquifers

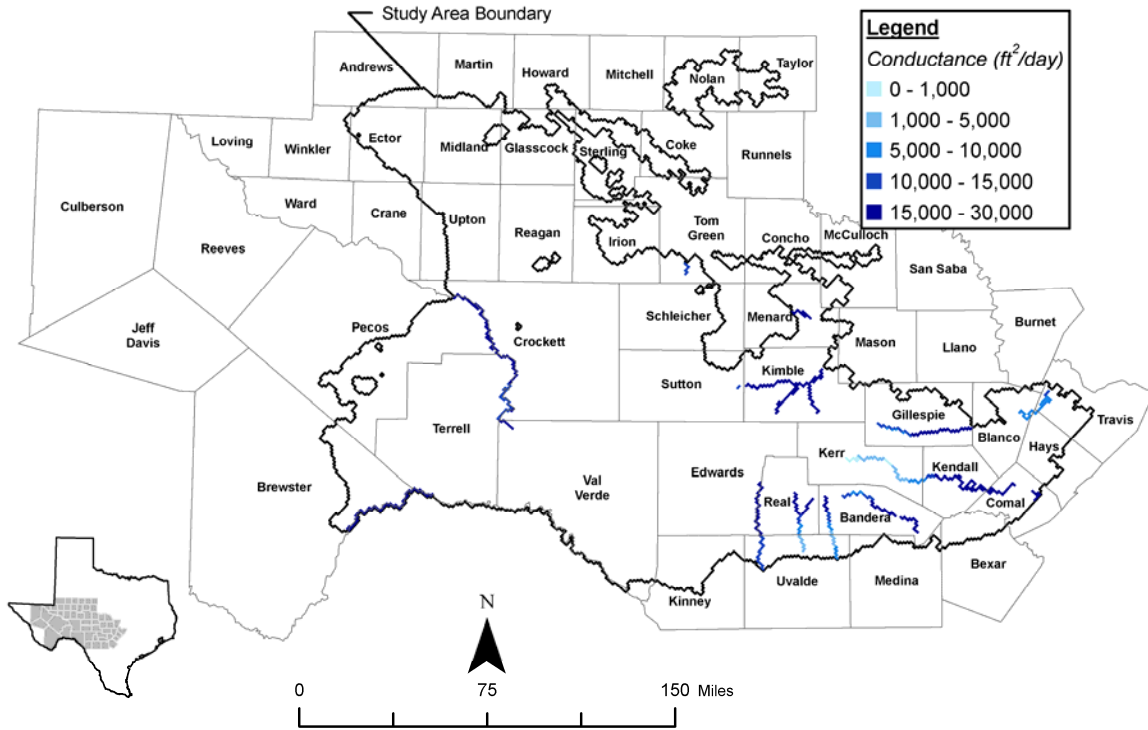


Figure 6-13. The conductance values (ft²/day) for streams in layer 2 from the PEST recalibration.

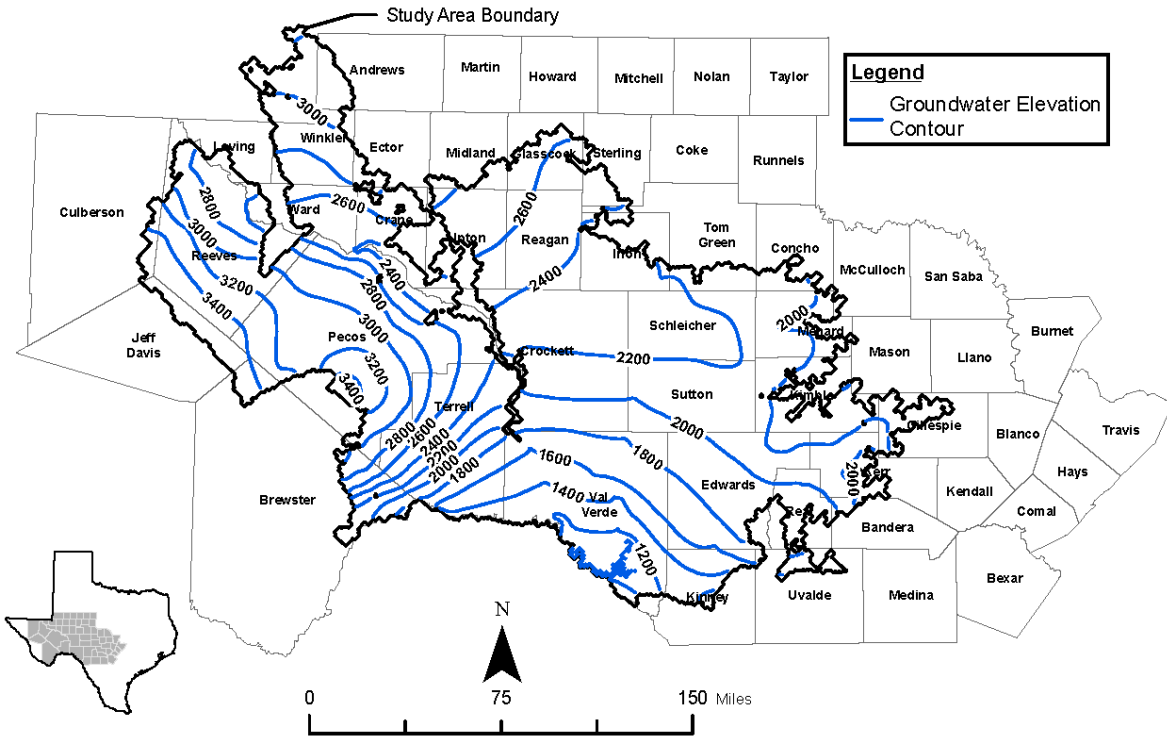


Figure 6-14. Simulated water levels for 1930 (presumed predevelopment conditions) in layer 1 from the PEST recalibration.

TWDB Report ##: Final – Application of PEST to Re-Calibrate the Groundwater Availability Model for the Edwards-Trinity (Plateau) and Pecos Valley Aquifers

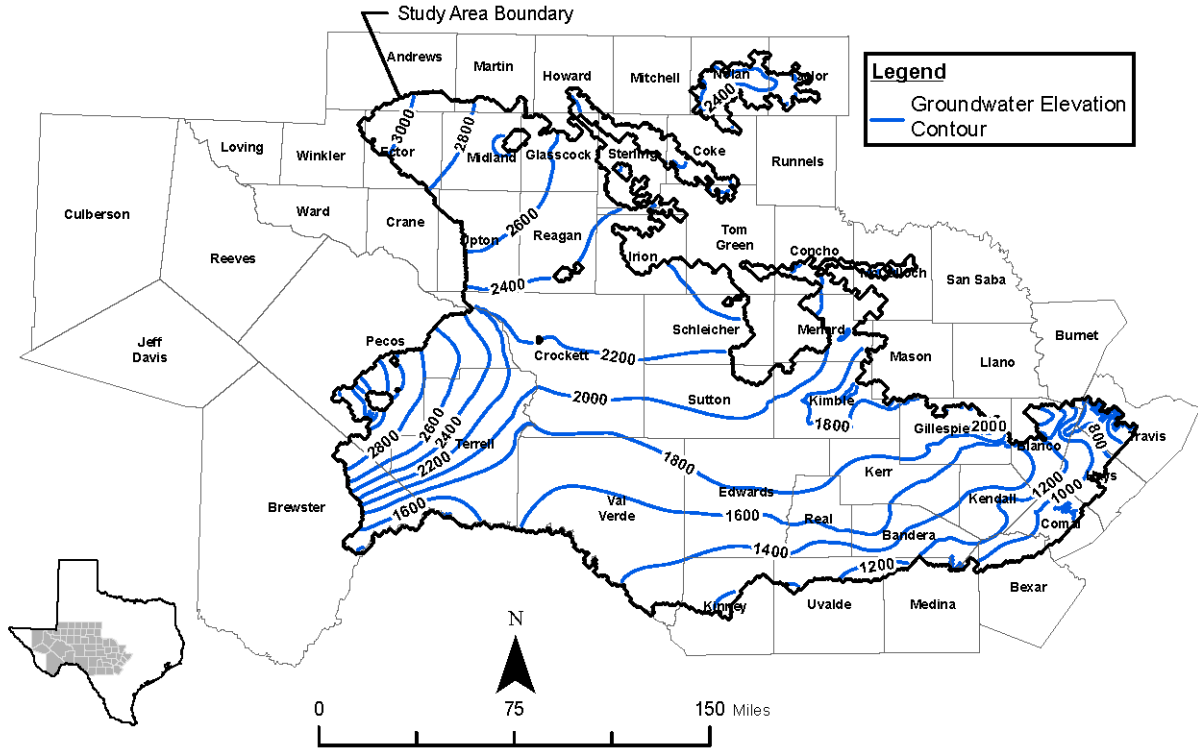


Figure 6-15. Simulated water levels for 1930 (presumed predevelopment conditions) in layer 2 from the PEST recalibration.

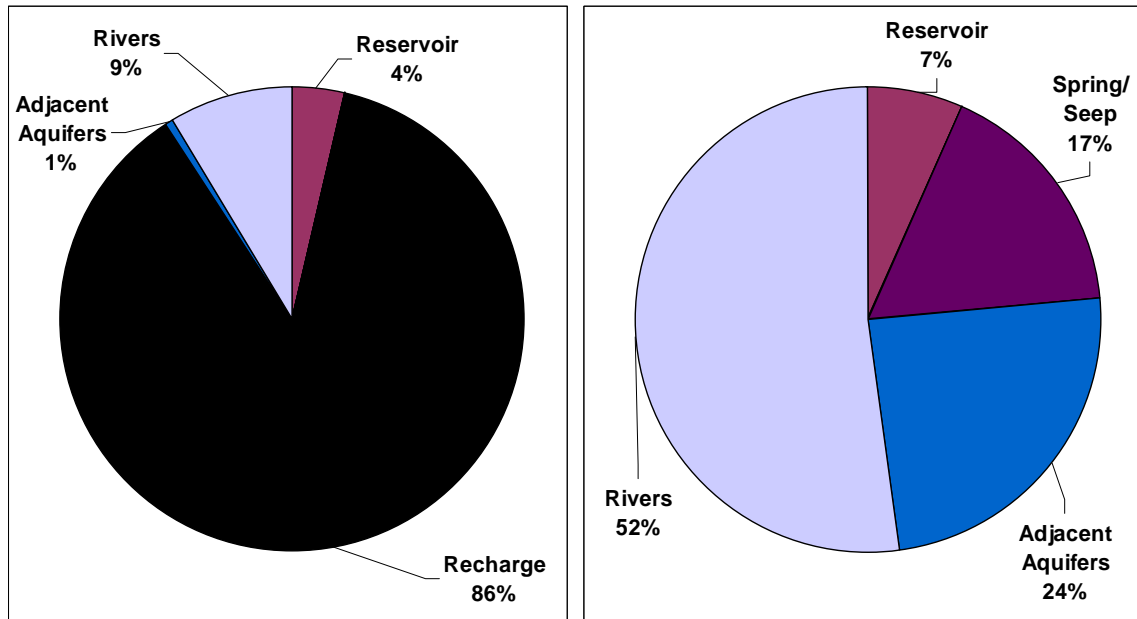


Figure 6-16. Water budget for the steady-state model for 1930. Inflows are shown in (a), and outflows are shown in (b).

TWDB Report ##: Final – Application of PEST to Re-Calibrate the Groundwater Availability Model for the Edwards-Trinity (Plateau) and Pecos Valley Aquifers

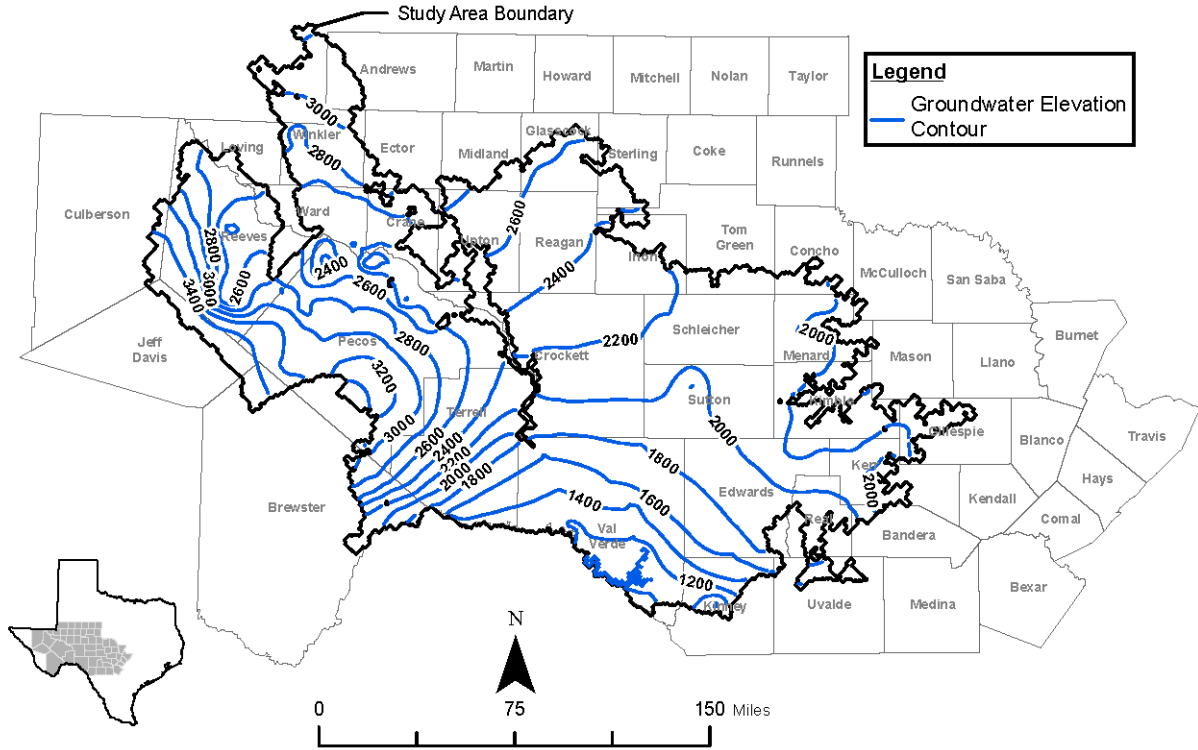


Figure 6-17. Simulated water levels for 1980 in layer 1 from the PEST recalibration.

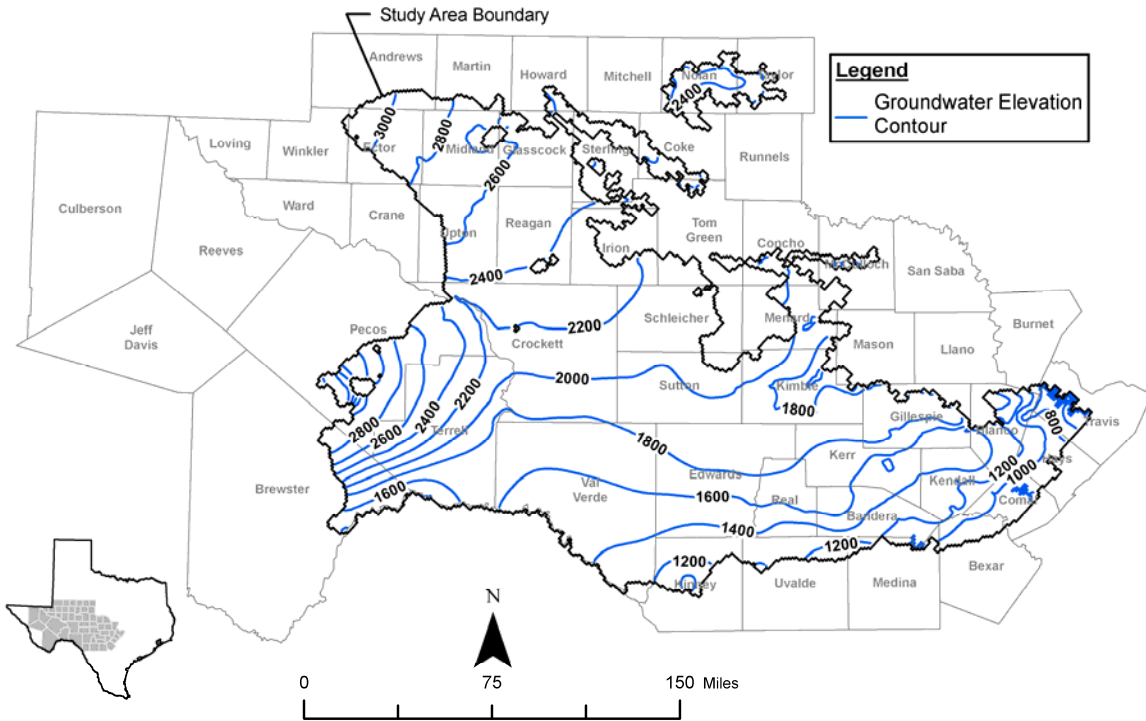


Figure 6-18. Simulated water levels for 1980 in layer 2 from the PEST recalibration.

TWDB Report #: Final – Application of PEST to Re-Calibrate
the Groundwater Availability Model for the Edwards-Trinity (Plateau) and Pecos Valley Aquifers

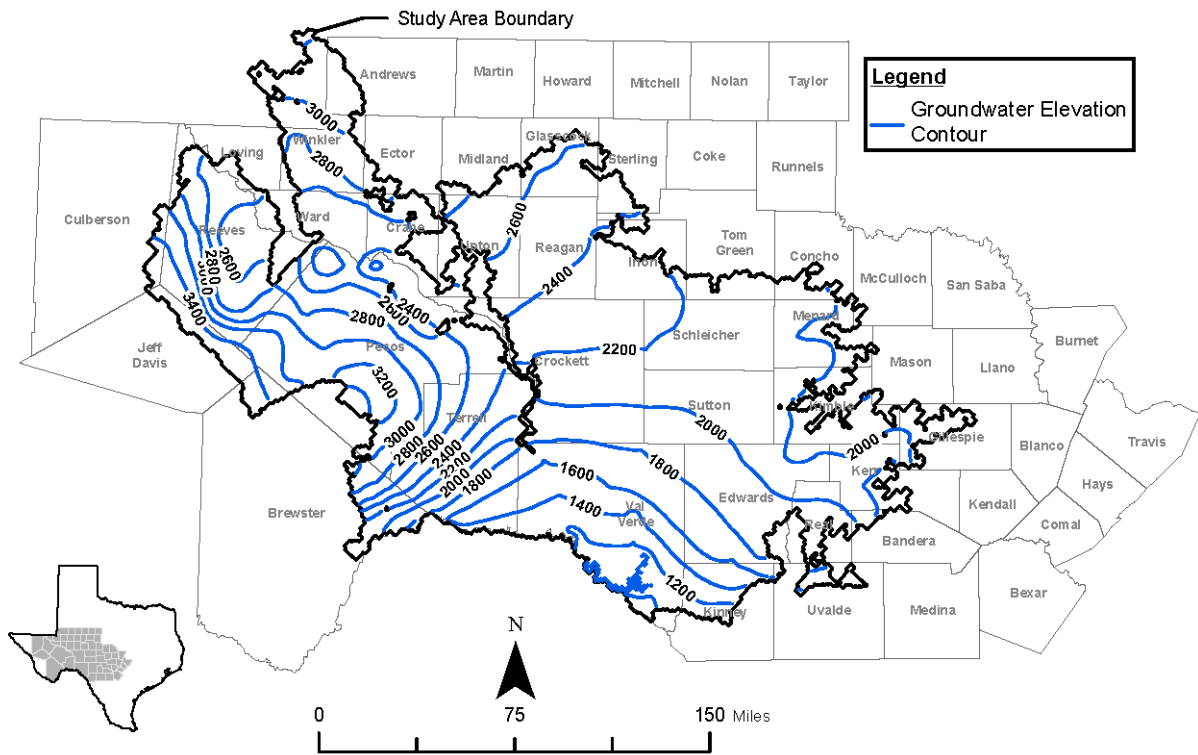


Figure 6-19. Simulated water levels for 2000 in layer 1 from the PEST recalibration.

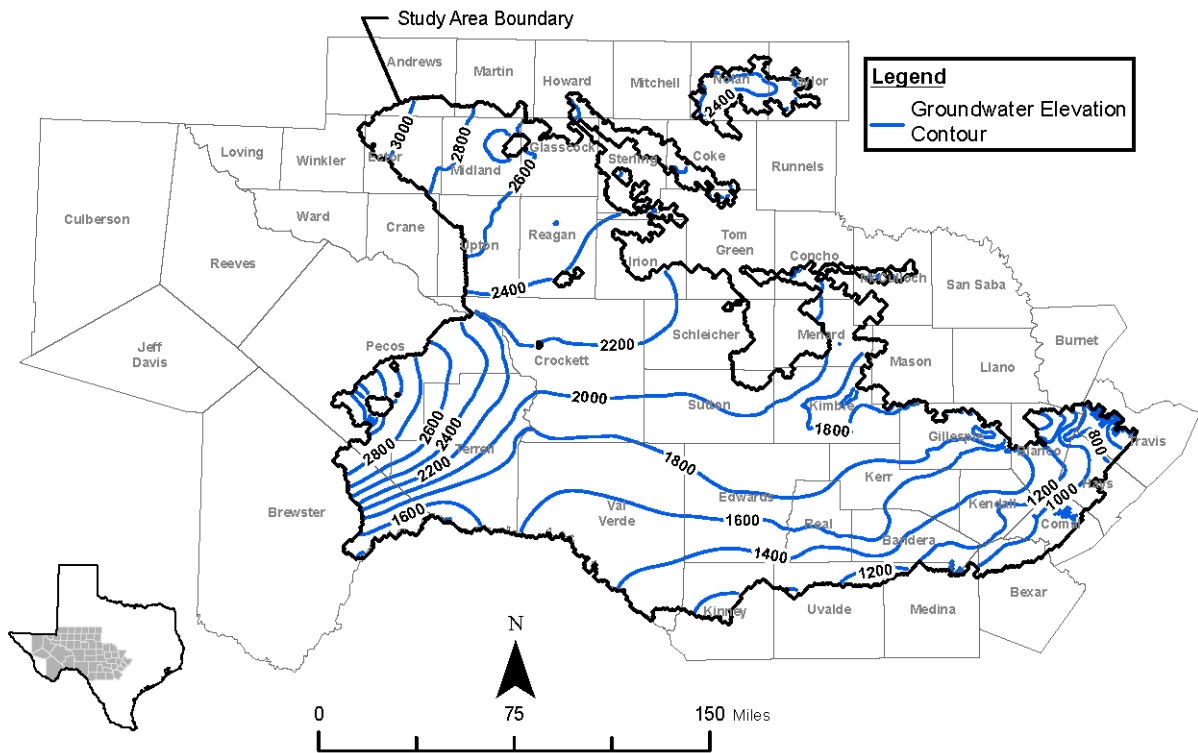


Figure 6-20. Simulated water levels for 2000 in layer 2 from the PEST recalibration.

TWDB Report #: Final – Application of PEST to Re-Calibrate the Groundwater Availability Model for the Edwards-Trinity (Plateau) and Pecos Valley Aquifers

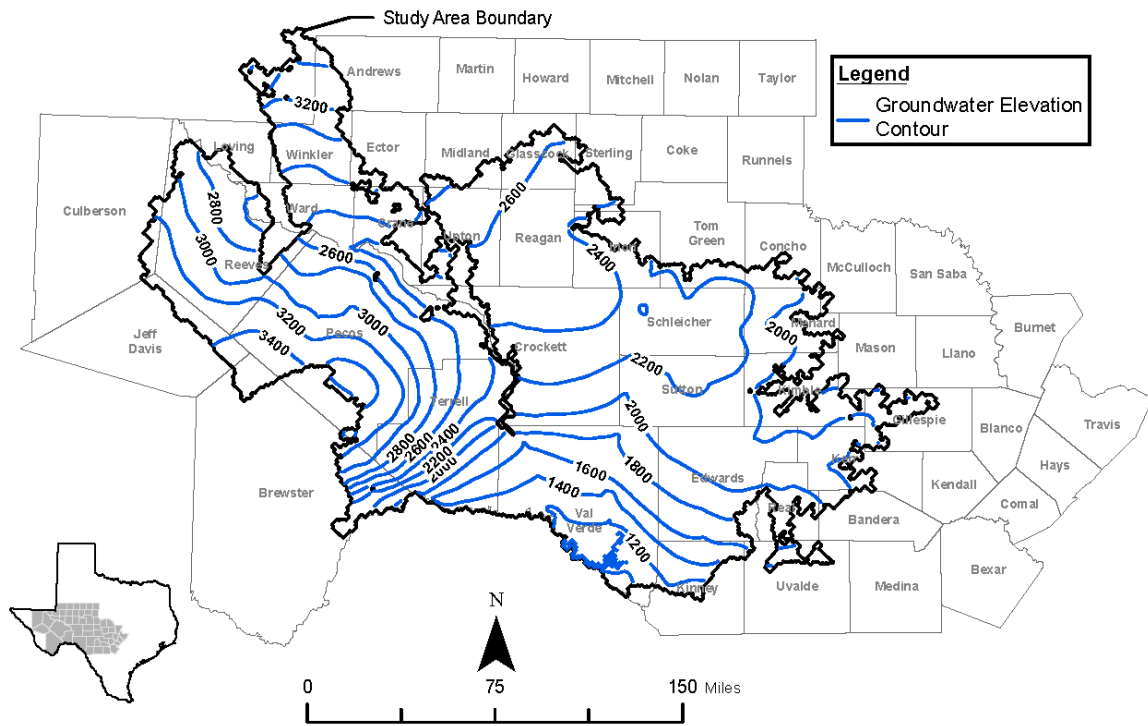


Figure 6-21. Simulated water levels for 2000 in layer 1 from the ETPV GAM. Source: Anaya and Jones, 2009.

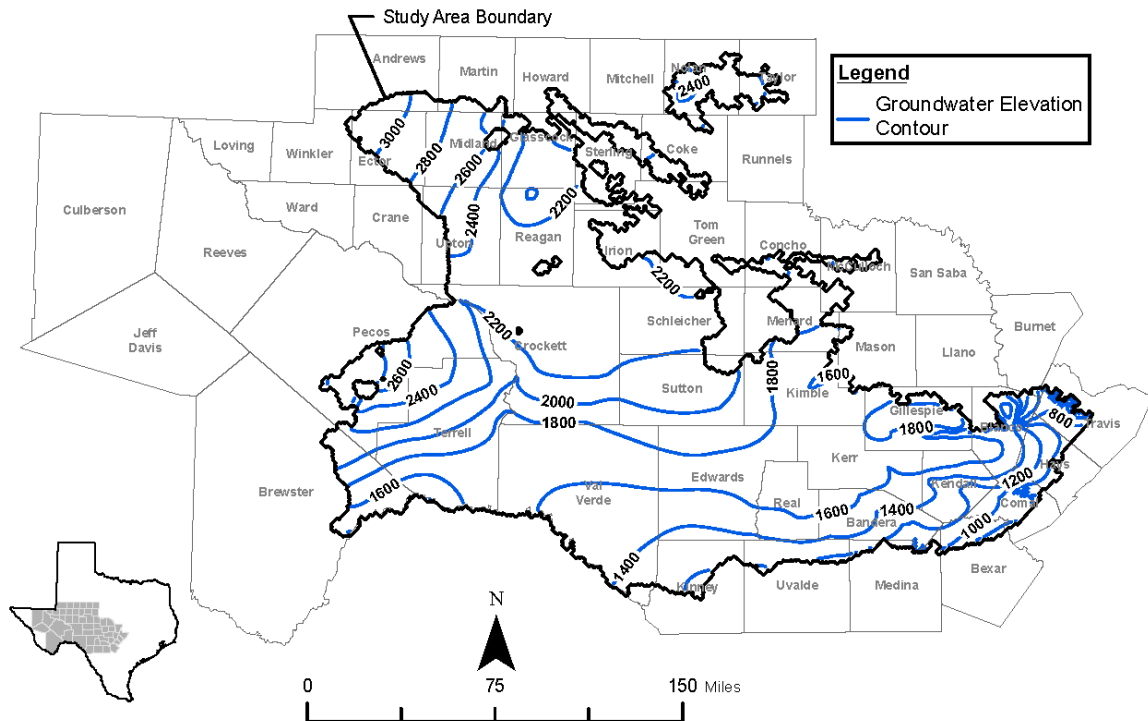
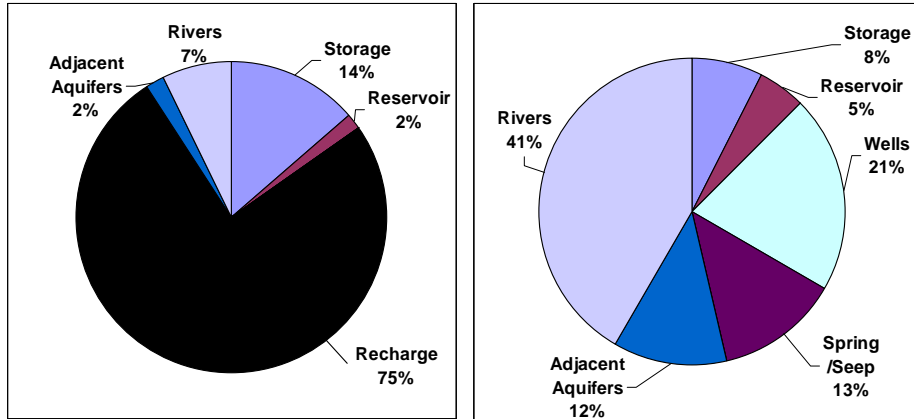


Figure 6-22. Simulated water levels for 2000 in layer 2 from the ETPV GAM. Source: Anaya and Jones, 2009.

Original ETPV GAM



Recalibrated ETPV GAM

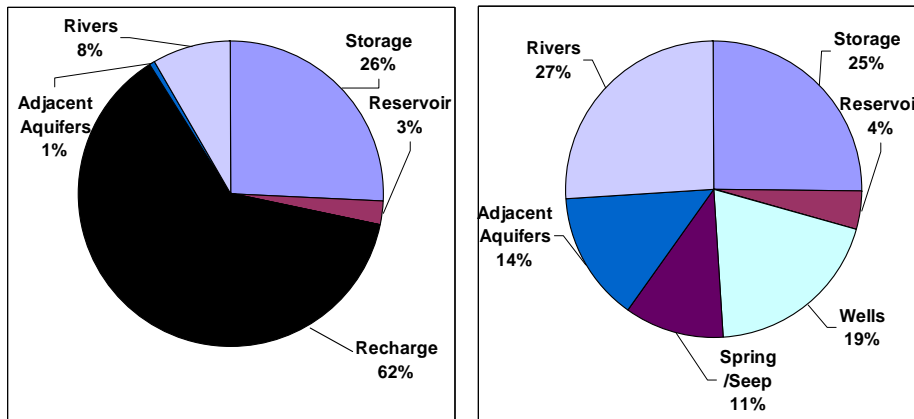


Figure 6-23. Average annual water budget for the 1980 to 2000 simulation for the original and recalibrated ETPV GAM (inflow is on left and outflow is on right).

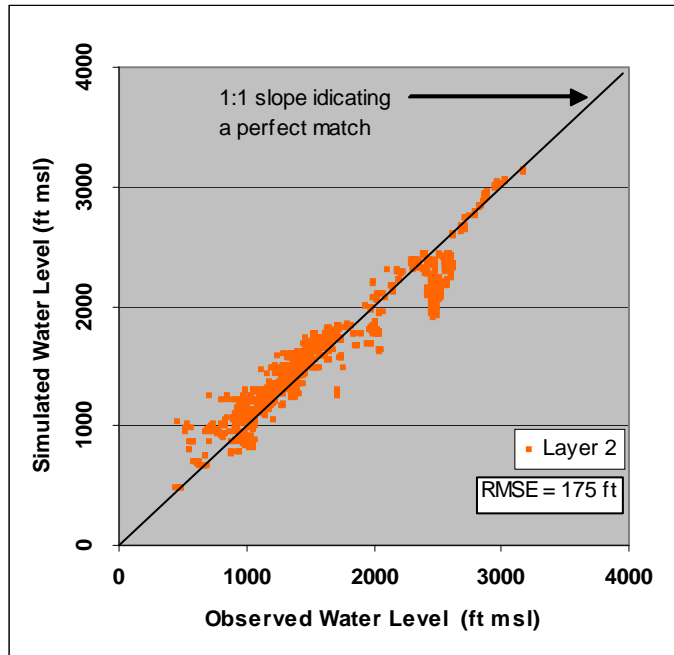
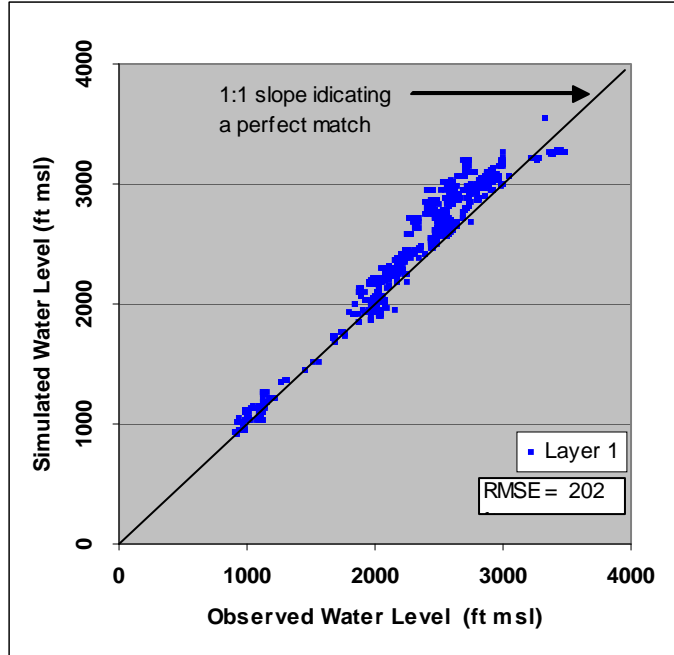


Figure 6-24. Comparison of observed versus simulated water levels for layers 1 and 2 from the original ETPV GAM.

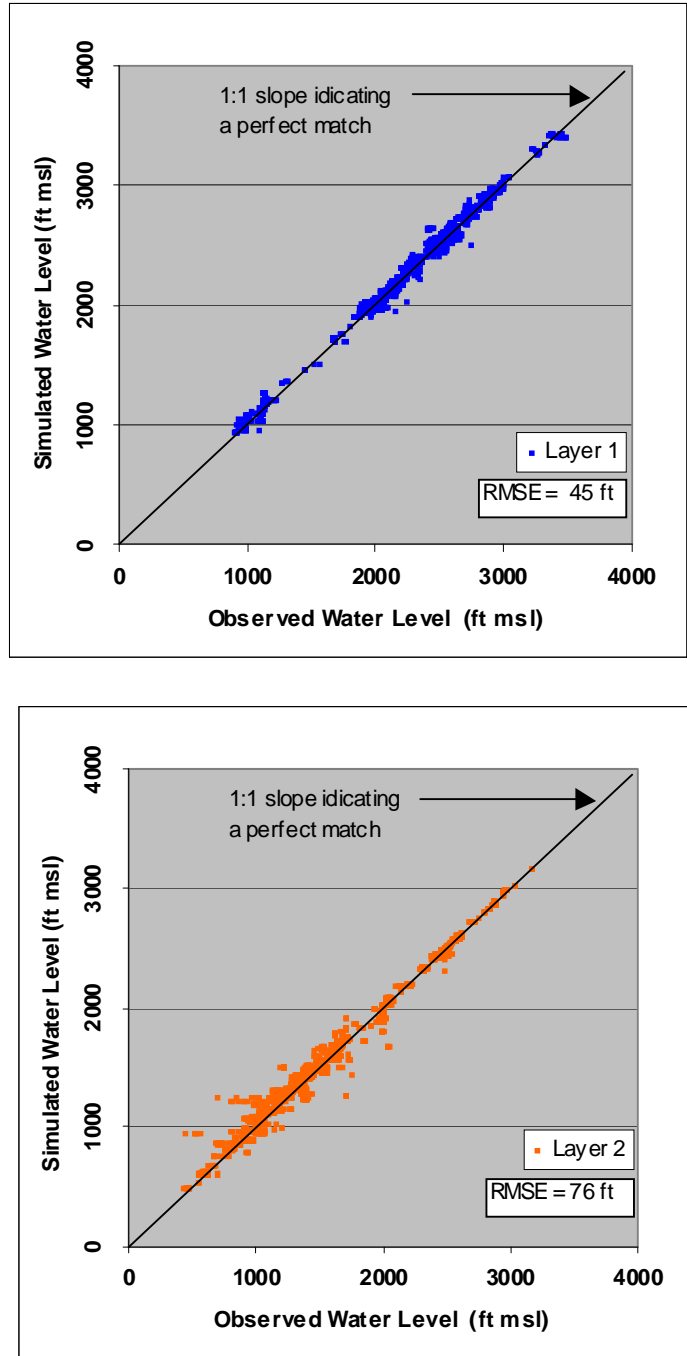


Figure 6-25. Comparison of observed versus simulated water levels for layers 1 and 2 from the PEST recalibration of the ETPV GAM

TWDB Report ##: Final – Application of PEST to Re-Calibrate
the Groundwater Availability Model for the Edwards-Trinity (Plateau) and Pecos Valley Aquifers

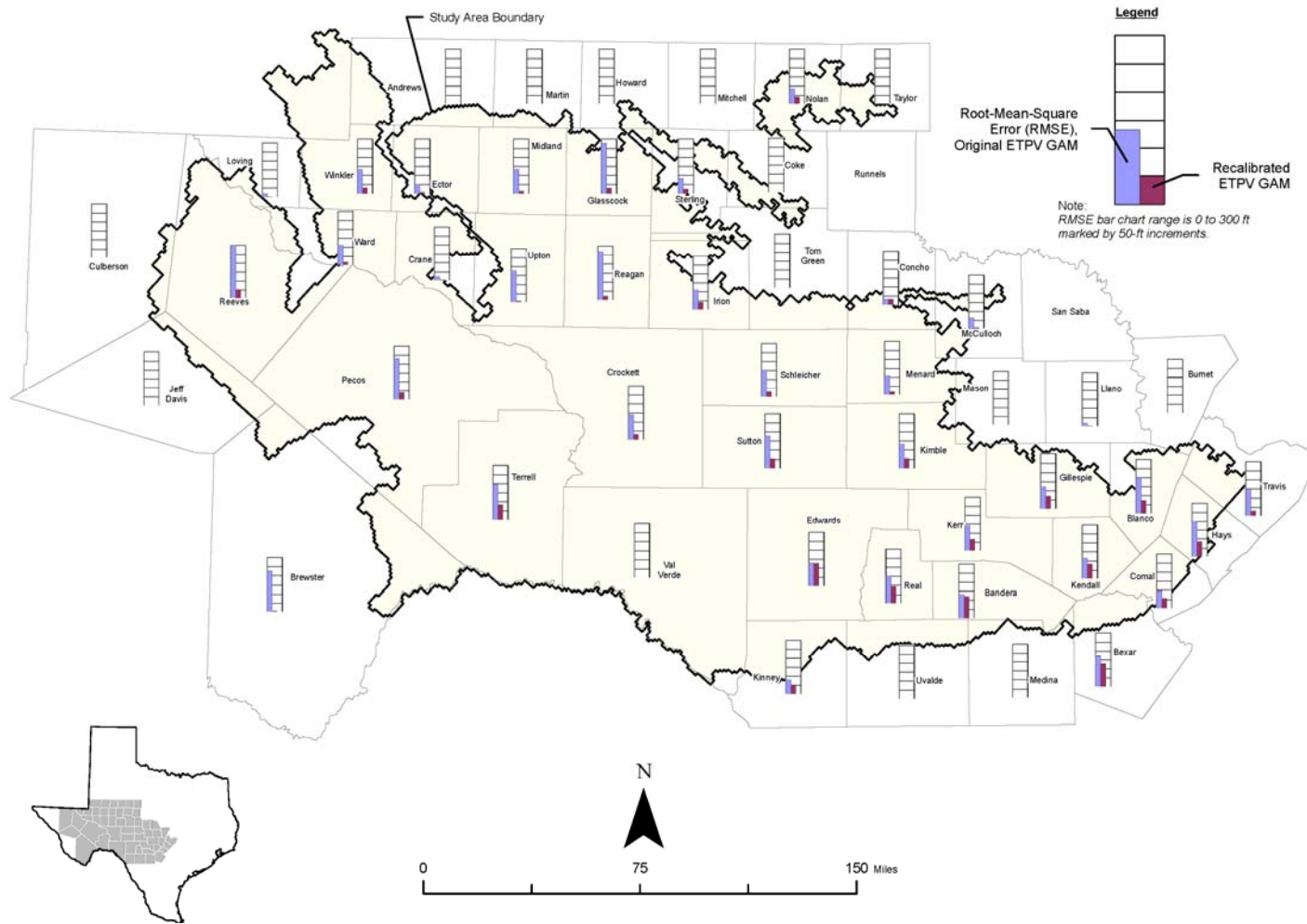


Figure 6-26. Comparison of RMSE calculated for the wells in each county for the original and the recalibrated model

TWDB Report ##: Final – Application of PEST to Re-Calibrate
the Groundwater Availability Model for the Edwards-Trinity (Plateau) and Pecos Valley Aquifers

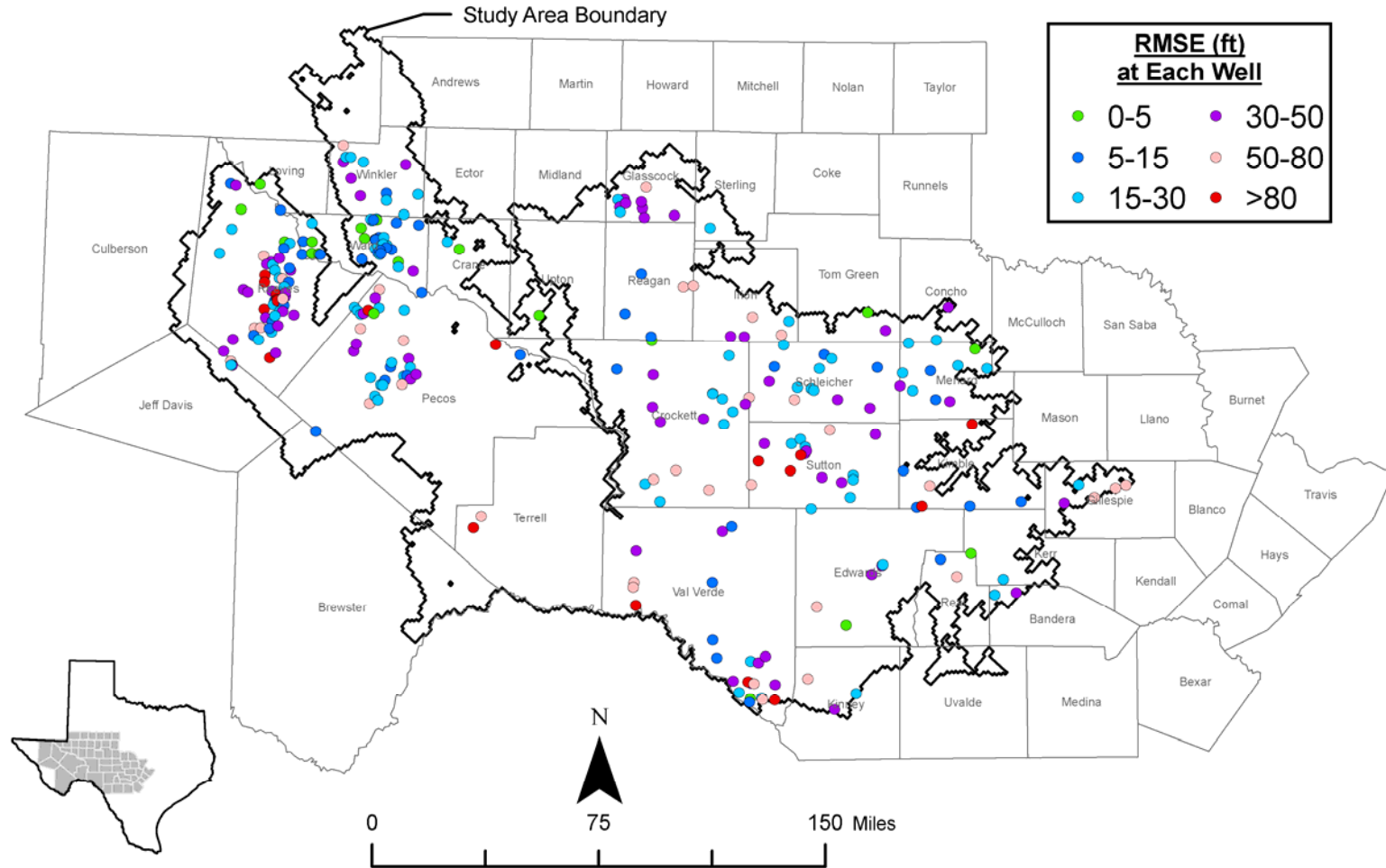


Figure 6-27. Calculated RMSE (ft) at each well location in layer 1 for the recalibrated model

TWDB Report #: Final – Application of PEST to Re-Calibrate
the Groundwater Availability Model for the Edwards-Trinity (Plateau) and Pecos Valley Aquifers

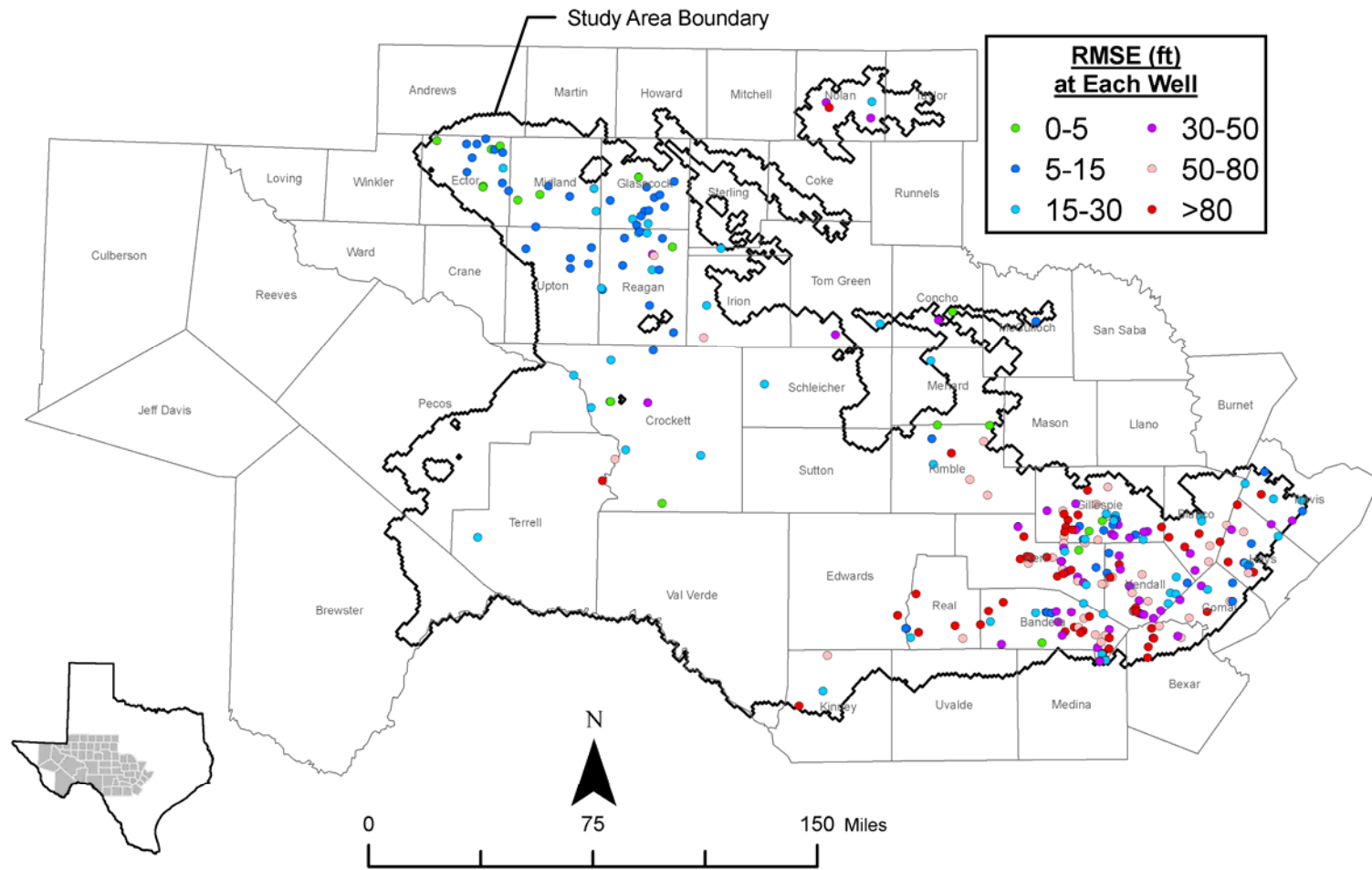


Figure 6-28. Calculated RMSE (ft) at each well location in layer 2 for the recalibrated model

TWDB Report ##: Final – Application of PEST to Re-Calibrate
the Groundwater Availability Model for the Edwards-Trinity (Plateau) and Pecos Valley Aquifers

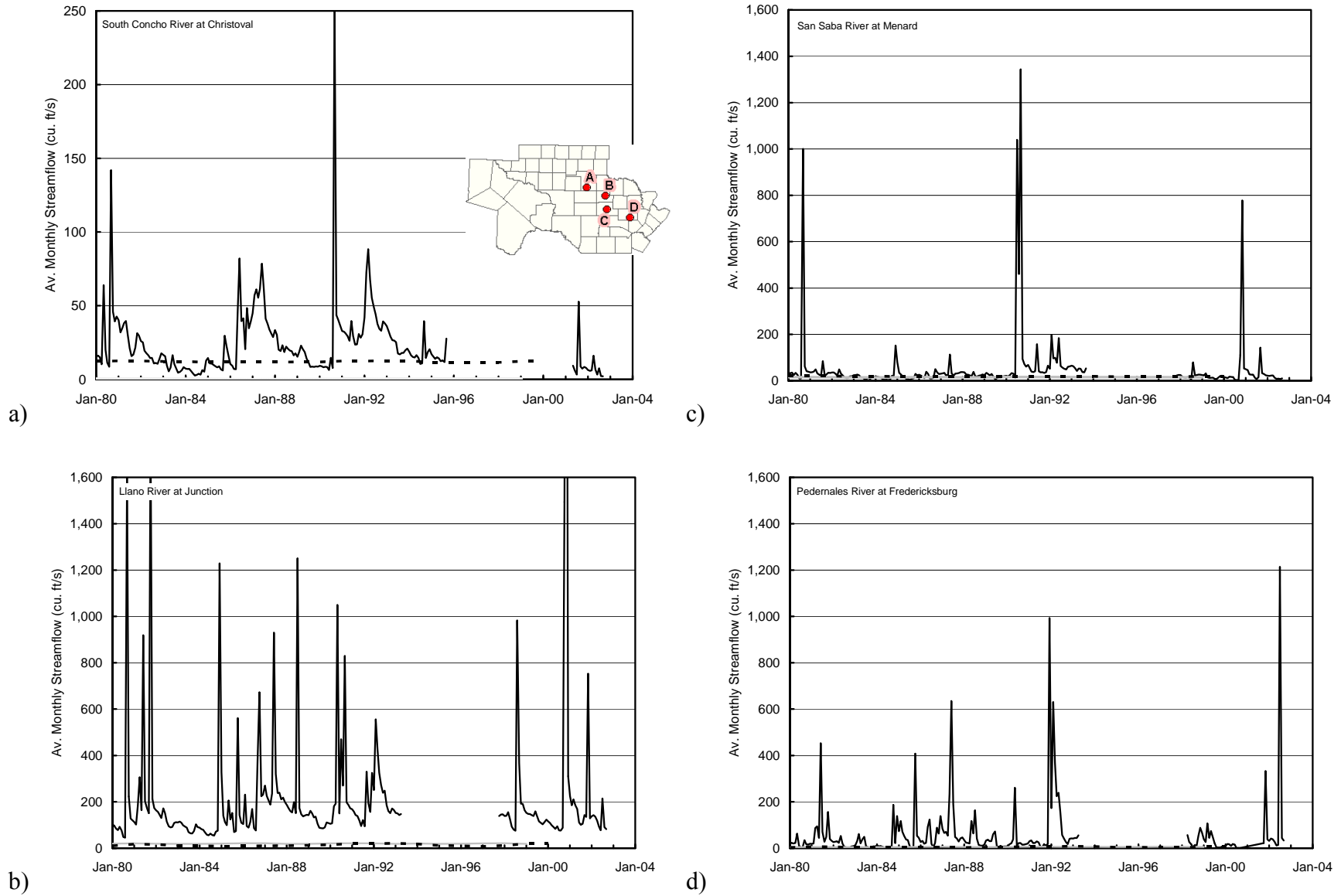


Figure 6-29. Comparison of groundwater contribution to stream flow from the original ETPV GAM (dashed black line) and the recalibrated ETPV GAM (solid grey line) and the river flow estimated from the cited river gage.

TWDB Report ##: Final – Application of PEST to Re-Calibrate
the Groundwater Availability Model for the Edwards-Trinity (Plateau) and Pecos Valley Aquifers

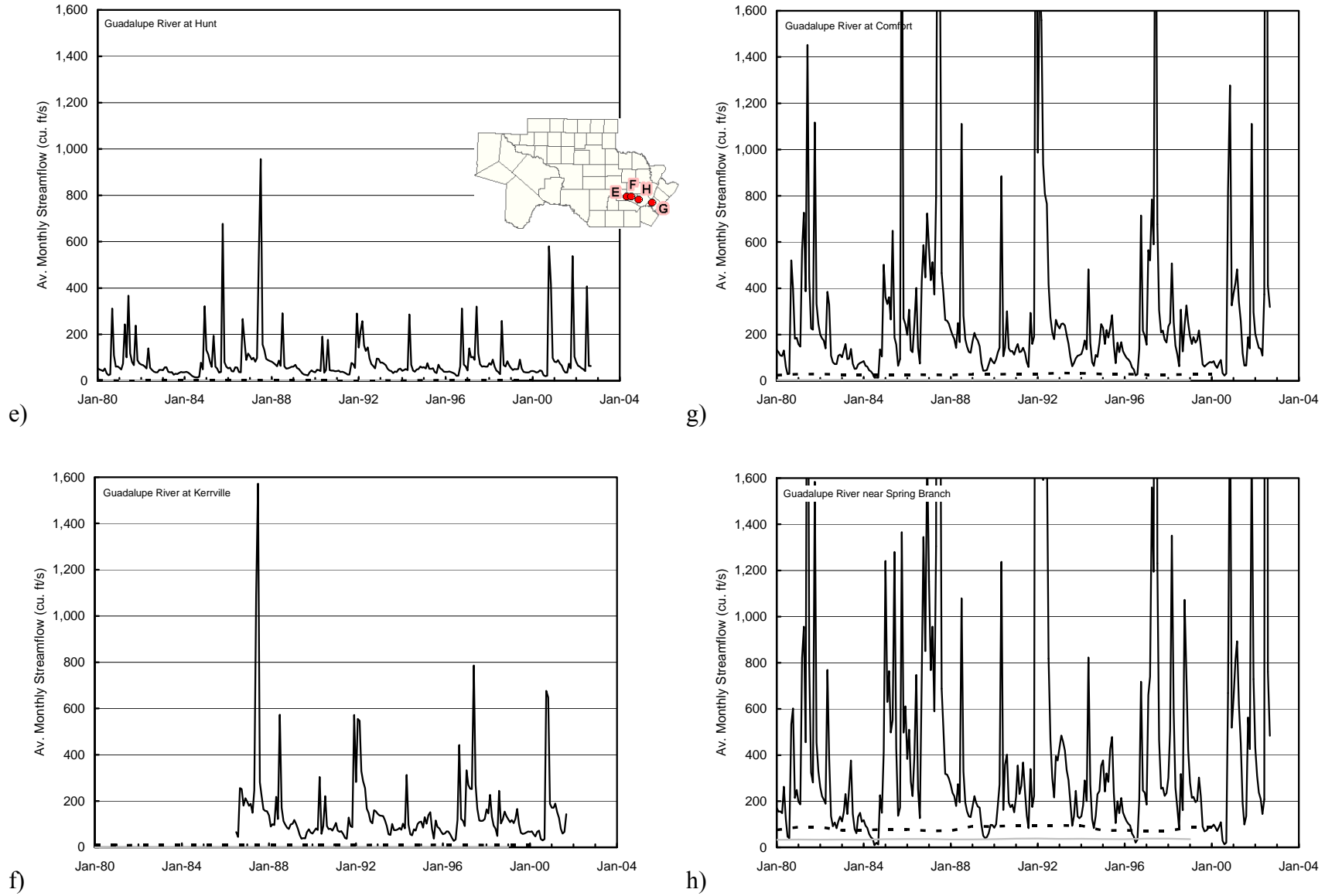


Figure 6-29. (continued).

TWDB Report ##: Final – Application of PEST to Re-Calibrate
the Groundwater Availability Model for the Edwards-Trinity (Plateau) and Pecos Valley Aquifers

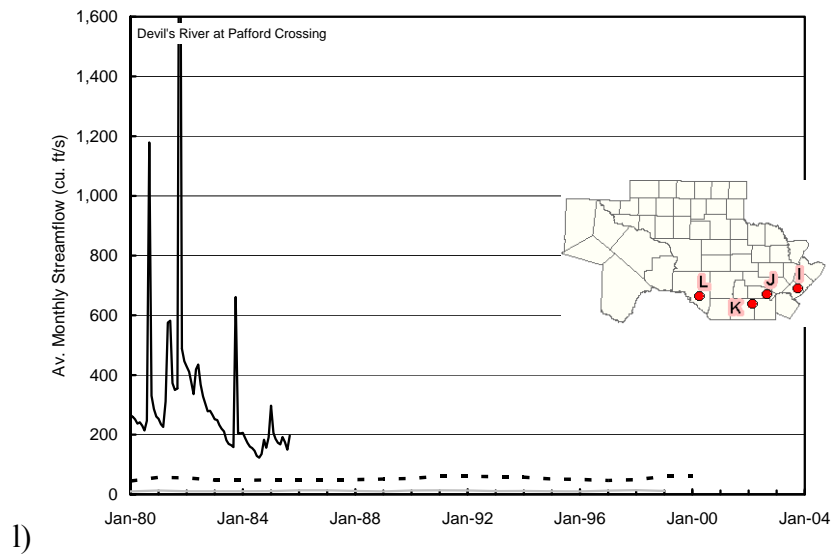
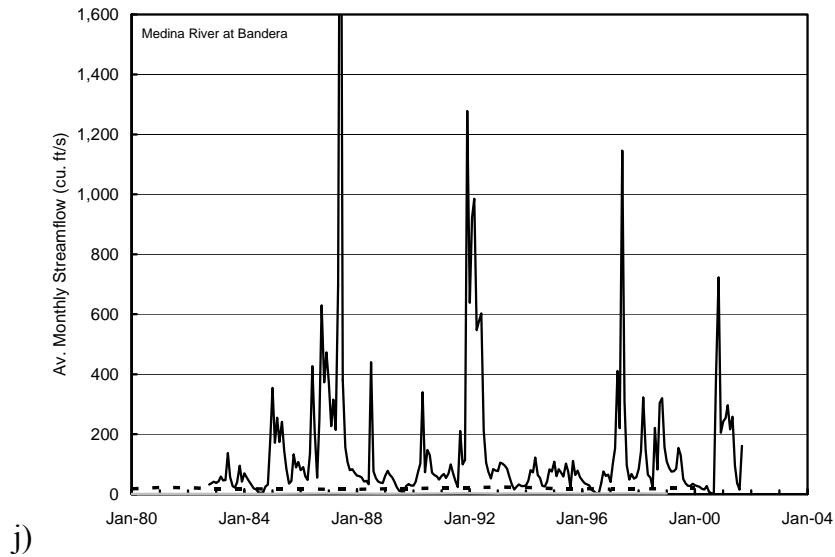
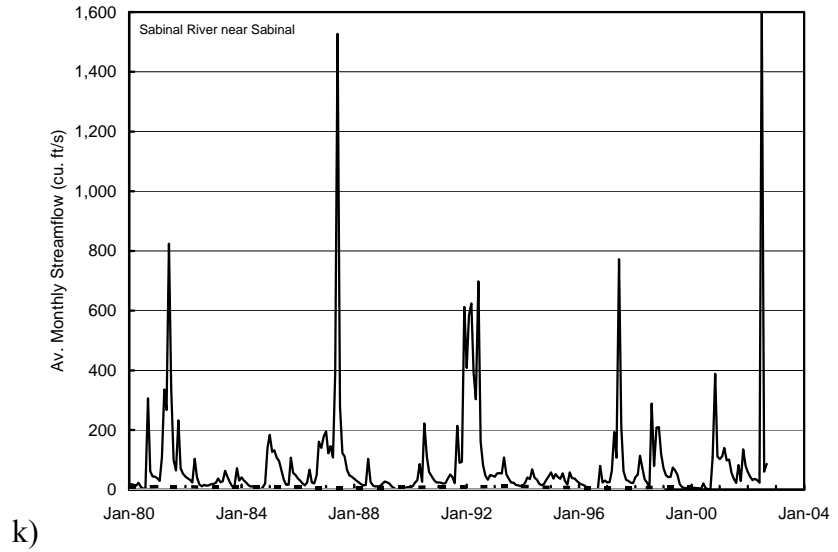
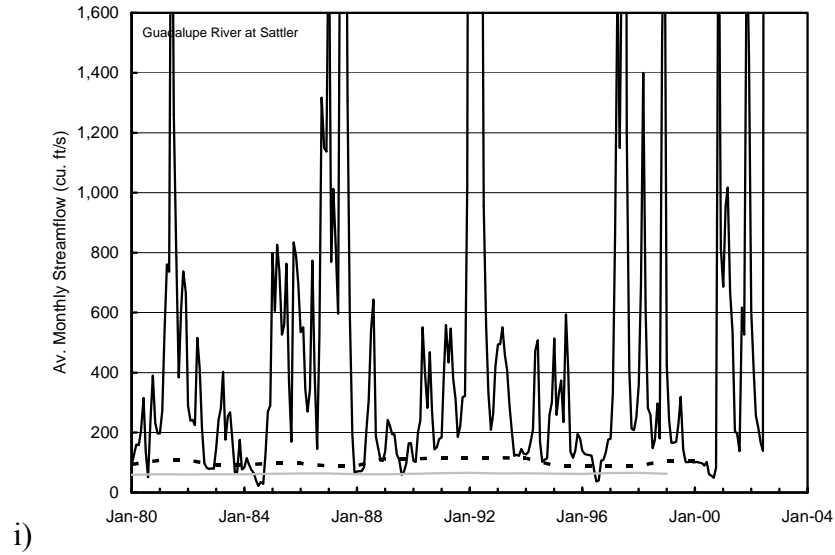


Figure 6-29. (continued).

TWDB Report ##: Final – Application of PEST to Re-Calibrate
the Groundwater Availability Model for the Edwards-Trinity (Plateau) and Pecos Valley Aquifers

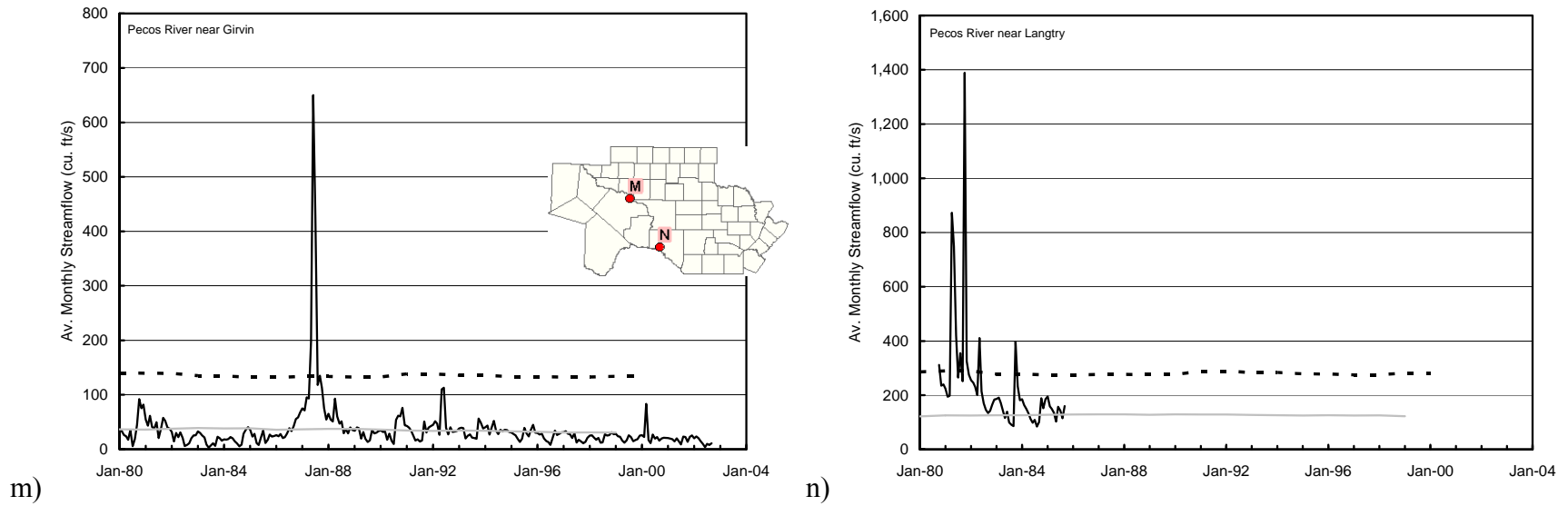


Figure 6-29. (continued).

TWDB Report ##: Final – Application of PEST to Re-Calibrate
the Groundwater Availability Model for the Edwards-Trinity (Plateau) and Pecos Valley Aquifers

7 Limitations of the model

A model can be defined as a representation of reality that attempts to explain the behavior of some aspect of it, but a model is always less complex than the real system it represents (Domenico, 1972). As a result, models have intrinsic limitations. These limitations can be grouped into several categories, including: 1) limitations in the data supporting the model; 2) limitations in the implementation of a model that may include assumptions inherent in the model application; and 3) limitations regarding model applicability (Kelley and other, 2004).

7.1 Data and assumptions supporting the model

As part of this study, no additional data analysis was performed to better characterize the groundwater systems. Therefore, all limitations discussed by Anaya and Jones (2009) regarding the original ETPV GAM equally apply to the recalibrated GAM. These limitations are significant and therefore any user of the recalibrated ETPV should review them carefully. Many of these limitations are associated with the general lack of data available for developing a conceptual framework and for constraining the model calibration.

As a point of emphasis, it should be noted that the recalibrated ETPV GAM's reliability in predicting future conditions is a combination of how well it represents reality and of how well it reproduces historical conditions. Thus, although the ability of the recalibrated GAM to match historical well data is significantly improved over the original ETPV GAM, a final judgment on the overall improvement in the model cannot be accurately assessed until additional field data have confirmed the key assumptions used to develop the model.

7.2 Implementation of model

The USGS MODFLOW96 (Harbaugh and McDonald, 1996) code is based on equations that sometimes fail to properly simulate groundwater flow in unconfined aquifers involving rewetting and desaturation of grid cells. Because of this limitation, Anaya and Jones (2009) modeled the ETPV by fixing the transmissivity values in the model so they do not change with changes in the water levels. This assumption was kept in the recalibration because, as of the writing of this report, there is no free version of MODFLOW that has addressed this limitation of MODFLOW96. As a result of this problem with MODFLOW96, both the original and recalibrated GAM could provide predictions that are physically impossible, such as water levels beneath the aquifer's base. In situations, however, where the drawdown associated with groundwater pumping remains small relative to the total saturated thickness of the aquifer, the approximations in the original and recalibrated ETPV are not likely to adversely affect water level predictions.

One of the factors that affects the ability of PEST to optimize the model parameters is the convergence and stability characteristics of MODFLOW96 over the range of parameter values to be investigated. With respect to these characteristics, MODFLOW96 performed well and PEST was therefore able to optimize the parameter values efficiently. However, during the course of exploring different parameter combinations, a problem with the MODFLOW simulations was identified that is unrelated to numerical convergence or stability. This problem occurred from

time to time when the PEST variable “vertaniso” (a global multiplier used to calculate vertical conductance) increased above 0.0001 and the ensuing MODFLOW simulation produced a solution with unacceptably high mass-balance errors in a few grid cells. Several options were investigated to identify and correct this problem including alternative matrix solvers and removal of multiple hydraulic boundary conditions in the same grid cell. The only option that was found to avoid this problem was to restrict the PEST variable “vertaniso” to values below 0.00001. As a result, the solution space available for calibrating the model was partially limited by the constraints placed on vertical conductances.

A limitation of the PEST application is a general lack of information available to constrain the solution and to minimize the non-uniqueness of the solution. As a result, even though the recalibrated ETPV model is likely a better predictor than the original ETPV model, there is the possibility that another PEST run with different model constraints could still produce a significantly better predictor of changes in water levels caused by pumping. Such a model could be produced because it could contain a different set of recharge values or aquifer parameter values that more closely reflect reality but do not necessarily lead to a noticeably better calibrated model.

7.3 Model applicability

The ETPV GAM was developed using grid cells of one square mile and one grid layer per aquifer. This discretization is designed to address regional water level changes at the scale of tens of miles. The ETPV GAM is not designed to predict aquifer responses at specific points or other points where groundwater flow is largely controlled by horizontal and vertical hydraulic gradients that occur over distances less than the length of the grid cell or the cell thickness. Another limitation of ETPV GAM applicability is that it does not represent spatial variability in the aquifer parameters at the local scale. This level of detail exists because of both a lack of aquifer property measurements and the water level data, as well as the grid cell size.

The ETPV GAM is most applicable and accurate for assessing regional-scale predictions such as aquifer-wide changes over decades caused by regional water management practices. The accuracy and applicability of the ETPV GAM decreases for more local-scale issues involving specific water projects isolated to a part of a county or to a city. Examples of where local-scale groundwater flow is important and therefore may not be adequately represented by the ETPV GAM include simulating the water levels near pumping wells, the groundwater contributions to streams, and changes in water quality over time.

Despite its limitation to address local-scale issues, the ETPV GAM provides a useful model from which a more detailed model can be developed to address local-scale issues involving wells, stream flows, and water quality. To create such a detailed model, a section of the ETPV GAM can be further refined by inserting additional grid cells and can be further calibrated by including additional aquifer information in an extended model calibration.

7.4 Recommended model improvements

All of the future model improvements discussed by Anaya and Jones (2009) regarding the original ETPV GAM equally apply to the recalibrated ETPV GAM. Among the key points in their discussion are identifying the benefits associated with collecting additional field data to support the model development, subdividing the ETPV GAM into two or more smaller regional models, and using a modeling approach that allows transmissivity to vary with saturated thickness.

During the project the authors investigated the options for developing a modeling approach that would allow transmissivity to vary with saturated thickness. The most promising approach involves a version of the USGS MODFLOW that should be released in late 2009. This MODFLOW version will include numerically enhanced options for predicting water levels in unconfined aquifers. These options are similar to those offered by MODFLOW-SURFACT (HGL, 2006), which is a commercially available version of MODFLOW. In a trial run, the authors successfully used MODFLOW-SURFACT to develop a 1930 steady-state solution that allows transmissivity to vary with saturated thickness with parameters developed from one of the PEST optimization runs.

Anaya and Jones (2009) lists numerous assumptions and approximations that had to be made to construct a hydrogeological framework for developing the original ETPV GAM. The numerous assumptions and approximations are needed because of the large model domain and the lack of information regarding the spatial variability in the aquifer properties. To expedite improvements to future models for the ETPV aquifer study area, we recommend partitioning the model into smaller regional models so that field work and analysis can be more effectively targeted to address specific areas of interest. As noted by Anaya and Jones (2009), the domain for the smaller regional models should be based on natural boundary conditions, such as major streams and groundwater divides.

One of the priorities in characterizing the aquifers is to identify key relationships within each aquifer that affect the spatial variability of the aquifer parameters. Such analysis would likely include a detailed analysis of geophysical logs and pumping test results to better understand the lithologic and stratigraphic variations across an aquifer and how these variations impact values for hydraulic conductivity, vertical conductance, and storage properties.

As the modeling sophistication with the ETPV GAM advances, it appears prudent to pursue an expanded use of PEST for both model development and model application. Among the benefits that PEST can provide over the previous practices of trial-and-error approaches are the following:

1. The ability to obtain better fits between model outputs and historical measurements than would otherwise be possible;
2. Through its regularization capabilities, an ability to better incorporate geological knowledge and expertise into the calibration process than would otherwise be possible;

3. Rapid turnaround on conceptual model development and refinement during early stages of the calibration process where a modeler uses local and global misfits to indicate areas and types of model inadequacies; and
4. Formulation of a base for parameter and predictive uncertainty analysis as an aid to collective model-based decision-making and in formulating strategies for future data acquisition.

Together, these benefits lead to model predictions that are likely to approach those of minimum error variance along with a capacity to quantify that variance. Variance reduction through optimization of future data acquisition can also be enhanced.

In summary, compared with the previous practices of trial-and-error model calibration approaches, PEST provides a significantly improved approach for determining optimal parameter values, investigating alternative data interpretations, and performing predictive uncertainty. Over the long-term development of a model, which may involve several groups of modelers and several updates of data, PEST has the potential to improve how GAMs are calibrated and applied.

8 Summary and conclusions

The primary focus of this report is to evaluate the potential for PEST and HPC clusters to benefit the GAM program. PEST is a computer program that provides the capability of semi-automatically calibrating groundwater models and integrating new information into the calibration process in a more robust and cost-effective manner than traditional manual approaches allow. An HPC cluster permits a single PEST application to run across a network of linked computers so that calibration run times can be greatly reduced.

The TWDB selected the ETPV GAM as the candidate for this demonstration project. Some of the reasons for selecting this GAM are that it is mathematically stable and the area around Upton and Reagan Counties would benefit greatly by additional model parameter adjustments to better simulate historical water levels. One of the objectives of the PEST application is to develop an improved ETPV GAM, particularly across Upton and Reagan Counties.

The transient calibration period for the original ETPV GAM is from 1980 to 2000. The PEST recalibration of the ETPV GAM is based on the same conceptual model and data presented by Anaya and Jones (2009) with one change. This change involves how the initial water levels are calculated for the transient simulation. The original ETPV GAM developed the initial water levels by running a steady-state simulation using 1980 pumping rates. The recalibrated ETPV GAM developed the initial water levels by running a transient model from 1930 to 1980. The pumping rates for the 1930-to-1980 transient model were developed using historical values for irrigation acreages and application rates in combination with PEST's parameter estimation capabilities.

In addition to the incorporation of PEST, the authors investigated several options for improving the value of the recalibration process. This investigation promulgated an increase in the number of hydrographs with more than three water level measurements from 10 to over 500. A benefit of including more hydrographs in the calibration is that the overall reliability of the model is improved and there is a greater likelihood of identifying where uncertainty or inaccuracies with the model predictions may be the greatest. Because of its robust input file formats, the inclusion of numerous hydrographs into a PEST calibration run incurs no logistical, numerical, or computational penalties.

A primary metric for evaluating the calibration of the ETPV GAM is the RMSE between the observed and simulated water levels. The difference between an observed and a simulated water level is called a residual. Analysis of all residuals indicates that the recalibrated GAM provides a significantly better fit to the 4,773 historical water level from 574 hydrographs than does the original ETPV GAM. For the entire historical water level dataset, the original GAM produces a RMSE of about 180 ft; the recalibrated GAM has a RMSE of about 60 feet. Across the entire model, the recalibrated GAM reduced the RMSE by 122 ft, the mean error by 66 ft, and the absolute error by 106 feet. From the perspective of how the recalibrated GAM improves the matches to historical water levels, the RMSE was reduced by more than a factor of 2 and 3 in 26 and 19 counties, respectively, and by more than 50 ft and 100 ft in 24 and 15 counties, respectively.

At the onset of the project, a specific goal of the project was to improve the model calibration for Upton and Reagan counties. The RMSE for Upton and Regan Counties produced by the original ETPV GAM is 176 ft and 269 feet, respectively. The RMSE for Upton and Regan Counties produced by the recalibrated ETPV GAM is 9 ft and 23 feet, respectively.

In regard to producing a model that is better at matching historical water level measurements and having a distribution of aquifer parameters consistent with the field data, the recalibration of the ETPV GAM using PEST has provided a significantly improved model compared to the original ETPV GAM.

Over the 20-year period, the recalibrated ETPV GAM averages about 1.1 million AFY of recharge compared to about 1.2 million AFY of recharge in the original ETPV GAM. In terms of the average total flow through the aquifer, the recalibrated ETPV GAM averages about 1.8 million AFY compared to 1.7 million AFY for the original ETPV GAM. For the recalibrated GAM, the sources of groundwater include recharge (62%), storage (26%), rivers (8%), reservoirs (3%), and adjacent aquifers (1%). The mechanisms for groundwater discharge in the recalibrated GAM include rivers (27%), storage (25%), wells (19%), adjacent aquifers (14%), springs/seeps (11%), and reservoirs (4%).

In the course of optimizing parameters for a model calibration, PEST needed to run the model many times as part of the process of calculating the Jacobean matrix, i.e., the matrix of derivatives of observations with respect to parameters. PEST updates the model parameters for the model calibration after the Jacobean matrix has been calculated. Between 120 hours and 250 hours of computing time are required to calculate the Jacobean. To expedite the PEST model calibration, the PEST calibration was run on a HPC cluster of six computers with duo core processors. The use of Parallel PEST across this HPC cluster reduced the total run time by a factor of about 12.

The HPC cluster operated smoothly throughout the project with minimal user intervention. The only user invention that occurred was the rebooting of the HPC following occasional power outages and the extraction of PEST output files to monitor the progress of a model calibration run. With the HPC cluster, a PEST run typically completed a model calibration in about 5 days, which is the time required to complete five to seven PEST optimization iterations.

All of the future model improvements discussed by Anaya and Jones (2009) regarding the original ETPV GAM equally apply to the recalibrated ETPV GAM. Among their key points are identifying benefits associated with collecting additional field data to support the model development, subdividing the ETPV GAM into two or more smaller regional models, and using a modeling approach that allows transmissivity to vary with saturated thickness.

Compared to the traditional practices of calibration models involving manual trial-and-error approaches, PEST provides a significantly improved methodology for determining optimal parameter values, investigating alternative data interpretations, and performing predictive uncertainty. Over the long-term development of a model, which may involve several groups of modelers and several updates of data, the PEST has the potential to improve how GAMs are calibrated and applied

9 References

- Anaya R., and Jones, I., 2009. Groundwater Availability Model for the Edwards-Trinity (Plateau) and Pecos Valley Aquifers of Texas. Report 373. Texas Water Development Board, Austin, TX, 103 pp.
- Ashworth, J.B., and Hopkins, J., 1995. Aquifers of Texas: Texas Water Development Board Report 345, 69 p.
- Ashworth J.B., 1990. Evaluation of ground-water resources in parts of Loving, Pecos, Reeves, Ward, and Winkler Counties, Texas: Texas Water Development Board Report 317, 51 p.
- Aster, R.C., Borchers, B., and Thurber, C.H., 2005. Parameter Estimation and Inverse Problems. Elsevier Academic Press, Amsterdam, 301pp.
- Beven, K., 2006. A manifesto for the equifinality thesis. *Journal of Hydrology* 320, 18-36.
- Certes, C. and de Marsily, G., 1991. Application of the pilot-points method to the identification of aquifer transmissivities. *Adv. Water Resour.*, 14(5), 284-300.
- Christensen, S. and Doherty, J., 2008. Predictive error dependencies when using pilot points and singular value decomposition in groundwater model calibration. *Advances in Water Resources*. Volume 31, Issue 4, April, pages 674-700.
- Clemo, T., 2007. User Guide to the Adjoint State Based Sensitivity Process (ADJ). Technical Report BSU CGISS 07-01. Center for the Geophysical Investigation of the Shallow Subsurface. Boise State University.
- Dausman, A.M., Doherty, J., Langevin, C.D., and Sukop, M.C., 2009. Quantifying data worth toward reducing predictive uncertainty. Submitted to *Groundwater*.
- De Marsily, G., Lavedan, C., Boucher, M., and Fasanino, G., 1984. Interpretation of interference tests in a well field using geostatistical techniques to fit the permeability distribution in a reservoir model. in *Geostatistics for Natural Resources Characterization*, edited by G. Verly, M. David, A.G. Journel and A. Marechal. NATO ASI, Ser. C. 182. 831-849
- Doherty, J., 2003. Ground-water model calibration using Pilot-points and Regularization. *Ground Water*. Vol 41 (2): 170-177
- Doherty, J., 2009a. PEST: Model-Independent Parameter Estimation User Manual: 5th Edition, Watermark Numerical Computing, Corinda Australia
- Doherty, J., 2009b. Addendum to the PEST manual. Watermark Numerical Computing, Corinda, Australia.

- Doherty, J., Fienen, M.N. and Hunt, R.J., 2009c. Pilot points as a parameterization device; theory guidelines and research directions. USGS Scientific Investigation Report, in prep.
- Domenico, P. A., 1972. Concepts and Models in Groundwater Hydrology, McCraw-Hill, New York.
- Duan, Q., Sorooshian, S. and Gupta, V., 1992. Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resources Research* 28 (4), 1015-1031.
- Duan, Q., Gupta, V.K. and Sorooshian, S., 1993. A Shuffled Complex Evolution approach for effective and efficient global minimization. *Journal of Optimization Theory and its Applications*, 76 (3), 501-521.
- Duan, Q., Sorooshian, S. and Gupta, V.K., 1994. Optimal use of the SCE-UA global optimization method for calibrating watershed models. *Journal of Hydrology*, 158 265-284.
- Gallagher, M. and Doherty, J., 2007. Predictive error analysis for a water resource management model. *Journal of Hydrology*, 34(3-4), 513-533.
- Hansen, N. and Ostermeier, A., 2001. Completely derandomized self-adaptation in evolution strategies. *Evol. Comput.*, 9, 159-195.
- Hansen, N., Muller, S.D., and Koumoutsakos, P., 2003. Reducing the time complexity of the derandomized evolution strategy with covariance matrix adaptation (CMA-ES). *Evol. Comput*, 9, 159-195.
- Harbaugh A. W., and McDonald, M. G. 1996. User's documentation of MODFLOW-96, an update to the U. S. Geological Survey modular finite-difference ground-water flow model: U.S. Geological Survey Open-File Report 96-485, 56 p.
- HGL, 2006. MODFLOW-SURFACT SOFTWARE (Version 3.0), Overview: Installation, Registration, and Running Procedures. HydroGeologic, Inc., Herndon, VA.
- Hill, M.C., 1998. Methods and Guidelines for Effective Model Calibration. U.S. Geological Survey, Water Resources Investigations Report 98-4005.
- Hunt, R.J., Doherty, J, and Tonkin, M.J., 2007. Are models too simple? Arguments for increased parameterization. *Ground Water* 45 (3), 254–262.
- Jiang, Y. and Woodbury, A.D., 2006. A full-Bayesian approach to the inverse problem for steady-state groundwater flow and heat transport. *Geophys. J. Int.* 167, 1501-1512.
- Kelley, V., Deeds, N., Fryar, D., and Nicot, JP., 2004. Groundwater Availability Models for the Queen City and Sparta Aquifers., prepared for the Texas Water Development Board, INTERA Incorporated, Austin, TX.

- Kitanidis, P. K., 1997. The minimum-structure solution to the inverse problem, *Water Resour. Res.*, 33(10), 2263-2272, 1997.
- Koch, K.R., 1988. *Parameter estimation and hypothesis testing in linear models* Springer-Verlag, Berlin. XVI+378 pp.
- LaVenue, A.M. and Pickens, J.F., 1992. Application of a coupled adjoint sensitivity and kriging approach to calibrate a ground-water flow model. *Water Resour. Res.*, 28(6), 1543-1569.
- McDonald, M.G., and Harbaugh, A.W., 1988. A modular three-dimensional finite-difference groundwater flow model: U.S. Geological Survey Techniques of Water-Resources Investigations Book 6: Model Techniques, Chapter A1.
- Moore, C. and Doherty, J., 2005. The role of the calibration process in reducing model predictive error. *Water Resour. Res.* Volume 41, No 5. W05050.
- Moore, C. and Doherty, J., 2006. The cost of uniqueness in groundwater model calibration. *Advances in Water Resources.* Volume 29, Issue 4, April, pages 605 – 623.
- Paige, C.C. and Saunders, M.A., 1982a. LSQR: an algorithm for sparse linear equations and sparse least squares. *ACM Trans. Math Softw*, vol8: 43-71.
- Paige, C.C. and Saunders, M.A. 1982b. LSQR: an algorithm for spare linear equations and sparse least squares. *ACM Trans. Math. Softw.* Vol 8: 192-209.
- Song, Z., Li, L., Nielsen, P. and Lockington, D., 2006. Quantification of tidal watertable over-height in a coastal unconfined aquifer. *Journal of Engineering Mathematics.* 56, 4, 437-444.
- Tonkin, M. and Doherty, J., 2005. A hybrid regularized inversion methodology for highly parameterized models. *Water Resour. Res.* Vol. 41, W10412, doi:10.1029/2005WR003995.
- Tonkin M., and Doherty, J., 2009. Calibration-constrained Monte Carlo analysis of highly parameterized models using subspace techniques, *Water Resour. Res.*, 45, W00B10, doi:10.1029/2007WR006678.
- Tonkin, M.J., Tiedeman, C. R., Ely, D.M., and Hill, M.C., 2007. OPR-PPR, a Computer Program for Assessing Data Importance to Model Predictions Using Linear Statistics: Reston Virginia, U.S. Geological Survey Techniques and Methods Report TM-6E2, 115 pages.
- Young, S. C., Kelley, V., Budge, T., Deeds, N., and Knox, P. 2009. Development of the LCRB Groundwater Flow Model for the Chicot and Evanageline Aquifers in Colorado, Wharton, and Matagorda Counties, prepared for Lower Colorado River Authority, Austin, TX.
- Young, S. C., and Kelley, V., editors. 2006. A Site Conceptual Model to Support the Development of a Detailed Groundwater Model for Colorado, Wharton, and Matagorda Counties, prepared for the Lower Colorado River Authority, Austin, Texas.

TWDB Report ##: Final – Application of PEST to Re-Calibrate
the Groundwater Availability Model for the Edwards-Trinity (Plateau) and Pecos Valley Aquifers

Appendix A Introduction to PEST

A.1 What is PEST?

A.1.1 General

PEST is the name given to a suite of programs that collectively undertakes calibration and uncertainty analysis for environmental and other numerical models. The motivation for its original development was to provide model calibration functionality in a model-independent manner, whereby it could interact with a model through the latter's own input and output files, thereby promulgating its use with any model without the need for recompilation of either PEST or the model. Since the time of PEST's original release in 1994, at least three other model-independent parameter estimation packages have become available. These are:

- OSTRICH developed by Shawn Matott of University of Waterloo, Ontario, Canada;
- AUTOCAL developed by Henrik Madsen of DHI, Denmark; and
- UCODE developed by Eileen Poeter and Mary Hill of USGS.

To the authors' knowledge, however, PEST retains many features that collectively make it unique among model-independent parameter estimation packages. These features include:

- PEST and its ancillary software allow parameter estimation to be undertaken interchangeably using a number of different methods, including both gradient-based methods and so-called global methods. Currently supported global methods are Shuffled Complex Evolution (SCE) (Duan et al., 1992; 1993; 1994) and Covariance Matrix Adaptation Evolution Strategy (CMA-ES) (Hansen and Ostermeier, 2001; Hansen et al., 2003).
- PEST is able to undertake parameter estimation in both over-determined and undetermined calibration contexts. In the latter context, mathematical regularization can be implemented using Tikhonov and/or subspace methods (including singular value decomposition (SVD) and least-squares using QR factorization (LSQR)) plus hybrids of these. It is also able to implement the highly efficient and unique "Singular Value Decomposition-assist (SVD-assist)" methodology that allows highly parameterized inversion to be undertaken with runtime efficiencies normally associated with parsimonious calibration.
- In addition to calibration, PEST can undertake calibration-constrained parameter and predictive uncertainty analyses. Both linear and nonlinear options are available. The latter includes the highly efficient (and unique) "null space Monte Carlo" method whereby many different parameter sets can be computed, all of which are realistic and all of which calibrate a model.
- Model runs can be parallelized across PC networks or Linux clusters. Third-party PEST developers have expanded these capabilities to include a variety of communications protocols between PEST and its model run supervisors.
- PEST is supported by a plethora of utility programs that expedite many data manipulation tasks associated with parameter estimation and uncertainty analyses. In addition, a large

suite of utilities is available to expedite the use of PEST in groundwater and surface water modeling contexts, including interfaces to popular groundwater models such as MODFLOW, MT3D, SEAWAT, and FEFLOW. A comprehensive time-series processor is available for use in the surface water modeling environment; this facilitates the handling of large datasets and implements automatic generation of PEST input datasets involving complex, multi-component objective functions.

- Comprehensive documentation is provided explaining all aspects of all algorithms employed by PEST and its ancillary software. Extensive training material is also available for those who attend PEST classes.
- Source code for PEST and all of its utilities is freely available, as are compilation instructions.
- PEST is supported by a number of popular groundwater graphical user interfaces, including Groundwater Vistas, GMS, Visual MODFLOW, and PMWIN. The level of supported PEST functionality varies between these interfaces. At the time of this writing, the level of support is in the order just listed (with the first providing the most support).

PEST and all of its utility software are free. PEST can be downloaded from:

<http://www.pesthomepage.org>

A.1.2 PEST utility software

When a user downloads and installs PEST, over 100 executable programs are placed in the PEST installation directory. These programs perform a variety of tasks associated with the use of PEST, including:

- Checking the integrity of part or all of a PEST input dataset;
- Pre- and post-processing for classical parameter estimation;
- Pre- and post-processing for regularized inversion;
- Observation weights and covariance matrix manipulation;
- Linear and nonlinear uncertainty and error analyses;
- Jacobean and sensitivity data manipulation;
- Matrix manipulation; and
- Global optimization.

A complete list of utility software supplied with PEST is provided in Appendix C.

Another suite of utility programs, many of which expedite the use of PEST in the groundwater modeling context (particularly in conjunction with the MODFLOW, MT3DMS, and SEAWAT groundwater models), can be downloaded with PEST. These are referred to as the “Groundwater Data Utilities”. (The word “modeling” would have been preferred instead of “data” in naming these utilities; however, at the time of their original release, another unrelated set of “groundwater modeling utilities” was available.) Not all of these utilities were written to expedite the use of PEST; some are useful in their own right as general modeling support software. At the time of writing, 97 Groundwater Data Utilities were available. Tasks performed by these utilities include:

- Implementation of pilot points parameterization for a variety of groundwater models;
- Manipulation of model (particular MODFLOW) two-dimensional arrays;
- MODFLOW/MT3D/SEAWAT pre- and post-processing;
- Processing and manipulation of time series;
- Automated construction of PEST input datasets;
- Construction of regularization constraints in complex modeling environments;
- Facilitation of use of the MODFLOW adjoint process;
- Stochastic field generation and uncertainty analysis; and
- Geographical data manipulation.

See Appendix C for a complete listing of programs belonging to the Groundwater Data Utility suite.

A third set of utility programs, known as the “PEST Surface Water Utilities”, was written to perform a role in the surface water modeling context similar to that which the Groundwater Data Utilities perform in the groundwater modeling context. However, a slightly different philosophy was adopted in their design in that most functionality resides in a single program named TSPROC. TSPROC is designed for use both as a general model postprocessor and a PEST input dataset generator. Tasks that it can undertake in its former capacity include:

- Interpolation of model outputs to times at which measurements were made;
- Arbitrary mathematical transformation of time series (time series do not need to be sampled at regular intervals);
- Arbitrary mathematical operations between time series;
- High-pass, low-pass, band-pass, and quickflow/baseflow digital filtering of time series;
- Accumulation of volumes over time series between sets of arbitrary dates;
- Computation of statistics for a single time series; and
- Computation of fit-statistics between two different time series.

In its role as a PEST pre-processor, TSPROC can create an entire PEST input dataset based on a multi-component objective function built from observations at one site or many sites, processed in some or all of the above ways.

See Appendix C for a listing of programs belonging to the Surface Water Utility suite.

A.2 Model independence

As stated above, PEST’s “model independence” describes the fact that it is able to communicate with a model through the model’s own input and output files. In the course of the parameter estimation, process a model must be run many times. Prior to each model run, PEST writes the parameter values that it wishes the model to use on that run to the model’s own input files; it is the user’s responsibility to identify the locations on those input files at which the parameters reside. After the model has completed execution, PEST reads from pertinent output files model-generated numbers whose differences with field measurements must be minimized. The user

must indicate to PEST the files from which these numbers must be read and the locations of these numbers on model output files.

The user identifies numbers on model input files that PEST is allowed to vary from model run to model run using templates of these files. An instruction file must be provided corresponding to each model output file that PEST must read, informing PEST of the location of numbers on these files which it must obtain. There is no limit to the number of model input files that PEST can write prior to a model run, nor to the number of output files from which it can read numbers after completion of a model run.

When PEST runs a model, it does so using a SYSTEM command.

A number of repercussions follow from this mode of communication with a model, some of which are positive and some of which are negative. These are listed below.

1. A model must read its parameters from ASCII files. It must also write numbers that are to be matched to field measurements to ASCII files.
2. For execution to be possible using a SYSTEM command, the model must exist as a discrete executable file. If the model requires user input from command-line prompts, these can be supplied in a text file to the model through redirection (i.e., using the “<” symbol on the command line). However, if the model is accessible only through a point-and-click style user interface, PEST access to the model becomes very difficult.
3. A “model” can in fact be a batch file comprised of many executables run in succession. This easily allows simultaneous calibration of multiple models. It also allows a model to be preceded by one or a number of parameter pre-processors and followed by one or a number of model output post-processors.
4. Parallelization of model runs across a network or cluster becomes a simple matter as long as it is possible to read and write to files across the network/cluster and send a signal to the operating system (or a proxy slave program) to run the model as required.

The third of the above features of PEST-model interaction was exploited in designing a calibration scheme for the model that is the subject of the present study. In fact, one of the main benefits of the model-independent approach encapsulated in PEST is the freedom that it provides to allow rapid design of a parameter estimation scheme that is tuned to the demands of the current modeling problem and to the nature of the data at hand. In most contexts, this is readily accomplished by adding appropriate pre/post-processing software to a model and incorporating that software in the overall model batch file as run by PEST.

A.3 Short history of PEST

A.3.1 In the beginning

The first version of PEST was released in 1994 by John Doherty. This initial version supported only traditional parameter estimation based on the Gauss-Marquardt-Levenberg (GML) method. Nevertheless, at the time of its release (and for 6 years afterwards), it was unique in its model-independence capabilities. Though featuring a particularly robust variant of the GML inversion

method, PEST's algorithm was not new. However, prior to the release of PEST, model calibration under software control required a subroutine interface between parameter estimation software and the model. Hence, programming skills and a compiler were required.

PEST originally was a commercial product sold as a stand-alone item. It was publicized through the Scientific Software catalogue, which at the time had a worldwide circulation of over 100,000. At that time, before the widespread use of the internet, this catalogue played a unique role in dissemination of information to the groundwater modeling community.

A.3.2 MODFLOW interfaces

Despite its good exposure in the Scientific Software catalogue, PEST sales were very slow until a set of utilities was written to expedite its use with MODFLOW and MT3D. Not too long after this, both PEST and these utilities were supported by the PMWIN MODFLOW graphical user interface. With the resulting further exposure and the ability for groundwater modelers to point and click through this interface, PEST use started to become widespread. A contributor to this widespread use was the fact that inclusion of PEST with PMWIN entailed no extra cost to the purchaser of PMWIN.

Shortly after this (in early 1996), a PEST interface was included in Visual MODFLOW. Because of the popularity of this graphical user interface, PEST usage became widespread. In 1997, PEST interfaces followed in both Groundwater Vistas and GMS.

A.3.3 Parallelization

Parallel PEST was introduced in 1998. The nature of parameter estimation as undertaken using the GML method is such that during significant phases of the parameter estimation process (in particular, those phases in which Jacobean matrix computation is undertaken), a large number of model runs must be undertaken independently of each other. If these runs must be done in serial, this creates a significant blockage to the rapid progress of the parameter estimation process. Fortunately, this is an eminently parallelizable activity. Because parallelization can be instigated at the model run level, it is relatively easy to implement. In Parallel PEST, parallelization is implemented through the agency of a PEST master program and a series of “slaves” that reside on different machines on a network to which the master has access. Communication between the two is effected through the agency of platform-independent message files. Hence, the same protocol can be employed on a single multi-processor machine, on a series of such machines comprising an office network, or arrays of PCs or nodes running WINDOWS, UNIX, or any variant of the latter. No other software is required.

A.3.4 Pilot points

In 1999, Tikhonov regularization was implemented in PEST to facilitate its use in a highly parameterized inversion of magnetotelluric geophysical data. Shortly thereafter, software was added to the expanding Groundwater Data Utility suite to implement pilot points calibration in conjunction with the use of MODFLOW and MT3D so that highly parameterized inversion could then form the basis of groundwater model calibration.

A.3.5 Surface water utilities

During 2001, while working as a research scientist at the University of Idaho, John Doherty wrote the TSPROC utility program specifically to enhance the use of PEST in conjunction with surface water models, particularly models such as HSPF that are commonly employed in total maximum daily load (TMDL) modeling. Of particular focus was the need to handle large amounts of data in construction of complex multi-component objective functions, thereby allowing maximum information to be transferred from surface water flow and quality datasets to model parameters. It was also hoped that the functionality provided for highly parameterized inversion that was at the time proving successful in groundwater modeling (and was rapidly gaining popularity in that field) would also prove successful in surface water model calibration, especially in TMDL studies where representation of multiple land uses in large, diverse watersheds encompassing multiple sub-watersheds necessitates the use of many parameters. Both PEST and TSPROC have since been incorporated into the U.S. EPA BASINS package, where they are widely used.

A.3.6 MODFLOW-2000

PEST has never been the only option for MODFLOW-based parameter estimation. Prior to PEST, MODINV (also developed by John Doherty) was a popular option, with over 300 copies used worldwide. MODFLOWP (developed by the USGS) was also available, but received little use outside USGS circles. With the advent of MODFLOW-2000 (which saw parameter estimation functionality built right into MODFLOW), another calibration option became available to MODFLOW users. Nevertheless, despite being supported by PMWIN, GMS, and Groundwater Vistas, the parameter estimation functionality of MODFLOW-2000 did not gain widespread adoption, mainly for the following reasons.

1. The inversion engine employed by MODFLOW-2000 did not prove to be as robust as was hoped. Furthermore, functionality such as the bounding of estimated parameters was not supported.
2. The intimate connection between parameter estimation and groundwater flow simulation provided in MODFLOW2000 proved to be a liability rather than an advantage, as it afforded no flexibility in the design of an inverse problem beyond that provided by the MODFLOW-2000 package itself. Thus, for example, there was no possibility of joint inversion of MODFLOW in conjunction with any other model or of defining parameters in any way other than that offered by MODFLOW-2000.
3. The design structure of MODFLOW-2000 made the cost of highly parameterized inversion in general, and the use of pilot points in particular, very expensive.

A.3.7 UCODE

In recognition of the limitations of MODFLOW-specific parameterization, the USGS released UCODE in 2003. UCODE, like PEST, provides model-independent parameter estimation (and in its latest version supports the use of template and instruction files identical to those of PEST). However, it did not provide (and at the time of writing, still does not provide) an extensive utility base to support its ready application with commonly used models. Nor does it provide

functionality for calibration and uncertainty analysis based on highly parameterized inversion which, in the senior author's opinion, is fundamental to practical groundwater model calibration. Nevertheless, in response to the release of UCODE, PEST was placed in the public domain in 2003, where it has remained until this day. Over the years since its release, the user-base of UCODE has remained small, being mainly restricted to USGS personnel and those with whom they collaborate. Limited support is provided for it by commercial MODFLOW graphical user interfaces.

A.3.8 Highly parameterized inversion

The years 2002 to 2006 saw a surge in PEST development. Following from modeling work undertaken by the authors and others, it was becoming increasingly clear that calibration of two- and three-dimensional groundwater models required a more sophisticated approach than that offered by traditional methods of model calibration based on parsimonious parameterization. Success had already been experienced in using pilot points in conjunction with Tikhonov regularization, but nagging problems with the numerical instability of the inversion process were being experienced in many contexts. Another problem was that the use of increasing numbers of parameters created an increasing computational burden. This provided an upper limit on the extent to which highly parameterized inversion could be used in the groundwater model calibration context, despite the obvious benefits of this approach and despite the availability of Parallel PEST to decrease overall calibration runtimes. (By this time, a number of institutions were using banks of PCs dedicated solely to running Parallel PEST to calibrate their models.)

In 2003, subspace regularization, in the form of truncated singular value decomposition (SVD), was introduced as an alternative to Tikhonov regularization as a methodology for highly parameterized inversion. This, and a number of PEST-specific enhancements introduced to the Tikhonov scheme, overcame problems with numerical instability. However, the run-time burden of highly parameterized inversion remained.

A.3.9 SVD-Assist

In 2003, John Doherty, the senior author of this report, was working in conjunction with consulting colleagues on a large multilayer model in Southern California built to assess the extent of contamination in an aquifer and to design a system for its remediation. Doherty began investigating the idea of “super parameters” based on Eigen components forthcoming from singular value decomposition of the matrix of sensitivities of model outputs to initial parameter values. Through redefining the parameter estimation problem in terms of these super-parameters, while retaining Tikhonov constraints on base parameters, it was possible to formulate a highly parameterized inverse problem that could be solved with unconditional numerical stability and with only a fraction of the number of model runs per iteration that would have otherwise been required. The fact that these model runs could be parallelized further reduced the run-time demands of highly parameterized inversion. With the development of this “SVD-assist” methodology (documented in Tonkin and Doherty, 2005), it was now possible to use hundreds or even thousands of parameters in the calibration of models with runtimes over an hour. This, indeed, constituted a revolution in groundwater model calibration practice, for now the process complexity embodied in large physically-based models could be matched by a

parameterization complexity that could be handled with ease using the SVD-assist methodology. (The advantages of highly parameterized inversion in contexts such as these are explained in the next subsection of this document.)

A.3.10 LSQR

In early 2007, further support for highly parameterized inversion was provided through introduction of the LSQR method to PEST. LSQR is in many ways similar to SVD. However, at some little cost in the precision of subspace definition, it allows much more efficient implementation of subspace regularization when parameters number more than about 2000. This development was supported by Boise State University to provide an inversion mechanism to accommodate the adjoint process (for fast Jacobean matrix computation) that personnel from this university had programmed for MODFLOW-2005.

A.3.11 Linear uncertainty analysis

With the ability to accommodate large numbers of parameters in the inversion process also came the ability to undertake model parameter and predictive uncertainty analysis. From 2005 to 2007, the focus of PEST development (both theoretical and programming) turned to quantifying the potential for error in parameters estimated for a model, and in predictions made by a model. In this PEST development, the author John Doherty received a great deal of support from his students Catherine Moore, Matt Tonkin, and Mark Gallagher, while serving as a Professor at University of Queensland in Australia. As explained in the next subsection, such analyses require that parameterization complexity be represented in a model to a level that is commensurate with the level of predictive dependence on this complexity. A highly parameterized approach to calibration and uncertainty analysis is therefore indispensable. Over this period, more than 15 utilities were added to the PEST suite that enable computation of parameter and predictive uncertainty based on a linear approximation to the action of the model.

Computation of uncertainty is useful not just for its own sake alone. Once the uncertainty of a specific prediction has been computed, so too can the reduction in this uncertainty accrued through gathering different types of data. Based on this premise, utility software written over this time allows a user to optimize acquisition of future data in light of predictions that are required of his/her model. Other software allows a user to quantify the contribution made to predictive uncertainty by different parameter and boundary condition types, this allowing rapid assessment of model strengths and weaknesses, and ways in which the former can be amplified and the latter mitigated. See Moore and Doherty (2005) and Christensen and Doherty (2008) for theory, and Gallagher and Doherty (2007) for some applications of these methods.

PEST-suite utilities for linear uncertainty analysis (and ancillary tasks) are not currently supported by any MODFLOW graphical user interface. However, based on files produced by these interfaces, it is easily implemented using pertinent members of the PEST suite of software. Meanwhile, a graphical user interface for many of these utility programs is, at the time of writing, being developed by personnel from the South Florida Water Management District.

A.3.12 Nonlinear uncertainty analysis

Ideally, a calibrated model should represent the simplest solution to the inverse problem that is compatible with all that is known about a system. As such it is the least prone to parameter and predictive error. Conceptually, that error can be explored by generating many different parameter fields, all of which are as complex as the system that they purport to represent, and all of which maintain the model in a calibrated state.

Prior to the introduction into PEST of the “null space Monte Carlo” (NSMC) method, the development of a suite of calibration-constrained Monte-Carlo based parameter fields was too computationally demanding to be considered feasible except in working environments with access to large, parallelized computing resources (such as government laboratories working on nuclear waste disposal problems). The introduction of the null space Monte Carlo methodology into PEST made such methodologies available to modelers working in consulting offices. The NSMC method combines a nonlinear extension of linear uncertainty analysis theory with computational advantages provided by the SVD-assist methodology. See Tonkin and Doherty (2009) for a description. The NSMC method is fully supported by PEST and its utilities, and partially supported by Groundwater Vistas.

A.3.13 The future

In some ways, a significant destination has been reached in PEST development. With the development of the SVD-assist methodology, highly parameterized inversion is now available to the general modeling community. With null space Monte-Carlo and associated linear uncertainty analysis methods, highly parameterized calibration can be complemented by highly parameterized uncertainty analysis.

While many issues need to be addressed, the major obstacle facing ubiquitous use of these technologies by the groundwater community is now one of education – a task that will be addressed in the future to a greater extent than it has been in the past. Complementary to this is a need to address the important issue of how model-based management should take place. Based on advances in PEST over the last 5 years, the computational context in which modeling takes place is now different from what it was previously. On the one hand, highly parameterized calibration methodologies provided with PEST allow more information to be extracted from calibration datasets than was hitherto possible, allowing superior model performance to that which was achieved even in the recent past. On the other hand, the ability to quantify model parameter and predictive uncertainty allows graphic exposure of model limitations. Improvements in model-based decision-making must make use of both of these. This is another area on which future development work will be focused.

A.4 Why highly parameterized inversion?

A.4.1 General

Some of the mathematical and conceptual bases of model calibration will be provided in the next two chapters of this document. However, to underscore the importance of the approach to

parameter estimation and concomitant uncertainty analysis that has underpinned recent PEST development, the benefits of a highly parameterized approach to model calibration and uncertainty analysis are now briefly listed.

A.4.2 Classical parameter estimation

Traditionally, modelers are urged to reduce the number of estimable parameters to a minimum prior to estimating values for these parameters using computer-based techniques; see for example Hill (1998). The parameter estimation process then informs the user whether the level of precalibration parsimonization undertaken by him/her has been too great or too small. If too small, computer-based parameter estimation may not work or, if it does, will compute post-calibration parameter correlations that are very high. If too much parsimonization has taken place, the fit between model outputs and their field-measured counterparts will not be as good as the potential they have to be. In neither case will maximum information content have been extracted from the calibration dataset, and in neither case will the uncertainty of predictions required of a model have been reduced to its theoretical minimum based on the dataset available.

Disadvantages associated with this approach include the following.

- Selection of an appropriate parsimonization strategy is often difficult. While zones of piecewise constancy can be employed based on geological subdivision, the exact location of geological boundaries is often unknown and problematic. Furthermore, in some circumstances it is obvious that hydrogeological heterogeneity lies within geological units. In other circumstances, there may be too many mapped units for estimation of parameters based on them to constitute a well-posed inverse problem.
- In most modeling contexts, more parameters require estimation than can be uniquely estimated on the basis of the calibration dataset, especially when account is taken of the values that must be assigned to boundary conditions and many model inputs. The modeler is then faced with the task of how best to simplify what is an inherently complex knowledge system. A litany of subjective decisions must then follow. An inevitable question is then “if these decisions had been made differently, would different values have been assigned to estimated parameters during the calibration process? Would model predictions then be different?”
- It is difficult to know in advance of the parameter estimation process just how many parameters should be fixed or tied to each other, or simplified in other ways. A significant amount of trial and error is required to discover the optimal level of simplification.
- The use of zones of piecewise constancy (especially when included subjectively to account for intra-geological-unit heterogeneity that becomes apparent during the calibration process) can give the model domain the appearance of a geologically unrealistic patchwork quilt. Furthermore, use of zones of piecewise constancy as a parsimonization device may lead to the introduction of unnecessary structural noise that detracts from the fit between model outputs and field measurements, and obstructs the passage of information from field measurements to model parameters.
- Just because calibration rarely allows unique representation of a high level of hydraulic property detail within a model domain, this does not negate the fact that important

predictions are often sensitive to such detail (for example predictions of contaminant movement and/or groundwater surface water interaction). While the calibrated parameter field may provide the simplest interpretation of field data, it is mathematically impossible for it to provide an accurate representation of hydrogeological complexity because accuracy requires unique representation of an unachievable level of parameterization detail. While predictions based on a calibration-inferred parameter set may be minimally wrong, they are often far from right. Parsimonious calibration provides no mechanism for analyzing the propensity for predictive error that arises from this fact because detail that cannot be uniquely estimated is banished from the model.

A.4.3 Highly parameterized inversion

Some benefits of highly parameterized inversion are discussed by Hunt, et al. (2007). These, and some others, are briefly listed below.

- The modeler is not faced with the need to tie some parameters to others, and to fix other parameters at arbitrary values to formulate an inverse problem that is well-posed. Instead, simplification is undertaken mathematically as part of the calibration process itself. If that process is properly formulated, this simplification endows the resulting parameter field with minimum error variance status.
- The use of Tikhonov constraints that encapsulate geological formulation of a “preferred parameter condition” from which parameters must depart only to the smallest extent possible in calibrating the model, provides further impetus for achievement of a parameter field of minimum error variance.
- While a modeler can designate that heterogeneity must be introduced preferentially at geological boundaries, intra-zonal heterogeneity can arise naturally at those locations where this is necessary for satisfactory calibration to be achieved. Furthermore, the extent and nature of heterogeneity introduced within geological units is determined as part of the calibration process itself; thus heterogeneity does not need to express itself in terms of possibly inappropriately pre-defined information receptacles such as zones of piecewise constancy.
- Maximum information is transferred from data to parameters because good fits between model outputs and field measurements are often readily achieved (thereby minimizing the occurrence of structural noise). However limits can be placed on the goodness of fit achieved through the calibration process to prevent over-fitting to that data.
- The amount of uncertainty associated with model predictions can be explored with integrity. Though a model domain may have been endowed with a high number of parameters, the parameter field emerging from the regularized inversion process may in fact be simple as an outcome of mathematical regularization. However, geologically-realistic heterogeneity can be introduced when and if needed to explore the uncertainty associated with any predictions that are sensitive to such potential heterogeneity.

TWDB Report ##: Final – Application of PEST to Re-Calibrate
the Groundwater Availability Model for the Edwards-Trinity (Plateau) and Pecos Valley Aquifers

Appendix B The model calibration process

B.1 Some considerations

B.1.1 General

The phrase “model calibration” means different things to different people. The word “calibration” has other meanings in other fields of science – meanings that are not necessarily relevant to the field of environmental modeling. In fields such as petroleum reservoir modeling, the term “calibration” is not used at all; there the term “history-matching” is used in its stead. This may actually be a better term, as it does not have the overtones of high precision that are associated with the word “calibration” when employed in the context of fine tuning of laboratory equipment. Some would go so far as to say that the word “calibration” should not be used in the environmental modeling context at all because it conveys a sense of finality that has no justification. See, for example, Beven (2006).

B.1.2 Bayes' equation

Calibration involves adjustment of model parameters (which can be defined loosely as model inputs that represent system hydraulic properties, input stresses and/or boundary conditions) to obtain a better fit between model outputs and historical observations of system state. A mathematical description of what can be achieved through history-matching is provided by Bayes' equation, which is often written as follows:

$$P(\mathbf{p}|\mathbf{h}) \propto P(\mathbf{h}|\mathbf{p})P(\mathbf{p}) \quad (\text{B.1})$$

In this equation, the vector \mathbf{h} represents the elements of the calibration dataset while the vector \mathbf{p} represents model parameters. The symbol $P()$ represents probability. The term on the left-hand side of Eq. 3.1, the so called “posterior probability distribution”, is the probability associated with a set of parameters given knowledge of the data expressed in the vector \mathbf{h} . The first term on the right (often referred to as the “data likelihood function”) is the probability of the data values, given the same set of parameter values. The last term on the right is the “prior probability” of the parameters. Eq. D.1 expresses the fact that the calibration process is a type of filtering process. Prior to calibration we do not know exact parameter values. But we do have some idea of their range of plausibility, this being expressed by their prior probability distribution $P(\mathbf{p})$. Through calibration this probability distribution is modified, hopefully being made more narrow, for parameter sets that result in a better fit between model outcomes and field measurements are endowed with a higher probability than those that do not. This filtering operation is performed by the data likelihood function; the better the fit between model outputs and field measurements, the higher the likelihood function.

Put simply, Bayes' equation states that all that can be achieved from the process of history-matching is a parameter probability distribution (this being a refinement of the prior probability distribution) in which parameters that allow a better fit to be made between model outputs and field measurements (while still being plausible) are deemed to be more likely. The process of calibration is illustrated diagrammatically in Figure B-1. What is important to note is that no

inference can be drawn from Bayes' equation that the outcome of the history-matching process should be a single set of parameter values.

B.1.3 The quest for parameter uniqueness

Despite its rigor, history-matching is not often carried out strictly on the basis of Bayes' equation. (There are some notable exceptions to this; see, for example, the work of Kitanidis, 1997; Jiang and Woodbury, 2006; and references cited therein). Part of the reason for this is that certain mathematical difficulties attend its use with large observation datasets. Another reason is that probability distributions of parameters are harder to deal with numerically than single parameter sets (especially where a model is complex and highly nonlinear). A third reason is cultural; modelers and managers prefer to deal with a facade of certainty as expressed by the calibrated parameter set, rather than with the explicit acknowledgement of parameter and predictive uncertainty as enshrined in Bayes' equation.

However, even if it is not explicitly used in the history-matching process, and even if a single set of parameter values is sought through the history-matching process, Bayes' equation provides an insight into how a single set of parameters should be chosen. Ideally, the set of values chosen as those that are deemed to “calibrate” a model (designated as \mathbf{p} herein) should be those that are of maximum likelihood, and that therefore give rise to peak probability as expressed by the posterior parameter probability distribution of Bayes' equation. Alternatively, they may be selected to be of minimum potential error such that:

$$\int (\mathbf{p} - \mathbf{p})^2 P(\mathbf{p}|h) d\mathbf{p} \tag{B.2}$$

is minimized through a particular choice of \mathbf{p} . Either of these conditions will lead to a simplified parameter set that represents only as much hydraulic property detail as can be supported by the calibration dataset. Thus the calibrated parameter field \mathbf{p} should not be regarded as representing hydrogeological reality, for this can vary on a cell-by-cell basis within a model domain. It should be regarded as the set of parameters for which, at any point in space, the true parameter at that point in space has as much probability of exceeding the calibrated value as it does of being exceeded by the calibrated value. Thus the calibrated parameter field is much like the kriged field of spatial interpolation. Kriging is a smooth interpolator, with smoothness increasing as the number of data points from which it is interpolated decreases. Different versions of “reality” conditioned on these same interpolation points can be generated using a stochastic field generator. Each of these stochastic fields is different because it expresses detail that cannot be expressed uniquely; instead, they can only be expressed probabilistically. If many such stochastic fields are generated and all averaged, the kriged field would result. The calibrated parameter field should try to achieve the same relationship with reality as does the kriged field.

B.1.4 Regularization

In most settings where a model is employed to underpin decision-making, part of the model construction process involves parameterization based on history-matching. Through such history-matching, values are inferred for at least some combinations of hydraulic properties

within the model domain. These combinations of properties are normally relevant to large areas, and are thereby representative of spatially averaged values of hydraulic properties, rather than the properties at discrete points in space.

Intuition readily informs us that there is an upper limit to the level of hydraulic property detail that can be inferred from historical measurements of system state comprising a calibration dataset. Hence, attempts that are made to infer hydraulic properties must be accompanied by an explicit or implicit parameter simplification step. Ideally, the level of simplification must be such that only a small number of parameters require estimation or at least small enough that, given the way in which they are defined and the manner in which “best fit” is defined, the set of parameter values that is associated with an optimized fit between model outcomes and corresponding field measurements is unique. Thus, the search for an optimized fit leads to a unique set of (broad scale) parameter values.

The process of simplification that is undertaken to achieve parameter uniqueness is broadly referred to as “regularization”. Parameter estimation methodologies differ in the way in which regularization takes place. Some of these are discussed below.

B.2 Practical calibration

B.2.1 Manual calibration

As the name implies, manual calibration takes place through manual adjustment of parameter values such that a good fit between model outcomes and field measurements is sought on a trial-and-error basis. Normally a modeler decides prior to embarking on a manual calibration exercise which parameters he/she will adjust through this process, and which parameters will be assigned values based solely on intuition and/or direct measurement of pertinent system properties. Adjusted parameters normally pertain to broad-scale system properties, for if their manual adjustment is to have any effect, the areas to which they pertain must often comprise a significant portion of the entire model domain.

In the manual calibration procedure, “fit” is rarely defined. This is not necessarily a bad thing because most modelers know a “bad fit” when they see it. However, a frustrating element of the manual calibration process is that a modeler rarely knows if he/she can achieve a better fit through further parameter adjustment and/or how to achieve it.

Some disadvantages of manual model calibration are as follows.

1. As stated above, a modeler can never be sure whether current model-to-measurement fit can be improved beyond its present level and, if so, what parameters should be adjusted by what amounts to achieve this fit. Its trial-and-error nature, and lack of knowledge of whether to engage in further trials, can cause a good deal of frustration. The magnitude of the challenge increases where data are plentiful and widespread, and a propensity exists for estimation of many parameters. Manual adjustment of many parameters rapidly becomes time consuming and burdensome.

2. Manual calibration is not reproducible. It is normally terminated not when model-to-measurement fit has been optimized, but when a modeler simply cannot face doing yet another model run and/or when the modeling budget decrees that calibration must now be complete.
3. If a modeler has problems achieving a desired level of fit, he/she cannot be sure whether the problem is caused by:
 - b. Failure to find the minimum through adjustment of parameters that he/she currently decrees as adjustable;
 - c. The need to adjust more parameters than were originally decreed as adjustable; or
 - d. Model conceptual inadequacies.

The use of the model calibration process as one of testing current hydrogeological hypotheses thus becomes very difficult.

4. Suppose that a good level of fit results from the adoption of a certain set of parameter values. When values for these parameters have been found manually, a modeler cannot be sure whether other (and possibly very different) sets of parameter values may also give rise to the same level of fit, or to a level of fit that is not very different when considered relative to measurement noise. Hence he/she cannot know of the level of uncertainty that is associated with inference of adjustable parameters (and hence with model predictions that depend on these).
5. The extent to which the calibrated parameter field approaches that of maximum likelihood and/or minimum error variance is impossible to assess.

B.2.2 Computer-assisted calibration – classical approach

In the classical approach to computer-assisted calibration, regularization is undertaken by hand, whereas optimization of parameters defined on the basis of manual regularization is undertaken using computer software. A large number of optimization algorithms can be used for this purpose, all with different advantages and disadvantages. However in the groundwater modeling context, where model run times tend to be large, model-run-efficiency is often the selection criterion of primary concern. The Gauss-Marquardt-Levenberg method, and variants of it, provide the best performance in this regard, and so have found most use in the groundwater model calibration context. This method is discussed in greater detail below.

Implementation of classical computer-based parameter estimation mitigates some of the problems associated with manual calibration listed above. However a repercussion of the fact that regularization must still be undertaken manually (normally through precalibration definition of zones of piecewise constancy), is that to some extent the calibration process is not reproducible. Optimization through adjustment of values assigned to a small number of parameter zones is conditional upon the way in which these zones are defined. Where based on known or inferred geological boundaries, subjectivity can be removed. However, rarely is the number of defined geological units and their mapped disposition exactly matched to the capacity of the data to inform parameter values. Sometimes intra-geological-unit heterogeneity must be introduced to accommodate obvious signs of its presence in historically measured groundwater heads. At other times, hydraulic properties assigned to units must be tied in certain assumed

ratios, or fixed at certain assumed values, when information contained within the calibration dataset is not sufficient for their estimation. Subjective decisions (with possibly large repercussions for inducement of parameter and predictive error) are thus inevitable. Furthermore, because of the large role played by such subjective decisions in the manual calibration process, the calibration process cannot be said to achieve an “optimal” parameter set according to any objective metric. The metric of maximum likelihood and/or minimum error variance as discussed above therefore cannot be said to have been achieved.

B.2.3 Computer-assisted calibration – highly parameterized inversion

Where a highly parameterized approach is taken to model calibration, regularization is undertaken automatically, along with estimation of parameters. When defining parameters within a model domain that will be subjected to adjustment through the calibration process, the modeler thus adopts an entirely different philosophy to that which he/she adopts when undertaking manual regularization (either as a precursor to manual or classical parameter estimation). Whereas manual regularization demands parameterization parsimony, the opposite is the case where parameters are defined prior to undertaking highly parameterized inversion. In the latter case a suitable maxim is “the more parameters the better”. This gives subsequent computer-based regularization maximum freedom in implementing a regularization scheme that is maximally receptive to information residing within the calibration dataset. Appendix D presents some of the theory that determines how PEST implements regularization.

While computer-based regularization is not normally implemented in strict adherence with Bayes' equation, it can be shown that, if properly implemented, it does lead to estimation of parameter fields that do approach that of maximum likelihood and/or minimum error variance. Furthermore, the extent of possible error associated with a parameter field, and with predictions that depend on it, can be analyzed. Much of this potential for error arises out of the necessity to simplify (i.e., regularize) so that a unique solution to the inverse problem of model calibration can be obtained. However because parameters can be represented within a model to a level of detail that is commensurate with the sensitivity of model predictions to real-world hydraulic property detail, the “cost of uniqueness” can be tallied. Where management decisions are based on model predictions, the risk that arises from the uncertainty associated with these predictions can then be incorporated into the decision-making process.

As will be discussed below, computer-based regularization can be implemented in a number of ways, leading to some differences in the parameter fields achieved through model calibration. However these differences are slight compared with differences that arise out of procedures such as null-space Monte Carlo where (calibration-constrained) stochastic detail is ascribed to these parameter fields for the purpose of uncertainty analysis.

B.3 Definition of parameters

B.3.1 General

In theory, a “parameter” can be any model input that is unknown. Many model inputs fall into this category to a greater or lesser degree. Traditionally, most of these inputs are fixed at values

that are deemed to be “reasonable”. A small number of other parameters are then adjusted through the calibration process. Where regularization is implemented mathematically as part of the calibration process, however, there is no need for such ad-hoc parameter value assignment. In fact, provided regularization is properly implemented, attainment of a minimum variance parameter field requires that all aspects of model parameterization that are not known be included in the parameter estimation process. The regularized inversion process then adjusts parameters in such a way that maximum parameter likelihood is achieved.

In practice, there are some practical limitations on the number of parameters that can be introduced to the regularized inversion process in the context of groundwater model calibration. This is not the case in other contexts; for example in geophysical data interpretation it is not uncommon for a separate parameter to be assigned to every cell of a model domain; parameters featured in the inversion process can therefore number in the tens of thousands. However models that can be parameterized in this way are often provided with elegant, and extremely fast, methods through which derivatives of model outputs can be computed with respect to adjustable parameters. This is not the case for groundwater models. However the adjoint process of MODFLOW-2005 (Clemo, 2007) can provide this functionality in certain circumstances.

Where traditional parameter estimation is undertaken, representation of parameter spatial variability is often zone-based. This has its roots partly in the ease with which zones can be defined, and partly because zones can be designed to coincide with mapped features such as geological formations, land units, and soil types. Where highly parameterized inversion is undertaken, the goal of which is often to promulgate the appearance of intra-unit heterogeneity if and where this is supported by the data, zones of piecewise constancy become a less appealing parameterization device. Mathematically, there is no reason why many zone-based parameters based on a high zonal spatial density could not be used for model parameterization purposes. However the resulting “patchwork quilt” would not look very appealing (nor provide the visual impression that parameterization has achieved maximum likelihood status).

For spatially varying properties such as hydraulic conductivity, pilot points constitutes a superior parameterization device. These are briefly discussed below.

B.3.2 Pilot point parameterization

The use of pilot points as a spatial parameterization device in groundwater model calibration was initiated by de Marsily et al. (1984), Certes and de Marsily (1991), and LaVenue and Pickens (1992). Doherty (2003) was the first to use pilot points in the context of highly parameterized model calibration. This marked a departure from previous pilot-point usage in that the need to greatly limit the number of pilot-point parameters could be relaxed, thus allowing them to be distributed liberally throughout a model domain. This is the context in which they are now most commonly used in the modeling industry. Pilot points parameterization in the MODFLOW context is supported by Visual MODFLOW, GMS and Groundwater Vistas. The PEST Groundwater Data Utilities support its use with other models (such as MT3D, SEAWAT, MICROFEM and FEFLOW).

Where pilot points are employed as a parameterization device, hydraulic property values are assigned at the locations of these points. Hydraulic property values are then assigned to model grid cells through spatial interpolation between these points. If implemented using software supplied with the Groundwater Data Utilities, spatial interpolation is undertaken using kriging. While other interpolation methods could be used, kriging has the advantages that:

1. It is a smooth interpolator;
2. It respects parameter values at pilot points themselves.
3. Once interpolation factors have been generated, interpolation is computationally fast.
4. It can readily accommodate the presence of horizontal anisotropy.

Pilot points can be employed in conjunction with zone-based parameterization if desired. In this context, they can parameterize a multiplier field (with a base value of 1.0) superimposed on the zonal field (as was done for the ETPV GAM). Alternatively, different sets of pilot points can be ascribed to different zones, with spatial interpolation prevented across zone boundaries thus guaranteeing discontinuous parameter fields across these boundaries.

As will be discussed in the following sections of this report, pilot points can also be employed with parameters that are typically supplied to MODFLOW in tabular form. These include drain, river, general head boundary, and stream conductances and elevations.

B.3.3 Expanded parameterization

It is common modeling practice, even where calibration is based on regularized inversion, for some parameters to be fixed during the calibration process while others are estimated. Fixed parameters often include those associated with model inputs and boundary conditions.

Where there is some doubt about the value that a fixed parameter should be assigned, fixing that parameter at an incorrect value may lead to erroneous estimates of other parameters during the calibration process as a compensatory measure that is necessary to obtain a good fit with measured data.

Ideally, the potential for errors in values assigned to fixed parameters should be taken into account when analyzing the uncertainty of key model predictions. This can be accomplished in a number of ways. One option is to declare them as adjustable during post-calibration predictive uncertainty analysis. As will be described below, such analysis can be undertaken through linear or nonlinear means. Where undertaken linearly it is normally a relatively easy matter to include in the predictive analysis process, parameters that would not normally be considered as adjustable. This could be done, for example, to compute the effect that erroneous provision of boundary conditions may have on a prediction of interest. If the contribution made to potential predictive error by these “parameters” is small compared with contributions made by other parameters, modelers (and reviewers) can be assured that the assumptions that underpin boundary condition assignment may be violated to at least some degree without compromising the predictive ability of the model.

Gallagher and Doherty (2007) demonstrate this process for a water management model. In that case, the contribution made by the “no-flow” assumption at pertinent model boundaries was tested by introducing flow parameters to those boundaries to test the contribution that they may make to potential predictive error. Similarly, parameter status was temporarily afforded to coastal fixed heads to accommodate the affect that the neglect of a “tidal over height correction” due to action of waves on the seashore (see Song et al., 2006) may have on model parameterization and predictive error.

Another instance of an enhanced parameter set is provided in the present study where spatially variable multipliers are applied to historical pumping rates. Their declaration, as adjustable in a highly parameterized calibration exercise based on regularized inversion, does not constitute an assumption that they are in fact estimable on the basis of information contained within the calibration dataset. However it does enable the dataset to inform these parameters if indeed there is information available to do so. The fact that historical pumping is largely unknown, and that assignment of values to this pumping is likely to be in error (and can thereby incur errors in estimation of parameters that are sensitive to this pumping) can then be accommodated in model predictive uncertainty analysis.

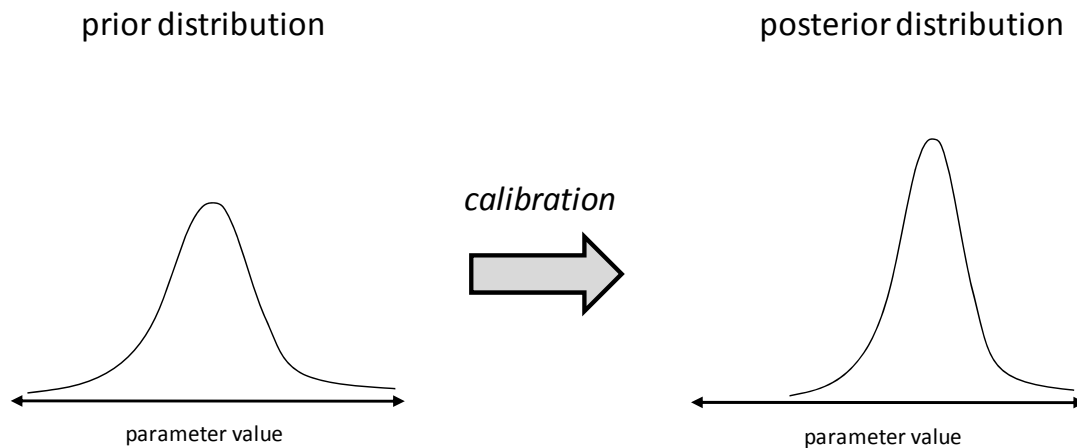


Figure B-1. The effect of history-matching as expressed by Bayes' equation.

TWDB Report ##: Final – Application of PEST to Re-Calibrate
the Groundwater Availability Model for the Edwards-Trinity (Plateau) and Pecos Valley Aquifers

Appendix C Utility software provided with PEST

C.1 PEST utilities

Following is a series of tables listing utility software installed with PEST, together with the function that each program serves. Programs are grouped into different tables according to similarity of function.

Parameter estimation

Program name	Purpose
PEST	Model-independent parameter estimation.
PPEST	Parallel version of PEST.

Checking utilities

Program name	Purpose
TEMPCHEK	Checks the integrity of a PEST template file.
INSCHEK	Checks the integrity of a PEST instruction file.
PESTCHEK	Checks an entire PEST input dataset for correctness and consistency.
JCOCHEK	Checks compatibility of PEST control file and corresponding JCO file.

Classical parameter estimation pre- and post-processing

Program name	Purpose
PARREP	Builds a new PEST control file whose initial values are optimized values from a previous PEST run.
PARAMFIX	Alters prior information pertaining to one or a number of parameters as these parameters are tied or fixed.
EIGPROC	Collects uncertainty, sensitivity and Eigen component information pertinent to a nominated parameter from PEST output files.
PCOV2MAT	Extracts a parameter covariance matrix from a PEST control file, re-writing it in matrix file format.
INFSTAT	Computes a suite of observation influence statistics, including DFBETAS and Cook's D.
PESTGEN	Builds a basic PEST control file based on a parameter value file and an INSCHEK output file.

Regularized inversion pre- and post-processing

Program name	Purpose
ADDREG1	Adds preferred value regularization to a PEST control file based on initial parameter values.
SUPCALC	Estimates number of super parameters to employ in SVD-assisted parameter estimation.
SVDAPREP	Writes a PEST input dataset for SVD-assisted parameter estimation.
PARCALC	Run as part of a model employed for SVD-assisted parameter estimation; computes base parameter values from super parameter values.
PICALC	Run as part of a model employed for SVD-assisted parameter estimation; computes prior information expressed in terms of base parameter values.

Program name	Purpose
IDENTPAR	Computes parameter identifiability.
PCLC2MAT	Computes base parameter composition of SVD-assist super parameters.
GENLINPRED	Automates running of PREDUNC and PREDVAR utilities. Undertakes linear predictive error analysis; also computes parameter identifiability and relative error variance and uncertainty reduction.
RESPROC	Processes information written in binary form to an *.rsd file during any PEST run in which regularization of any kind is employed.
RESWRIT	Processes information written by RESPROC; stores resolution and G matrices in PEST matrix file format.
REGERR	Computes the covariance matrix of regularization-induced structural noise.

Weights and covariance matrix manipulation

Program name	Purpose
COV2COR	Calculates a correlation coefficient matrix from a covariance matrix.
COVCOND	Calculates a conditioned covariance matrix from an unconditioned covariance matrix.
PWTADJ1	Alters weights in a PEST control file so that the contributions to the initial objective function by all observation groups are equal.
PWTADJ2	Attempts to create observation-group-specific weights which are the inverse of measurement error standard deviations.
WTFACTOR	Multiplies the weights pertaining to all observations belonging to a selected observation group by a specified factor.

Linear uncertainty analysis

Program name	Purpose
PREDUNC1	Computes the uncertainty of a user-specified prediction.
PREDUNC4	Computes contributions to predictive uncertainty by different parameters or parameter groups.
PREDUNC5	Computes observation worth through its effect in lowering predictive uncertainty.

Linear error analysis

Program name	Purpose
PARAMERR	Computes the covariance matrix of parameter error after a calibration exercise involving any form or regularization.
PREDERR	Computes the error variance of a prediction whose sensitivities are available after a calibration exercise involving any form or regularization.
PREDERR1	Similar to PREDERR, but slightly different in its input file requirements.
PREDERR2	Similar to PREDERR, but slightly different in its input file requirements.
PREDERR3	Similar to PREDERR, but slightly different in its input file requirements.
PREDVAR1	Computes the error variance of a model prediction based on a notional calibration exercise undertaken using truncated SVD; also finds the minimum of the predictive error variance curve.
PREDVAR1A	As for PREDVAR1, but undertakes SVD on $\mathbf{Q}^{1/2}\mathbf{X}$ rather than $\mathbf{X}^t\mathbf{Q}\mathbf{X}$.
PREDVAR4	Computes contribution made to the error variance of a prediction by different parameters and/or groups of parameters.
PREDVAR5	Computes observation worth through its effect on lowering predictive error variance.

Non-linear error analysis

Program name	Purpose
VECLOG	Computes the log of all elements of a vector (normally used as part of nonlinear highly parameterized predictive maximization/minimization).
PEST2VEC	Facilitates preparation for nonlinear highly parameterized predictive uncertainty analysis undertaken through constrained maximization /minimization.
VEC2PEST	Facilitates preparation for nonlinear highly parameterized predictive uncertainty analysis undertaken through constrained maximization /minimization.
OBSREP	Replaces observations in a PEST control file with best-fit model-generated equivalents. (This is normally run just prior to REGPRED.)
REGPRED	Builds a PEST control file in which post-calibration nonlinear predictive uncertainty analysis is effected by constrained prediction maximization/minimization.
RANDPAR	Computes random parameter values, placing these values into a series of parameter value files.
PNULPAR	Undertakes null-space projection of random parameter fields to remove solution space component; replaces it with solution space component from calibrated model.
RDMULRES	Reads multiple output files produced as an outcome of Monte-Carlo analysis, and collates results.
MULPARTAB	Builds a table of multiple sets of parameter values produced through null space Monte Carlo analysis.
COMFILNME	Facilitates post-null-space-Monte-Carlo file management.

Sensitivity data manipulation

Program name	Purpose
JACTEST	Undertakes serial or parallel model runs to test the integrity of finite-difference-calculated derivatives.
POSTJACTEST	JACTEST postprocessor; provides index of derivatives corruptness for different model outputs.
JACWRIT	Rewrites the contents of a *.jco file in ASCII format.
JCO2JCO	Writes a Jacobean matrix corresponding to a new PEST control file on the basis of information contained in an existing *.jco/*.pst file pair.
JCO2MAT	Rewrites the contents of a *.jco file in PEST matrix file format.
MAT2JCO	Writes a matrix file as a *.jco file; useful in converting between platforms and/or compilers with different binary file protocols.
JCOADDZ	Adds sensitivities to an existing *.jco file.
JCOCOMB	Builds a new *.jco file from an existing one, in which observations from the first are combined in user-supplied ratios in the second.
JCODIFF	Subtracts the contents of one *.jco file from that of another.
JCOORDER	Re-orders rows and/or columns in a *.jco file.
JCOPCAT	Concatenates two *.jco files; thus sensitivities with respect to some parameters can be computed on one machine and those with respect to other parameters can be computed on another.
JCOSUB	Replaces sensitivities in an existing *.jco file with those residing in another.
JCOTRANS	Translates from old *.jco storage format to new (compressed) storage format employed by PEST.
JROW2MAT	Extracts a row of a Jacobean matrix file and writes it in PEST matrix file format.
JROW2VEC	Extracts a row of a Jacobean matrix file, transposes it, and writes it in PEST matrix file format.
DERCOMB1	Combines two external derivatives files (supplied by models that can calculate

Program name	Purpose
MULJCOSEN	their own derivatives) into one, prior to being read by PEST. Reads multiple *.jco files as written on a single PEST run (if PEST is instructed to write such multiple files); calculates composite sensitivity of nominated parameter or observation from iteration to iteration.
WTSENOUT	Computes a weighted Jacobean matrix and a weighted observation vector.

Matrix manipulation

Program name	Purpose
MAT2SRF	Writes a matrix in SURFER grid format.
MATADD	Performs matrix addition.
MATCOLEX	Extracts a column of a matrix.
MATDIAG	Extracts the diagonal of a matrix.
MATDIFF	Performs matrix differencing.
MATINVP	Computes the inverse of a positive definite matrix.
MATJOINC	Joins matrices which possess the same number of columns.
MATJOIND	Joins two matrices in a diagonal sense (useful in forming a composite covariance matrix).
MATJOINR	Joins matrices which possess the same number of rows.
MATORDER	Re-orders the rows or columns of a matrix.
MATPROD	Performs matrix multiplication.
MATQUAD	Evaluates the quadratic form $\mathbf{y}^t \mathbf{M} \mathbf{y}$.
MATROW	Extracts a single row of a matrix.
MATSMUL	Multiplies a matrix by a scalar.
MATSPEC	Lists matrix specifications.
MATSVD	Undertakes singular value decomposition of an arbitrary matrix.
MATSYM	Forms a symmetric matrix as $(\mathbf{M} + \mathbf{M}^t)/2$.
MATTRANS	Computes the transpose of a matrix.
MATXTXI	Computes $(\mathbf{X}^t \mathbf{X})^{-1}$ where \mathbf{X} has more rows than columns.
MATXTXIX	Computes $(\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}$ where \mathbf{X} has more rows than columns.

Global optimization

Program name	Purpose
SCEUA_P	Global optimization using the SCEUA algorithm.
CMAES_P	Global optimization using the CMAES algorithm.

General

Program name	Purpose
PAR2PAR	Undertakes arbitrary mathematical manipulation of model parameters; normally run as part of a model calibrated by PEST.
SCALEPAR	Builds a PEST input dataset based on parameters scaled by their innate variability.
GENLIN	Generalized linear model.
PESTLIN	Reads a general PEST input dataset and accompanying *.jco file; creates a GENLIN model and accompanying PEST input dataset for calibration of that model.
SUPOBSPREP	Builds a new PEST input dataset comprised of super observations in place of native observations.
OBSCALC	Runs as part of a model to compute native observation-equivalents from model outputs corresponding to super observations.

Program name	Purpose
SIMCASE	Writes input files for a simplified PEST input dataset based on those for a more complex one.
SENSAN	Undertakes basic sensitivity analysis through repeated model runs.
SENSCHEK	Checks the integrity of a SENSAN input dataset.
PPAUSE	Pauses PEST execution.
PUNPAUSE	Unpauses PEST execution.
PSTOP	Stops PEST execution.
PSTOPST	Instructs PEST to cease execution with a full statistical printout.
PSLAVE	PEST slave program.
PSLAVE1	PEST slave program - slightly more efficient than PSLAVE.

C.2 Groundwater data utilities

Following is a series of tables listing software provided with the PEST Groundwater Data Utilities suite. Programs are grouped into different tables according to similarity of function.

Implementation of pilot points parameterization

Program	Purpose
FAC2FEFL	Uses PPKFAC_FEFL-generated kriging factors to modify a FEFLOW model input data file on the basis of spatial interpolation from a set of pilot points.
FAC2FEM	Uses PPK2FAC-generated kriging factors to produce a MicroFEM input file on the basis of spatial interpolation from a set of pilot points.
FAC2MF2K	Modifies an existing set of MODFLOW-2000 input files, replacing parameters cited in that file with pilot-point-based parameters (often a first step in pilot-point-based model calibration).
FAC2REAL	Uses PPKFAC or PPK2FAC1-generated kriging factors to produce a MODFLOW-compatible real array on the basis of spatial interpolation from a set of pilot points.
FAC2RSM	Uses PPKFACR-generated kriging factors to produce an RSM model input data file on the basis of spatial interpolation from a set of pilot points.
PPK2FAC	Calculates kriging factors for use in spatial interpolation from a set of pilot points to model grid cell centers.
PPK2FACF	Calculates kriging factors for use in spatial interpolation from a set of pilot points to the nodes of a MicroFEM finite element mesh.
PPK2FAC1	Identical to PPK2FAC except for the fact that the regularization data file which it writes is suitable for the use of PPKREG1.
PPK2FAC2	Identical to ppk2fac1 except for the fact that it prompts for a blanking radius.
PPK2FACR	Calculates kriging factors for use in spatial interpolation from a set of pilot points to the nodes of an RSM mesh. Regularization data file protocol is identical to that of PPK2FAC1.
PPK2FAC_FEFL	Calculates kriging factors for use in spatial interpolation from a set of pilot points to the elements of a FEFLOW mesh. Regularization data file protocol is identical to that of PPK2FAC1.
PARM3D	Assists in pilot-point parameterization of a 3-D model domain where hydrogeological units intersect grid layers.

MODFLOW/MT3D array manipulation

Program	Purpose
ARR2BORE	Undertakes spatial interpolation from a single array to a set of points.
INT2MIF	Generates MAPINFO MIF and MID files based on a MODFLOW/MT3D-compatible integer array.
INT2REAL	Builds a MODFLOW/MT3D-compatible real array based on the contents of a MODFLOW/MT3D-compatible integer array.
LOGARRAY	Evaluates the log (to base 10) of all elements of a real array.
PT2ARRAY	Builds a MODFLOW-compatible real array; the value assigned to each array element is calculated from information pertaining to points lying within the respective element.
REAL2INT	Builds a MODFLOW/MT3D-compatible integer array based on the contents of a MODFLOW/MT3D-compatible real array.
REAL2MIF	Generates MAPINFO MIF and MID files based on a MODFLOW/MT3D-compatible real array.
REAL2SRF	Translates a MODFLOW/MT3D-compatible real array into a SURFER grid file.
REAL2TAB	Translates a MODFLOW/MT3D-compatible real array into three-column real array table format.
SRF2REAL	Re-writes a SURFER grid file as a MODFLOW-compatible real array.
TAB2INT	Generates a MODFLOW/MT3D-compatible integer array from an integer array stored within a GIS.
TAB2REAL	Generates a MODFLOW/MT3D-compatible real array from a real array stored within a GIS.
TABCONV	Translates between integer or real array table files using row/column identifier format and those using cell number identifier format.
TWOARRAY	Combines two real arrays by addition, subtraction, multiplication, division and partial replacement.

MODFLOW/MT3D/SEAWAT pre-processing

Program	Purpose
MOD2ARRAY	Reads a MODFLOW or MT3D input file, extracting real or integer arrays from that file and storing them in separate files.
ELEV2CONC	Computes the elevation of the freshwater/saltwater interface on the basis of a sequence of concentration arrays.
ELEV2CONC1	Similar to ELEV2CONC, but computes zero flow head arrays as well.
REPARRAY	Pastes a MODFLOW or MT3D compatible real array into an existing MODFLOW or MT3D input file.

MODFLOW/MT3D/SEAWAT/FEFLOW post-processing

Program	Purpose
ARRDET	Lists the contents of a MODFLOW or MT3D binary head/drawdown/concentration output file.
BUD2HYD	Extracts flow data from a MODFLOW binary cell-by-cell flow term file. Rewrites this data in a form suitable for plotting against time.
BUD2SMP	Extracts flow data from a MODFLOW binary cell-by-cell flow term file. Rewrites this data in bore sample file format.
CONC2ELEV	Computes the elevation of the freshwater/saltwater interface on the basis of a sequence of concentration arrays.
DAR2SMP	Translates system states computed by a FEFLOW model to bore sample file format.

Program	Purpose
GETMULARR	Extracts arrays from MODFLOW/MT3D binary output files at user-nominated simulation times and stores these arrays in separate binary files.
GETMULARR1	Extracts all arrays for a nominated simulation time from a MODFLOW/MT3D binary output file and writes these to another binary MODFLOW/MT3D output file.
MANY2ONE	Splits MODFLOW/MT3D-generated binary files comprised of multiple two-dimensional results arrays into individual ASCII/binary files.
MOD2OBS	Interpolates model-generated data to the same times and locations as those cited in a user-supplied bore sample file; writes another bore sample file.
LAYDIFF	Evaluates head value differences in different layers based on contents of a bore sample file, bore coordinates file and bore listing file.
MOD2SMP	Interpolates the information contained in a binary MODFLOW/MT3D output file to a set of user-specified bores, rewriting the bore-specific data as a bore sample file.
MOD2SMPDIFF	Interpolates the information contained in a binary MODFLOW/MT3D output file to user-specified bores, calculating the difference or ratio between heads/concentrations at user-nominated pairs of bores.
SECTION	Interpolates the data contained in multiple MODFLOW-compatible real arrays to an arbitrary transect line through all or part of the finite-difference grid.

Processing and manipulation of field and model time series

Program	Purpose
FEM2SMP	Converts MicroFEM output to bore sample file format.
PMP2INFO	Builds a bore information file from a bore pumping file, the former containing cumulative pumped volumes between two user-specified dates for a user-supplied list of bores.
PMPCHEK	Checks the integrity of the data contained in a bore pumping file.
SMP2HYD	Rewrites the contents of a bore sample file for a user-specified list of bores in a form suitable for plotting borehole data against time.
SMP2INFO	Time-interpolates the information contained in a bore sample file to a user-specified date for a list of user-specified bores, thus writing a bore information file ready for access by commercial contouring software.
SMP2SMP	Interpolates data contained within one bore sample file to the dates and times represented in another bore sample file.
SMPCAL	Calibrates one time-series dataset on the basis of another.
SMPCHEK	Checks the integrity of a bore sample file.
SMPDIFF	Writes a new bore sample file in which differences are taken between successive values in an existing bore sample file, or between values in an existing file and a reference value.
SMPTREND	Writes a new bore sample file in which differences are taken between samples within an existing bore sample file and either the first sample for each bore in that file or a reference sample. However sampling is restricted to a yearly sample window.

Construction of a PEST input dataset

Program	Function
ADJOBS	Adjusts observation weights for different observation groups in a PEST control file according to user-defined formulae.
ARRAYOBS	Facilitates the introduction of model outputs comprised of MODFLOW/MT3D-compatible real arrays into a PEST parameter estimation process.
PESTPREP	Automates construction of a PEST control file and PEST instruction file for a

Program	Function
	model comprised of MODFLOW and/or MT3D followed by MOD2OBS, or MODFLOW followed by BUD2SMP followed by SMP2SMP.
PESTPREP1	Similar to PESTPREP. However provides extra flexibility in observation naming.
PESTPREP2	Similar to PESTPREP1. However allows extra observation data to be added to an existing PEST input dataset.

Adding regularization to a PEST input dataset

Program	Purpose
GENREG	Inserts prior information pertaining to many different types of regularization into an existing PEST control file.
PPCOV	Builds a covariance matrix pertaining to pilot point parameters based on one or a number of geostatistical structures.
PPKREG	Adds a prior information and regularization section to a PEST control file where parameterization is based on pilot points.
PPKREG1	Similar to PPKREG but more powerful in that it facilitates the use of both difference regularization (same as PPKREG) and preferred value regularization.
ZONE2VAR1	Computes a parameter variogram where parameterization is based on a large number of zones of piecewise constancy, and is defined through a ZONMDEF output file. Assists in undertaking variogram regularization as described by Johnson et al. (2007).
ZONE2VAR2	Computes a parameter variogram much more quickly than ZONE2VAR1 because it employs the results of the parameter search process undertaken by the latter program as read from a binary file written by it.
VERTREG	Adds vertical regularization prior information equations to a PEST control file where parameterization is based on pilot points.

Working with the MODFLOW adjoint process

Program	Function
ASENPROC	Reads a distributed parameter sensitivity file written by the adjoint state version of MODFLOW; formulates sensitivities for PEST parameters and writes them to a PEST external derivatives file.
MKMHOBBS	Reads a bore sample file. Writes a MODFLOW 2005 heads observation file, as well as an instruction file to read a MODFLOW heads output data file and a PEST building block file containing pertinent fragments of a PEST control file.
PPMDEF	Builds a parameter definition file for the use of ASENPORC, linking distributed parameters as employed by the adjoint process of MODFLOW to pilot point parameters.
ZONMDEF	Assists in the preparation of input files for the use of PEST in conjunction with the MODFLOW-2005 adjoint process where parameters are based on a large number of zones of piecewise constancy.

Uncertainty analysis

Program	Function
FIELDGEN	Generates a stochastic field in each zone of a model domain using the sequential Gaussian simulation method.
PPSAMP	Used in calibration-constrained Monte Carlo analysis. Samples stochastic fields at pilot point locations, interpolates between the pilot points, and generates difference fields.
PARCOV	Builds a geostatistically-based covariance matrix for a set of parameters whose coordinates are provided.

Geographical data manipulation

Program	Function
GRID2ARC	Writes ARCINFO generate files of the active part of the finite-difference grid as defined by a user-supplied integer array.
GRID2BLN	Writes a SURFER blanking file of the active part of the finite-difference grid as defined by a user-supplied integer array.
GRID2DXF	Writes a DXF file of the active part of the finite-difference grid as defined by a user-supplied integer array.
GRID2PT	Tabulates the coordinates of the cell centers of the finite-difference grid within an active window defined by a user-supplied integer array.
INT2MIF	Generates MAPINFO MIF and MID files based on a MODFLOW/MT3D-compatible integer array.
PTINGRID	Locates the finite-difference cells in which arbitrary, user-supplied points lie; optionally provides the value of an integer or real array element pertaining to the cell containing each such point.
QDIG2DXF	Translates the output of the shareware digitizing program, QDIGIT, into DXF format.
QDIG2XYZ	Translates the contours output of QDIGIT to an xyz data file.
RDAT2TAB	Reads an RSM element data file or index file. Adds mesh centroid coordinates to respective data elements and re-writes data in tabular format.
ROTBLEN	Rotates a SURFER blanking file about the top left corner of a finite-difference grid so that the component elements of the file can be overlain over the grid when the latter has been rotated such that its row direction is oriented directly east.
ROTDAT	Rotates a data file about the top left corner of a finite-difference grid so that the component elements of the file can be overlain over the grid when the latter has been rotated such that its row direction is oriented directly east.
ROTDXF	Rotates a DXF file about the top left corner of a finite-difference grid so that the component elements of the file can be overlain over the grid when the latter has been rotated such that its row direction is oriented directly east.
RSM2SRF	Reads an RSM (also GMS) 2D mesh file. Writes files through which SURFER can plot mesh design, outer mesh boundary, as well as nodes and element centroids.
ZONE2BLN	Writes a SURFER blanking file of finite-difference grid zonation as defined by a user-supplied, MODFLOW-compatible integer array.
ZONE2DXF	Writes a DXF file of finite-difference grid zonation as defined by a user-supplied, MODFLOW-compatible integer array.

C.3 Surface water utilities

Following is a series of tables listing software provided with the PEST Surface Water Utilities suite.

Program	Purpose
ADJOBS	Adjusts observation weights for different observation groups in a PEST control file using a simple user-adjustable expression.
PD_MS1	A PEST driver that undertakes many parameter estimation or predictive analysis runs based on random parameter starting values.
PD_MS2	A PEST driver that combines random sampling and parameter tracking to find the global minimum of the objective function.
PESTPRP1	Automates construction of PEST instruction and control files for calibration of models which generate output in site sample file format.
PLT2SMP	Builds a site sample file on the basis of a HSPF-generated plot file. Used as part of a composite model run by PEST.
SMP2HYD	Rewrites the contents of a site sample file for a user-specified list of sites in a form suitable for plotting against time.
SMP2SMP	Interpolates data contained within one site sample file to the dates and times represented in another site sample file.
SMP2VOL	Calculates volumes between arbitrary dates and times for flow samples listed in a site sample file.
SMPCAL	Calibrates one time series dataset against another.
SMPCHEK	Checks the integrity of a site sample file.
TSPROC	A comprehensive time series processor, designed for use as a stand-alone data processor, an environmental model post-processor, and a PEST input file generator.

Appendix D Some of the theoretical foundations of PEST

D.1 Introduction

D.1.1 General

This section provides a brief overview of the theory on which PEST is based. This presentation is not meant to be thorough; it is only meant to provide enough background information for a reader to understand how PEST estimates parameters, particularly in highly parameterized settings. For further details, see PEST documentation. See also Aster et al. (2005) for a broad coverage of the theoretical underpinnings of highly parameterized inversion.

All theory provided herein is linear, being based on manipulation of vectors and matrices. Non-linear parameter estimation and uncertainty analysis are based on these same equations. However, they are applied iteratively, with the Jacobean matrix (matrix of sensitivities) replacing model matrices discussed below.

D.1.2 Statistical preliminaries

Before describing the theory underlying PEST's parameter estimation and predictive uncertainty analysis functionality, a few useful equations and concepts will be provided.

Let x be a random variable. Because it is random, we can never know its exact value. By knowing its mean (represented herein by \underline{x}) and its standard deviation (represented herein by σ_x) and the probability distribution to which it belongs, we know as much about that variable as we can know.

The square of the standard deviation is referred to as the variance. As we shall see below, variance is a more natural quantity to work with than standard deviation because it is easily transmitted and manipulated through linear equations. The variance of x is denoted by σ_x^2 .

Let \mathbf{x} be a vector of random variables. Thus, if \mathbf{x} has n elements x_i , each of these is an individual random variable in its own right. Therefore, each has a mean \underline{x}_i and a variance $\sigma_{x_i}^2$. We label the vector of means $\underline{\mathbf{x}}$. However, to fully describe the stochastic properties of all x_i , we need more than a vector of variances for this takes no account of stochastic interaction between these elements; for example, the tendency of x_i to be high if x_j is low. A full description of the stochasticity of \mathbf{x} requires a matrix which we label $C(\mathbf{x})$ and call the covariance matrix of \mathbf{x} . The diagonal elements of $C(\mathbf{x})$ are indeed the variances of the elements of \mathbf{x} . Off-diagonal elements, if non-zero, express the degree of statistical interdependence of the elements x_i of \mathbf{x} . $C(\mathbf{x})$ is always symmetric (and positive definite).

Suppose that the vector \mathbf{y} is calculated from the vector \mathbf{x} using the equation:

$$\mathbf{y} = \mathbf{Ax} \tag{D.1a}$$

Then \mathbf{y} too is a random vector. Therefore, it possesses a mean $\underline{\mathbf{y}}$ and covariance matrix $C(\mathbf{y})$. These are calculated from respective quantities pertaining to the vector \mathbf{x} using the equations:

$$\underline{\mathbf{y}} = \mathbf{A}\underline{\mathbf{x}} \quad (\text{D.1b})$$

$$C(\mathbf{y}) = \mathbf{A}C(\mathbf{x})\mathbf{A}^t \quad (\text{D.1c})$$

D.2 Classical parameter estimation

D.2.1 Problem definition

Let the vector \mathbf{p} define a set of model parameters. It will be assumed that these are small in number so that they can be estimated uniquely on the basis of a calibration dataset. This dataset comprises the elements of the vector \mathbf{h} . Let it be assumed that \mathbf{p} contains m elements and that \mathbf{h} contains n elements. For uniqueness of estimation, n must be at least equal to m and should mostly be considerably greater than m .

Let the $n \times m$ matrix \mathbf{X} represent the action of a model on its parameters to produce a set of outputs \mathbf{o} corresponding to elements of the calibration dataset \mathbf{h} . Thus:

$$\mathbf{o} = \mathbf{X}\mathbf{p} \quad (\text{D.2})$$

where \mathbf{p} in Eq. D.2 is an arbitrary set of parameters, and \mathbf{o} is a corresponding set of model outputs. Let \mathbf{p}^* represent the parameters that exist in the real world. These are the parameters that we need to estimate, but whose real values we will never know. Furthermore, let the n -dimensional vector $\boldsymbol{\varepsilon}$ represent noise associated with measurements \mathbf{h} ; this is a random vector for which the covariance matrix will be designated as $C(\boldsymbol{\varepsilon})$. Then the basic equation for our model system, and the one that defines the calibration context in which we are operating, is:

$$\mathbf{h} = \mathbf{X}\mathbf{p}^* + \boldsymbol{\varepsilon} \quad (\text{D.3})$$

D.2.2 Objective function

Let the vector of residuals \mathbf{r} define the misfit between model outputs and field measurements calculated on the basis of an arbitrary set of parameters \mathbf{p} . Then:

$$\mathbf{r} = \mathbf{h} - \mathbf{o} = \mathbf{h} - \mathbf{X}\mathbf{p} \quad (\text{D.4})$$

To characterize overall model-to-measurement misfit, let us define an objective function Φ as:

$$\Phi = \mathbf{r}^t\mathbf{Q}\mathbf{r} = (\mathbf{h} - \mathbf{X}\mathbf{p})^t\mathbf{Q}(\mathbf{h} - \mathbf{X}\mathbf{p}) \quad (\text{D.5a})$$

where \mathbf{Q} is a weight matrix which in most real-world cases is chosen to be diagonal. In the latter case, let the i th diagonal element of \mathbf{Q} be designated as w_i^2 ; this being the weight assigned to the i th observation. Eq. D.5a then becomes:

$$\Phi = \sum (w_i r_i)^2 \quad (\text{D.5b})$$

this being the sum of squared weighted residuals.

D.2.3 Parameter estimates

We will declare a model to be calibrated when Φ is reduced to a minimum. We will declare the parameter set \mathbf{p} for which this minimum occurs to be the parameter set which calibrates the model. With only a little vector calculus it is easy to show that:

$$\mathbf{p} = (\mathbf{X}^t \mathbf{Q} \mathbf{X})^{-1} \mathbf{X}^t \mathbf{Q} \mathbf{h} \quad (\text{D.6})$$

where \mathbf{X}^t is the transpose of \mathbf{X} , and the -1 superscript designates the matrix inverse.

D.2.4 Parameter error

If Eq. D.3 is now substituted into Eq. D.6, we obtain:

$$\mathbf{p} = (\mathbf{X}^t \mathbf{Q} \mathbf{X})^{-1} \mathbf{X}^t \mathbf{Q} \mathbf{X} \mathbf{p}^* + (\mathbf{X}^t \mathbf{Q} \mathbf{X})^{-1} \mathbf{X}^t \mathbf{Q} \boldsymbol{\varepsilon} \quad (\text{D.7a})$$

That is:

$$\mathbf{p} = \mathbf{p}^* + (\mathbf{X}^t \mathbf{Q} \mathbf{X})^{-1} \mathbf{X}^t \mathbf{Q} \boldsymbol{\varepsilon} \quad (\text{D.7b})$$

from which the error in estimated parameters is calculated as:

$$\mathbf{p} - \mathbf{p}^* = (\mathbf{X}^t \mathbf{Q} \mathbf{X})^{-1} \mathbf{X}^t \mathbf{Q} \boldsymbol{\varepsilon} \quad (\text{D.7c})$$

We cannot know parameter error because we cannot know $\boldsymbol{\varepsilon}$. However, we can know its stochastic properties, and hence, the propensity for parameter error. From Eq. D.1 and the fact that $\boldsymbol{\varepsilon}$ is a random vector, we can obtain the covariance matrix of parameter error $C(\mathbf{p} - \mathbf{p}^*)$ as:

$$C(\mathbf{p} - \mathbf{p}^*) = (\mathbf{X}^t \mathbf{Q} \mathbf{X})^{-1} \mathbf{X}^t \mathbf{Q} C(\boldsymbol{\varepsilon}) \mathbf{Q}^t \mathbf{X} (\mathbf{X}^t \mathbf{Q} \mathbf{X})^{-1} \quad (\text{D.8})$$

Suppose that the weight matrix \mathbf{Q} was specifically chosen to satisfy the equation:

$$\mathbf{Q} = \sigma_h^2 C^{-1}(\boldsymbol{\varepsilon}) \quad (\text{D.9})$$

If noise $\boldsymbol{\varepsilon}$ associated with the elements of \mathbf{h} is statistically independent so that $C(\boldsymbol{\varepsilon})$ is diagonal, then \mathbf{Q} chosen according to Eq. D.9 is also diagonal; furthermore, each of its elements w_i^2 is proportional to the inverse of the variance of the noise associated with the corresponding element of \mathbf{h} . Thus, the weight associated with each h_i is proportional to the inverse of the standard deviation of noise associated with h_i .

Using Eq. D.9, Eq. D.8 becomes:

$$C(\mathbf{p} - \mathbf{p}^*) = \sigma_h^2(\mathbf{X}^t\mathbf{Q}\mathbf{X})^{-1} \quad (\text{D.10})$$

It can be shown that this choice of \mathbf{Q} leads to minimized variance of any element of $\mathbf{p}-\mathbf{p}^*$ and of any linear combination of elements of $\mathbf{p}-\mathbf{p}^*$ (including model predictions).

D.2.5 Predictive error

Let s be a model prediction, calculable from model parameters as:

$$s = \mathbf{y}^t\mathbf{p} \quad (\text{D.11a})$$

where \mathbf{y} is a vector of sensitivities of the prediction to the parameters \mathbf{p} of the model. The true value of the prediction s^* is made on the basis of the true model parameters \mathbf{p}^* . That is:

$$s^* = \mathbf{y}^t\mathbf{p}^* \quad (\text{D.11b})$$

The prediction \underline{s} made on the basis of the calibrated model is given by:

$$\underline{s} = \mathbf{y}^t\mathbf{p} \quad (\text{D.11c})$$

From Eqs. D.11b and D.11c, the error incurred through making the prediction using the calibrated model is given by:

$$\underline{s} - s^* = \mathbf{y}^t(\mathbf{p} - \mathbf{p}^*) \quad (\text{D.12})$$

As for parameters, we cannot know predictive error. But from Eqs. D.1, D.10, and D.12, we can evaluate the covariance matrix of predictive error (which for a scalar becomes the variance of predictive error) as:

$$\sigma_{\underline{s} - s^*}^2 = \sigma_h^2 \mathbf{y}^t(\mathbf{X}^t\mathbf{Q}\mathbf{X})^{-1}\mathbf{y} \quad (\text{D.13})$$

D.2.6 Nonlinear parameter estimation

Estimation of parameters for a nonlinear model is based on Eq. D.6. However, in the nonlinear context, some changes are necessary.

1. The model matrix \mathbf{X} is replaced by the Jacobean matrix \mathbf{J} . Each column of the Jacobean matrix is the partial derivatives of every model output (i.e., every element o_i of \mathbf{o}) with respect to a single parameter (i.e., a single element p_i of \mathbf{p}). This matrix is often computed using finite parameter differences, and therefore, requires at least as many model runs to fill as there are parameters p_i to be estimated.
2. Parameter upgrades are computed instead of optimized parameters. Thus, the nonlinear parameter estimation process is an iterative process in which:
 - a. a Jacobean matrix \mathbf{J} is computed on the basis of current parameter values;
 - b. an improved set of parameters is computed using Eq. D.6.

- Eq. D.6 is modified slightly by addition of a diagonal term to the matrix $\mathbf{X}^t\mathbf{Q}\mathbf{X}$ before inversion. The revised equation becomes:

$$\mathbf{p} = (\mathbf{X}^t\mathbf{Q}\mathbf{X} + \lambda\mathbf{I})^{-1}\mathbf{X}^t\mathbf{Q}\mathbf{h} \quad (\text{D.14})$$

λ is known as the Marquardt lambda. It performs a number of roles including that of stabilizing the parameter estimation process if the matrix $\mathbf{X}^t\mathbf{Q}\mathbf{X}$ becomes difficult to invert. Its value is recalculated during every iteration of the nonlinear parameter estimation process. Ideally, it should fall as the process progresses. If it rises instead, this indicates to the user that he/she is attempting to estimate more parameters than the measurement dataset \mathbf{h} can sustain.

D.2.7 Discussion

A number of important features of classical parameter estimation can be gleaned from the above brief mathematical coverage.

Firstly, successful estimation of parameters using classical parameter estimation techniques is predicated on the assumption that the matrix $\mathbf{X}^t\mathbf{Q}\mathbf{X}$ of Eq. D.6 can be inverted. If it cannot be inverted, or is only just invertible, then pre-calibration regularization has been insufficient. Under these circumstances parameter error as calculated using Eq.D.10 will be large. The modeler must then re-formulate the inverse problem and attempt parameter estimation again. Alternatively, a modeler may find that the objective function has not been lowered as much as he/she hoped that it would be, with the minimized objective function resulting from the parameter estimation process being too high. In that case the modeler may have undertaken too much pre-calibration regularization.

A second important feature of classical parameter estimation follows from Eqs. D.10 and D.13. It is apparent from these equations that (according to the theory which underpins classical parameter estimation) all parameter and predictive error is ultimately derived from measurement noise. As Moore and Doherty (2005; 2006) show, this is not the case. In fact, the largest contributor to parameter error is the fact that parameters must be grossly simplified before their estimation is attempted; the less data that is available for estimation of parameters, the greater must this simplification be. As classical parameter estimation can only take place after manual pre-calibration regularization has been undertaken by the user, it knows nothing of this regularization and hence cannot accommodate it when computing the propensity for parameter and predictive error.

Miscalculation of parameter and predictive error variance in this manner occurs because Eq. D.3 was substituted into Eq. D.6 in deriving Eq. D.10. In fact, Eq. D.3 is not correct, for reality is described by an equation that involves many more parameters than are seen by the classical parameter estimation process. Specifically, reality is better described by the following equation.

$$\mathbf{h} = \mathbf{Z}\mathbf{k} + \boldsymbol{\varepsilon} \quad (\text{D.15})$$

where \mathbf{k} is a vector with many more elements than \mathbf{p} , this reflecting the fact that temporal and spatial variability of hydraulic properties is much more complicated, and requires many more parameters for its expression, than that which can be accommodated by classical parameter estimation. In ignoring this fact, equations such as D.10 and D.13 lead to gross underestimation of the potential for error of parameters estimated through the calibration process, and predictions which depend in them.

D.3 Highly parameterized inversion

D.3.1 Problem definition

Highly parameterized inversion commences with Eq. D.15. It acknowledges that the vector \mathbf{k} of model parameters contains far more elements than can ever be estimated uniquely. However, when formulating the inverse problem, a very different philosophy is followed. It is that if a parameter is: 1) at least partially unknown, and 2) salient to predictions required by the model, then it should be featured in the parameter estimation process and in the model predictive uncertainty analysis that follows it.

We will designate by $C(\mathbf{k})$ the covariance matrix of parameter variability. This is similar to the prior parameter probability distribution of Eq. B.1. It provides a stochastic description of innate parameter variability, based on expert knowledge (or lack of knowledge) alone. Diagonal elements of this matrix express their pre-calibration uncertainty range. Off-diagonal elements express spatial and temporal correlation between parameters of the same or different types. In some situations at least part of $C(\mathbf{k})$ may be informed by a variogram or other outcome of detailed geostatistical site characterization. In most situations however $C(\mathbf{k})$ will be the outcome of a modeler's intuition.

While $C(\mathbf{k})$ is not directly used in the model calibration process, ideally parameters specified in \mathbf{k} of Eq. D.15 should be scaled by their innate variability before being introduced to the regularized inversion process as this leads to parameter and prediction estimates of minimum error variance.

D.3.2 The null space

The matrix \mathbf{Z} of Eq. D.15 may have more rows than columns, or more columns than rows; numerical predominance of observation numbers over parameter numbers is not required for highly parameterized inversion. In both cases, so long as parameters k_i comprising the vector \mathbf{k} are too great in number to be uniquely estimable, the matrix \mathbf{Z} possesses a null space. This means that there are at least some vectors $\delta\mathbf{k}$ for which:

$$\mathbf{0} = \mathbf{Z}\delta\mathbf{k} \tag{D.16}$$

This means that if any parameter set \mathbf{k} can be found that solves the inverse problem of model calibration such that:

$$\mathbf{h} \approx \mathbf{Z}\mathbf{k} \tag{D.17}$$

then $\underline{\mathbf{k}} + \delta \mathbf{k}$ constitutes another solution to the inverse problem, as this vector also obeys Eq. D.17.

Given the fact that an infinite number of solutions exist to the so-called under-determined inverse problem of highly parameterized inversion, how does one select a single parameter set $\underline{\mathbf{k}}$ which is endowed with the status of the calibrated parameter set? Different methods of mathematical regularization achieve slightly different $\underline{\mathbf{k}}$ vectors. However, the two methods that are discussed herein, namely Tikhonov and singular value decomposition, provide $\underline{\mathbf{k}}$ vectors which approach minimum variance estimates of the true \mathbf{k}^* vector of model parameters. These two matrices can be shown to yield identical solutions to the highly parameterized inverse problem of model calibration as measurement noise ϵ approaches zero.

D.3.3 Tikhonov regularization

Tikhonov regularization creates uniqueness of the inverse problem of model calibration by supplementing the information contained in the measurement set \mathbf{h} with a suite of information that pertains directly to parameters \mathbf{k} . This expresses either preferred values for \mathbf{k} , or preferred values for relationships between the elements of \mathbf{k} . Collectively, these relationships should define a preferred parameter condition, this condition being carefully defined in accordance with current hydrogeological understanding. Tikhonov regularization creates uniqueness in solution of the inverse problem of model calibration by seeking that parameter set $\underline{\mathbf{k}}$ that matches observations \mathbf{h} to within a specified tolerance, and for which departure from preferred parameter conditions as expressed by Tikhonov constraints is a minimum.

Mathematically, concepts underpinning Tikhonov regularization are expressed as follows. Let Tikhonov constraints be expressed by the matrix equation:

$$\mathbf{T}\mathbf{k} = \mathbf{0} \tag{D.18}$$

and let a regularization objective function Φ_r express the extent to which parameter values \mathbf{k} depart from their preferred condition as:

$$\Phi_r = (\mathbf{T}\mathbf{k})^t \mathbf{Q}_r (\mathbf{T}\mathbf{k}) \tag{D.19}$$

where \mathbf{Q}_r is an appropriate user-specified weight matrix.

Let the measurement objective function Φ_m be defined as per Eq. D.5a, i.e.

$$\Phi_m = (\mathbf{h} - \mathbf{Z}\mathbf{k})^t \mathbf{Q} (\mathbf{h} - \mathbf{Z}\mathbf{k}) \tag{D.20}$$

In PEST, Tikhonov regularization is implemented as a constrained optimization problem as follows:

Minimize Φ_r subject to the constraint that Φ_m rises no higher than a target measurement objective function Φ_m specified by the user.

It can be shown that solution to this problem is obtained as:

$$\underline{\mathbf{k}} = (\mathbf{Z}^t\mathbf{Q}\mathbf{Z} + \beta^2\mathbf{T}^t\mathbf{Q}\mathbf{T})^{-1}\mathbf{Z}^t\mathbf{Q}\mathbf{h} \quad (\text{D.21})$$

where β^2 is a Lagrange multiplier calculated during the inversion process.

Tikhonov regularization has the desirable quality that geological expertise is built into the inversion process through definition of a preferred parameter condition. Over-fitting is prevented through selection of a target measurement objective function Φ_m which is commensurate with measurement/structural noise. However despite its conceptual appeal, numerical instability can sometimes be encountered in its implementation, especially towards the end of the inversion process where parameters are approaching their optimized values. PEST provides a number of devices that attempt to overcome this problem. Depending on the calibration context these achieve varying degrees of success.

D.3.4 Singular value decomposition

As stated above, Tikhonov regularization achieves uniqueness through *adding* to the inverse problem a suite of information that pertains directly to parameters. Singular value decomposition (SVD) achieves uniqueness by *subtracting* from the inverse problem parameters, and combinations of parameters, that cannot be estimated because of information inadequacy in the measurement dataset \mathbf{h} .

Any matrix \mathbf{Z} used in Eq. D.15 can be decomposed as follows:

$$\mathbf{Z} = \mathbf{U}\mathbf{S}\mathbf{V}^t \quad (\text{D.22})$$

where \mathbf{U} is a matrix whose columns are orthogonal unit vectors spanning the range space of the model \mathbf{Z} , \mathbf{V} is a matrix of orthogonal unit vectors spanning the domain of the model \mathbf{Z} (i.e., parameter space), and \mathbf{S} is a diagonal matrix of singular values, normally arranged from highest to lowest. Where \mathbf{Z} has a null space, some of these singular values are zero. In most cases, many other singular values are close enough to zero to be considered as such when estimating parameters. If non-zero singular values are assigned to the \mathbf{S}_1 matrix and zero-valued singular values are assigned to the \mathbf{S}_2 matrix, Eq. D.22 becomes:

$$\mathbf{z} = [\mathbf{U}_1 \quad \mathbf{U}_2] \begin{bmatrix} \mathbf{S}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_2 \end{bmatrix} \begin{bmatrix} \mathbf{V}_1 \\ \mathbf{V}_2 \end{bmatrix} \quad (\text{D.23})$$

Declaring \mathbf{S}_2 as zero and substituting into Eq. D.15, we obtain:

$$\mathbf{h} = \mathbf{U}_1\mathbf{S}_1\mathbf{V}_1^t\mathbf{k} + \boldsymbol{\varepsilon} \quad (\text{D.24})$$

Recognizing that:

$$\mathbf{U}_1^t\mathbf{U}_1 = \mathbf{I} \quad (\text{D.25})$$

because it is an orthogonal matrix, and that \mathbf{S}_1 is invertible because it is a diagonal matrix with non-zero elements, a little manipulation of Eq. D.24 leads to:

$$\mathbf{V}_1 \mathbf{V}_1^t \mathbf{k} = \mathbf{V}_1 \mathbf{S}_1^{-1} \mathbf{U}_1^t \mathbf{h} + \mathbf{V}_1 \mathbf{S}_1^{-1} \mathbf{U}_1^t \boldsymbol{\varepsilon} \quad (\text{D.26})$$

$\mathbf{V}_1 \mathbf{V}_1^t \mathbf{k}$ is the projection of \mathbf{k} onto the subspace of parameter space spanned by the orthogonal unit vectors comprising the columns of \mathbf{V}_1 . This becomes our estimated parameter set $\underline{\mathbf{k}}$. As we do not know the noise associated with the calibration dataset \mathbf{h} , $\underline{\mathbf{k}}$ is calculated from \mathbf{h} alone, so that (D.26) becomes:

$$\underline{\mathbf{k}} = \mathbf{V}_1 \mathbf{S}_1^{-1} \mathbf{U}_1^t \mathbf{h} \quad (\text{D.27})$$

The projection operation through which SVD is used to solve the inverse problem of highly parameterized model calibration is illustrated in Figure D-1. Parameter space (spanned by the columns of \mathbf{V}) is three-dimensional in this figure, while the calibration solution space (spanned by the columns of \mathbf{V}_1) is two-dimensional. The \mathbf{v}_i in Figure D-1 are the unit vectors comprising the columns of \mathbf{V} .

In summary, when SVD is employed as an agent for highly parameterized inversion, it first subdivides parameter space into two subspaces. The first is comprised of orthogonal combinations of parameters which are not estimable on the basis of the current calibration dataset; these comprise the calibration null space. The orthogonal complement of this is the solution subspace, this being spanned by orthogonal combinations of parameters that are indeed estimable on the basis of the current calibration dataset. SVD restricts itself to estimation of these estimable parameter combinations, leaving the other combinations unestimated (and hence retaining their initial values, which should be set in accordance with hydrogeological expertise).

An advantage of using SVD for highly parameterized inversion is that it is unconditionally numerically stable. It can be shown to yield the solution to the inverse problem which is of minimum norm. If $\mathbf{C}(\mathbf{k})$ is diagonal with elements of equal value, minimum norm is equivalent to maximum likelihood (for a symmetrical parameter probability distribution). A disadvantage of SVD is that parameter fields forthcoming from this process do not have quite the same visual appeal as those forthcoming from Tikhonov regularization, as an expert's knowledge does not occupy the same pivotal position in SVD-based inversion as it does in Tikhonov-based inversion. Also, the user does not have the ability to set a target measurement objective function as he/she does in implementing Tikhonov regularization; however the user can select the cut-off point between singular values that are assigned to the \mathbf{S}_1 matrix and those that are assigned to the \mathbf{S}_2 matrix, thereby defining the dimensions of the solution and null spaces for use during the parameter estimation process. The smaller are the dimensions of the solution space, the fewer parameter combinations are estimated, and the higher will be the final objective function achieved through the calibration process.

Some of the above problems can be rectified by using Tikhonov regularization simultaneously with SVD.

As an alternative to SVD, PEST offers the LSQR algorithm of Paige and Saunders (1982a, 1982b) for estimation of parameter projections onto the calibration solution space. This is much more numerically efficient than SVD where parameters number more than about 2000. However a slight cost in numerical definition of solution and null spaces is incurred.

D.3.5 Nonlinear parameter estimation

Formulas presented above pertain to linear models, that is, models that can be represented as a matrix. When undertaking highly parameterized inversion for nonlinear models, an iterative process is adopted in which the model is successively relinearized through computation of the Jacobean matrix \mathbf{J} ; this is then used in place of \mathbf{Z} in the above equations. During each such iteration, an improved parameter set is computed, rather than the final, calibrated parameter set. A Marquardt lambda can also be employed to facilitate rapid parameter movement at early stages of the parameter estimation process, and to safeguard against incipient nonuniqueness when undertaking Tikhonov regularization on its own.

D.3.6 SVD-Assist

The main obstacle to implementation of highly parameterized inversion in the groundwater modeling context is the numerical burden of calculation of the Jacobean matrix. This requires as many model runs per iteration as there are estimable parameters. Where parameters number in the hundreds or thousands, and where model run times are large, this can impose an intolerable computational burden, even where model runs are parallelized.

PEST provides a unique hybrid scheme named SVD-assist which reduces the numerical burden of highly parameterized inversion to a level that is commensurate with that required for classical parameter estimation. Using this scheme, a full Jacobean matrix is computed only once - this being at the commencement of the parameter estimation process. SVD is then undertaken on the weighted Jacobean matrix to define a set of super parameters, these being estimable parameter combinations which span the calibration solution space associated with the current inverse problem. The inverse problem is then re-formulated in terms of these super parameters, so that only as many model runs are required per iteration as there are super parameters. At the same time, Tikhonov regularization (defined in terms of native model parameters) can be employed, thereby allowing the user to define a suitably chosen target measurement objective function and to build hydrogeological expertise into the inversion process.

Through use of SVD-assisted calibration (which is unique to PEST) in conjunction with parallelization of model runs, the benefits of highly parameterized inversion can now be provided to large models with high run times. Without the use of the SVD-assist scheme, this would not be possible.

D.3.7 Parameter and predictive error

Eqs. D.21 and D.27 are of the form:

$$\mathbf{k} = \mathbf{Gh} \tag{D.28}$$

where \mathbf{G} is a matrix that, as $\boldsymbol{\varepsilon}$ approaches $\mathbf{0}$, should approach a generalized inverse of \mathbf{Z} (and ideally should approach the Moore-Penrose pseudo-inverse of \mathbf{G}). See for example Koch (1988) for further details. Regularized inversion undertaken using SVD-assist can be formulated in identical terms. If reality is described by Eq. D.15 with \mathbf{k}^* (the true, but unknown set of real-world parameter values) substituted for \mathbf{k} , and if this equation is in turn substituted into Eq. D.28, we obtain:

$$\underline{\mathbf{k}} = \mathbf{GZk}^* + \mathbf{G}\boldsymbol{\varepsilon} = \mathbf{Rk}^* + \mathbf{G}\boldsymbol{\varepsilon} \quad (\text{D.29})$$

where \mathbf{R} is the so-called resolution matrix. In the absence of measurement noise, each row of this matrix expresses factors by which elements of the real-world \mathbf{k}^* are multiplied before summation (in accordance with matrix multiplication by \mathbf{R}) to yield the element of $\underline{\mathbf{k}}$ corresponding to that row. Thus, $\underline{\mathbf{k}}$ is in fact a simplification of \mathbf{k}^* , with this simplification involving averaging and integration as expressed by the rows of \mathbf{R} ; it is through this simplification that regularization occurs. It is through regularization that a unique solution to the inverse problem of highly parameterized model calibration is achieved.

From Eq. D.29, parameter error can be expressed as:

$$\underline{\mathbf{k}} - \mathbf{k}^* = -(\mathbf{I} - \mathbf{R})\mathbf{k}^* + \mathbf{G}\boldsymbol{\varepsilon} \quad (\text{D.30})$$

where \mathbf{I} is the identity matrix.

Where inversion is undertaken using SVD, the resolution matrix \mathbf{R} is given by the projection operator onto the calibration solution subspace:

$$\mathbf{R} = \mathbf{V}_1\mathbf{V}_1^t \quad (\text{D.31})$$

Parameter error can then be visualized as shown in Figure D-2. The first term in Eq. D.30 describes parameter error resulting from loss of the null space component of \mathbf{k}^* when it is projected onto the parameter solution space; this is labeled as the null space component in Figure D-2. The second term in Eq. D.2 describes error arising from the fact that the dataset \mathbf{h} from which parameters are estimated is contaminated by measurement noise $\boldsymbol{\varepsilon}$; this is labeled as the solution space component in Figure D-2. This is the only term represented where parameter uncertainty is computed by classical means - see Eq. D.7c. However it is the authors' experience that in just about all real-world contexts, parameter (and predictive) error is dominated by the null space term.

Parameter error cannot be known as \mathbf{k}^* cannot be known. However, the propensity for parameter error can be computed from Eq. D.30 using Eq. D.1 as:

$$\mathbf{C}(\underline{\mathbf{k}} - \mathbf{k}^*) = (\mathbf{I} - \mathbf{R})\mathbf{C}(\mathbf{k})(\mathbf{I} - \mathbf{R})^t + \mathbf{G}\mathbf{C}(\boldsymbol{\varepsilon})\mathbf{G}^t \quad (\text{D.32})$$

If the sensitivities of a prediction s to parameters \mathbf{k} are encapsulated in the vector \mathbf{y} , the error variance of that prediction is expressed as:

$$\sigma^2_{s-s^*} = \mathbf{y}^t(\mathbf{I} - \mathbf{R})\mathbf{C}(\mathbf{k})(\mathbf{I} - \mathbf{R})^t\mathbf{y} + \mathbf{y}^t\mathbf{G}\mathbf{C}(\boldsymbol{\varepsilon})\mathbf{G}^t\mathbf{y} \quad (\text{D.33})$$

D.3.8 Analysis of parameter and predictive error

Eqs. D.32 and D.33 can be used directly for analysis of parameter and predictive error variance. The fact that they contain only sensitivities (and quantities computed from sensitivities) brings certain advantages to their use. In particular:

1. A model does not need to be calibrated prior to use of Eqs. D.32 and D.33 in conjunction with that model; in computing all of the quantities appearing in these equations, the model can be endowed with any parameter values judged to be realistic by the modeler.
2. The contribution that different parameter types make to the overall uncertainty of any prediction can be readily computed by assuming notional perfect knowledge of those parameter types and calculating the reduction in potential for predictive error thereby accrued. This can be particularly useful in assessing the potential for error incurred by assumptions pertaining to different model boundary conditions.
3. The worth of different types of data in reducing the uncertainty of particular predictions can be appraised through including/excluding that data from the notional calibration exercise that is implicit in Eq. D.33. Because data values are not actually used in this equation (only sensitivities of model outputs corresponding to that data), Eq. D.33 can be employed to judge the comparative worth of different data acquisition strategies before investing money in implementing those strategies.

See Gallagher and Doherty (2007) and Dausman et al. (2009) for examples of the use of Eq. D.33 (and similar equations implemented in PEST utility software). Note that the USGS OPR-PPR program of Tonkin et al. (2007) purports to provide similar functionality for optimization of data acquisition. However, because it is based on classical parameter estimation and therefore ignores the (mostly dominant) null space term, it is flawed.

Nonlinear analysis of parameter and predictive uncertainty is most easily undertaken using the null space Monte Carlo (NSMC) methodology implemented in PEST. Using this methodology a suite of parameter fields is generated, each of which can be considered to calibrate the model, as it results in a model-to-measurement misfit which is acceptably low when judged in terms of $\mathbf{C}(\boldsymbol{\varepsilon})$. However each member of this suite includes a level of heterogeneity that is compatible with that which actually exists in the real world as judged in terms of the covariance matrix of innate parameter variability $\mathbf{C}(\mathbf{k})$. By making a particular model prediction with all of these fields, the uncertainty associated with that prediction can be assessed.

Generation of calibration-constrained parameter fields has historically been an extremely computer-intensive process. However implementation of the NSMC methodology reduces this burden considerably. The algorithmic basis of this technique lies in Eq. D.32; further details are provided in Tonkin and Doherty (2009) and in PEST documentation. The process is diagrammatically illustrated in Figure D-3.

D.4 Summary

The contents of this section are now summarized diagrammatically.

Figure D-4 provides a pictorial demonstration of the situation confronting a modeler prior to calibration of a model. A heterogeneous hydraulic property field within a model domain is represented in that figure. Historical head measurements are available from a number of wells within that domain.

In traditional model calibration, as is illustrated in Figure D-5, whether it be manual or computer-based, the model domain is subdivided into zones of piecewise constancy. Sometimes such subdivision is guided by geological mapping. On other occasions it is done in an ad-hoc manner to accommodate intra-geological-unit heterogeneity that becomes apparent during the calibration process, as an acceptable fit between model outputs and field measurements could not be achieved without it.

Alternatively, a highly parameterized approach could be taken to calibration. In Figure D-6, hydraulic conductivity is representing using pilot points. There is no need for manual pre-calibration parameter parsimonization in this case, for the simplification necessary for attainment of a unique solution to the inverse problem of model calibration is implemented as part of the calibration process itself. However this simplification expresses itself in a smooth, rather than piece-wise constant, parameter field. This field expresses as much of the underlying hydraulic property heterogeneity as it is capable of doing, given the information content of data on which calibration is based. The use of many pilot points provides the calibration process with the flexibility that it needs to transfer this information from the calibration dataset to the parameter field employed by the model.

Mathematical regularization falls into two broad groups, namely Tikhonov and subspace regularization. Tikhonov regularization promulgates the estimation of many parameters by providing a fallback position for those parameter which, either individually or in various combinations, cannot be estimated. Parameters are allowed to vary from this fallback position only to the minimum extent necessary to achieve a desired level of fit between model outcomes and historical field measurements.

Subspace regularization takes the opposite approach; individual parameters, and/or combinations of parameters which are discovered to be inestimable (because they lie in the null subspace of parameter space) are removed from the parameter estimation process, leaving only those parameter combination for which estimation is possible on the basis of the current calibration dataset. Singular value decomposition is the most numerically accurate way to accomplish this task. However where a large number of parameters are featured in the calibration process, LSQR can provide a more efficient means of achieving nearly the same thing.

Ideally, both Tikhonov and subspace methods should be used in concert. However the use of subspace methods (either alone or in concert) presents the modeler with a means through which the computational efficiency of the overall parameter estimation process can be increased dramatically (sometimes with a cost in performance for highly nonlinear models). This is

illustrated in Figure D-8. A modeler has the choice of whether to compute derivatives with respect to ALL model parameters during every iteration of the parameter estimation process, or simply with respect to a limited number super parameters which span the solution subspace. The latter option comprises SVD-Assist functionality available through PEST.

Once a calibrated parameter field has been obtained, subspace methods can be further employed to find many more parameter fields, all of which calibrate the model as well as the original parameter field obtained through regularized inversion, but all of which encompass, in a stochastic sense, hydraulic property heterogeneity at a level which may exist within the subsurface; see Figure D-9. By making a model prediction with all of these parameter fields, the uncertainty associated with that prediction can be explored.

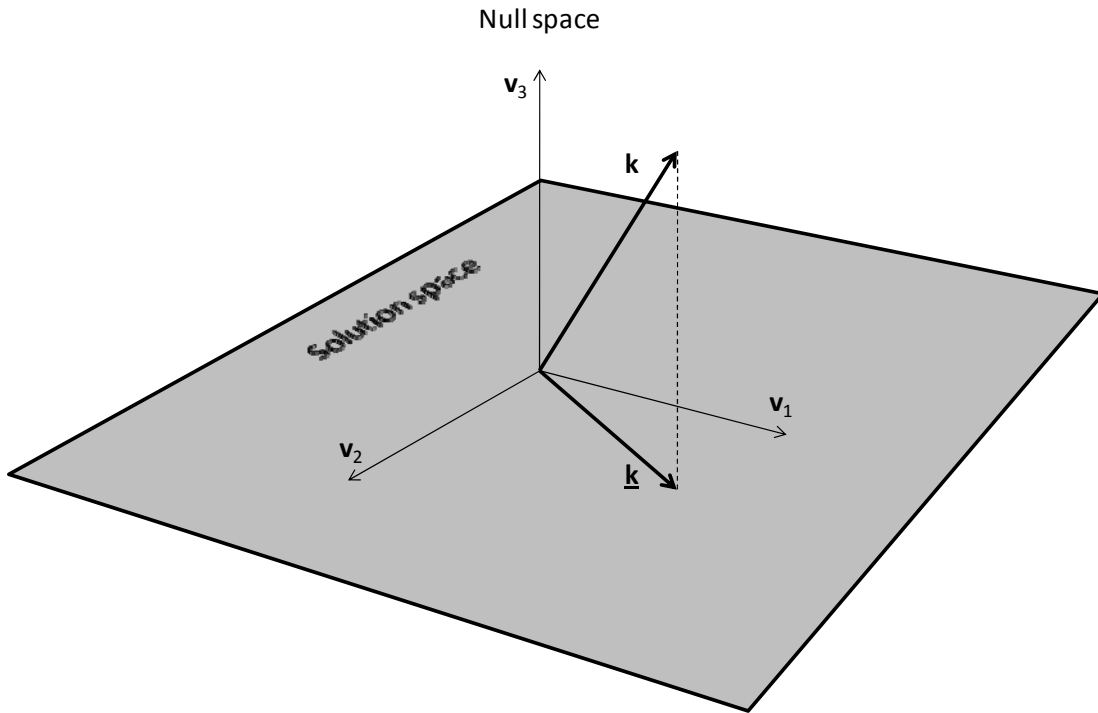


Figure D-1. Parameter estimation as a projection operation.

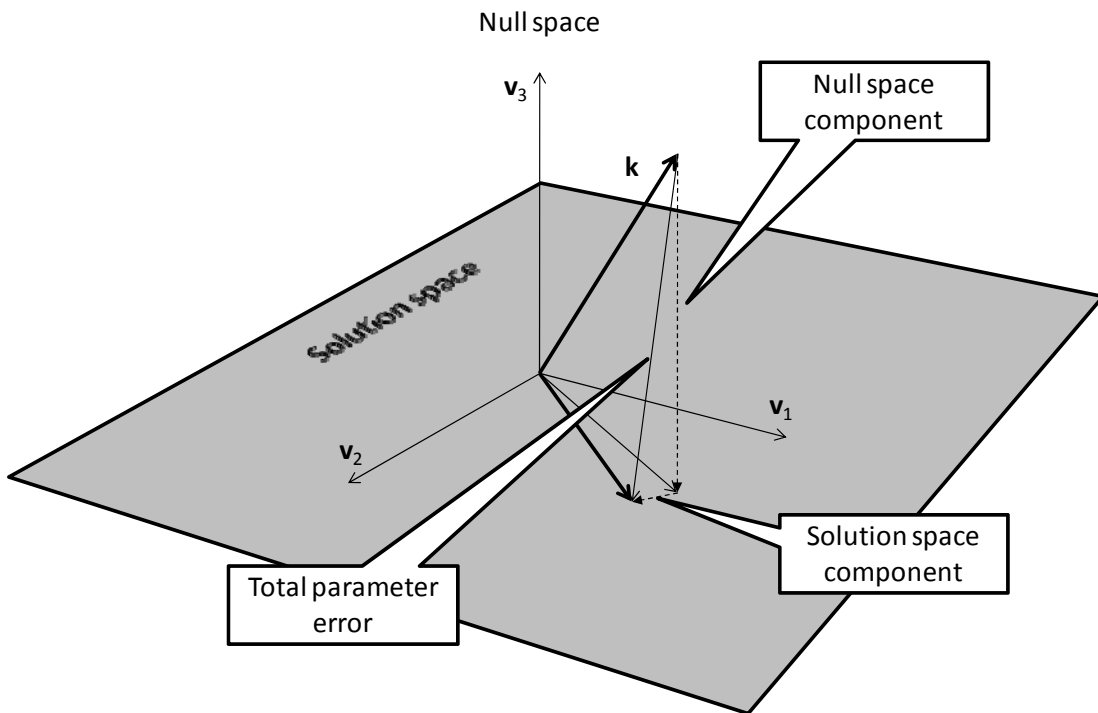


Figure D-2. Components of parameter error.

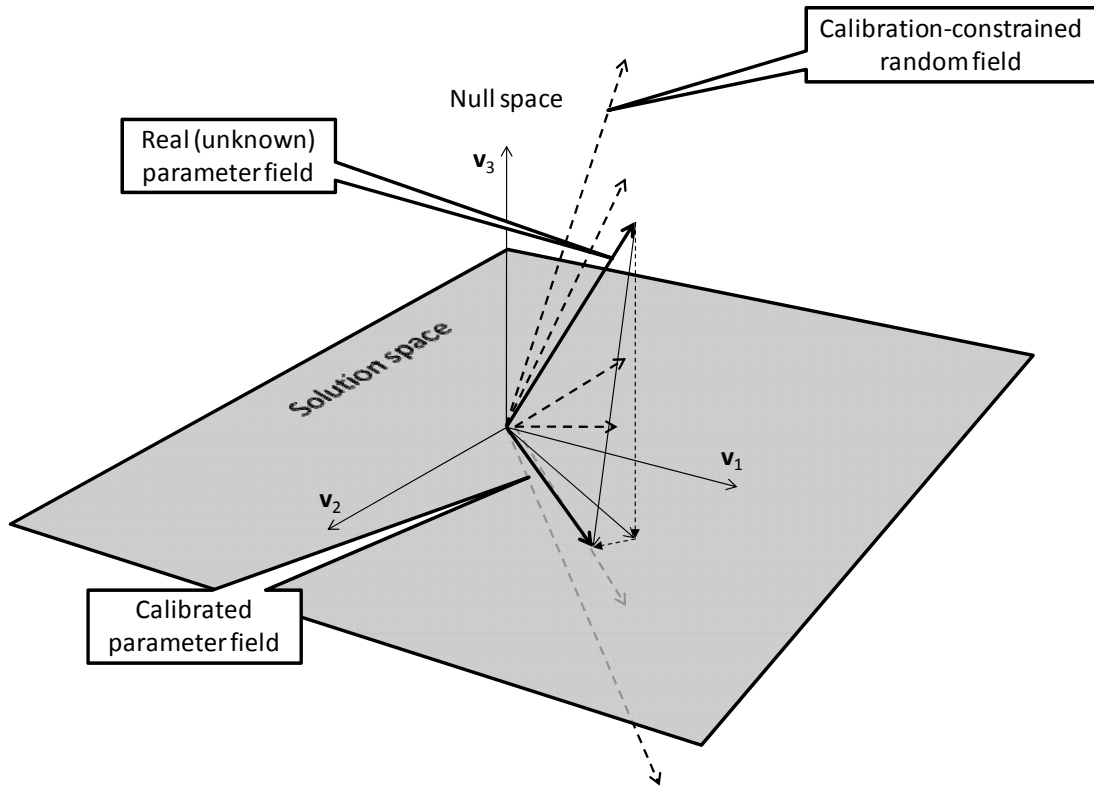


Figure D-3. Null space Monte Carlo generation of calibration constrained random fields.

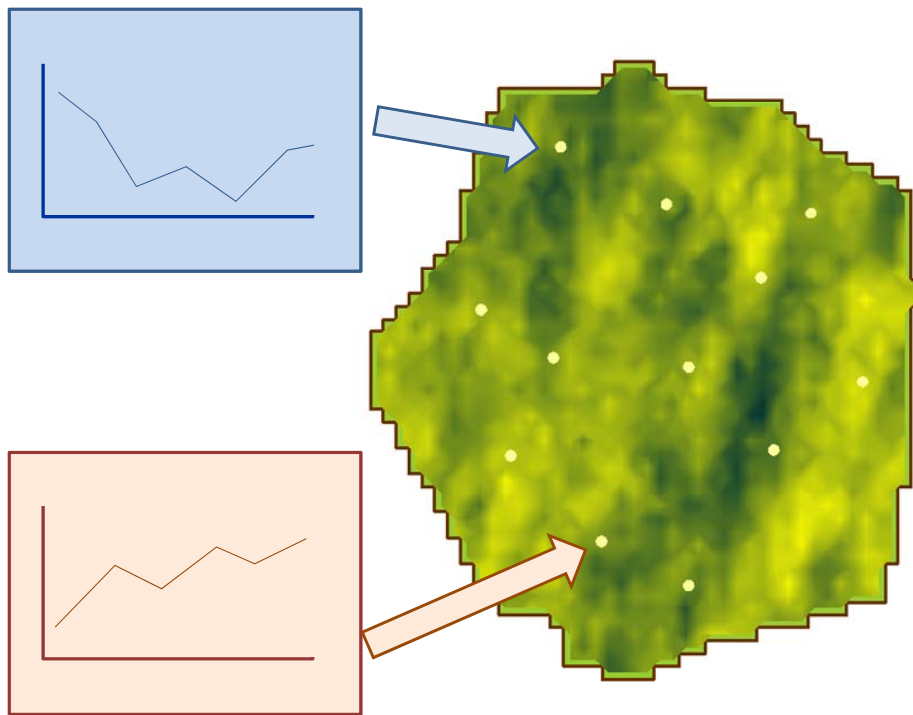


Figure D-4. Heterogeneous hydraulic property field within a model domain. Historical head measurements are available at wells located within the domain.

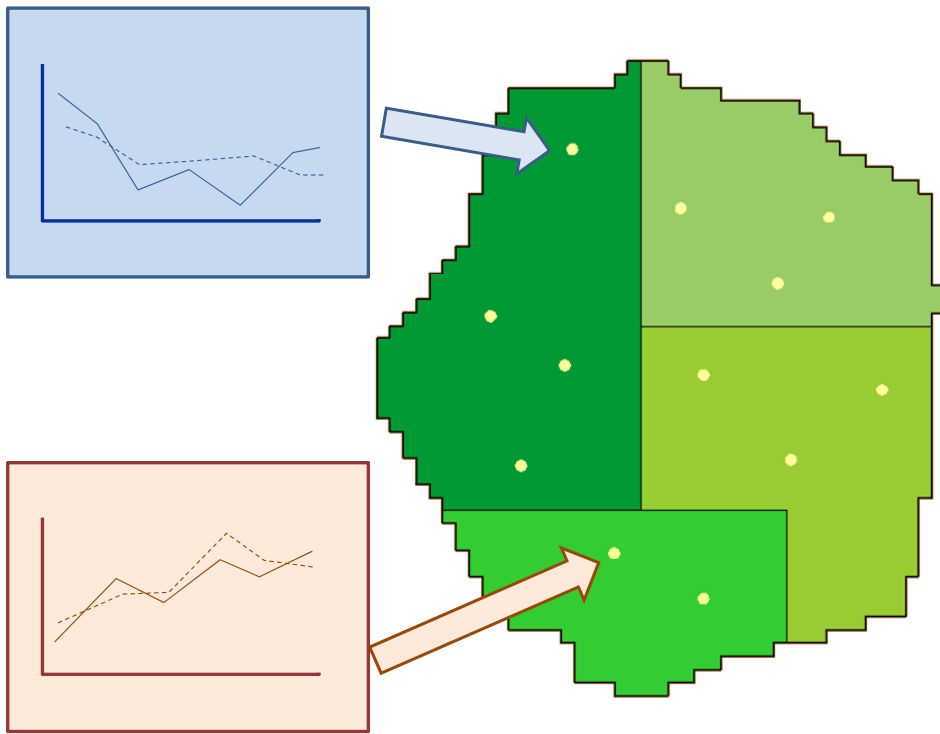


Figure D-5. Calibration based on zones of piecewise constancy.

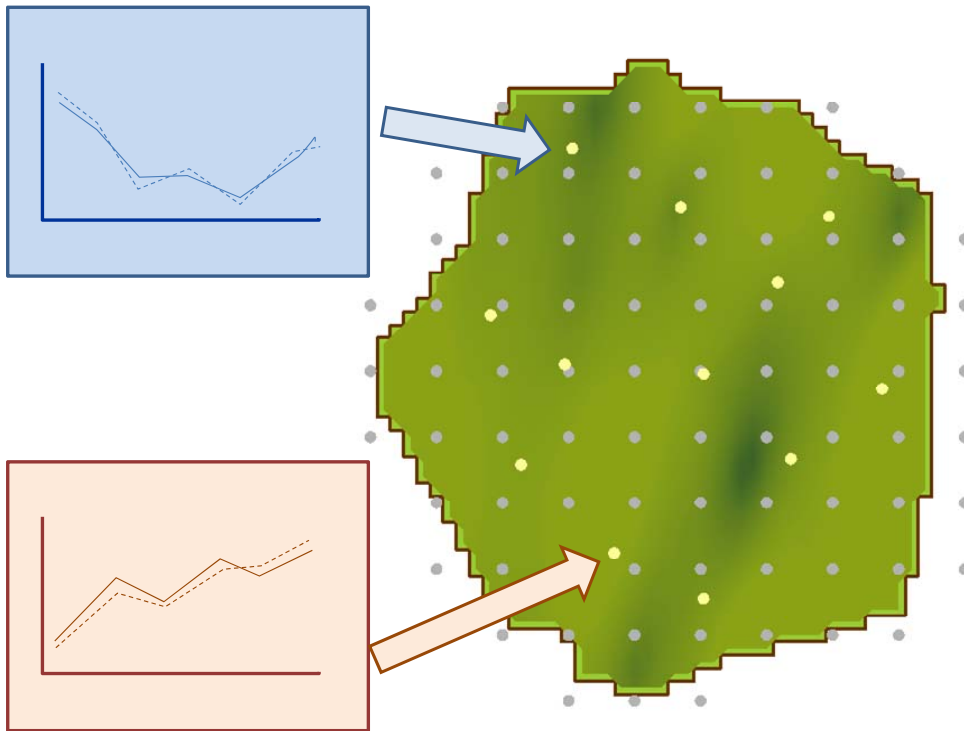


Figure D-6. Pilot-point-based calibration. Pilot points are shown in grey.

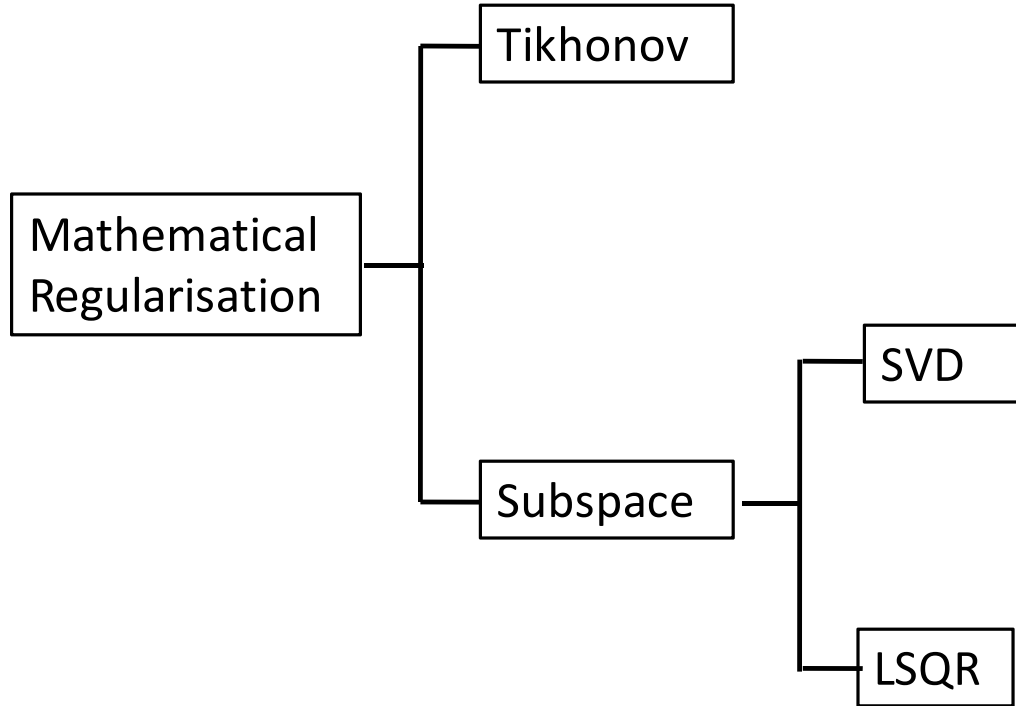


Figure D-7. Types of mathematical regularization used by PEST.

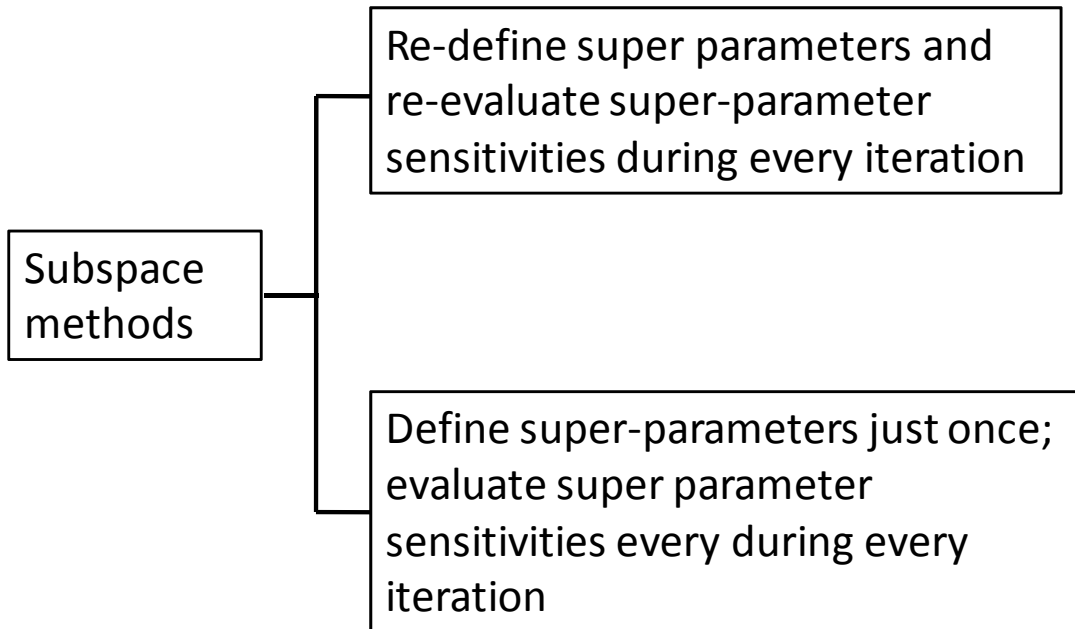


Figure D-8. Use of super parameters to enhance the efficiency of the parameter estimation process.

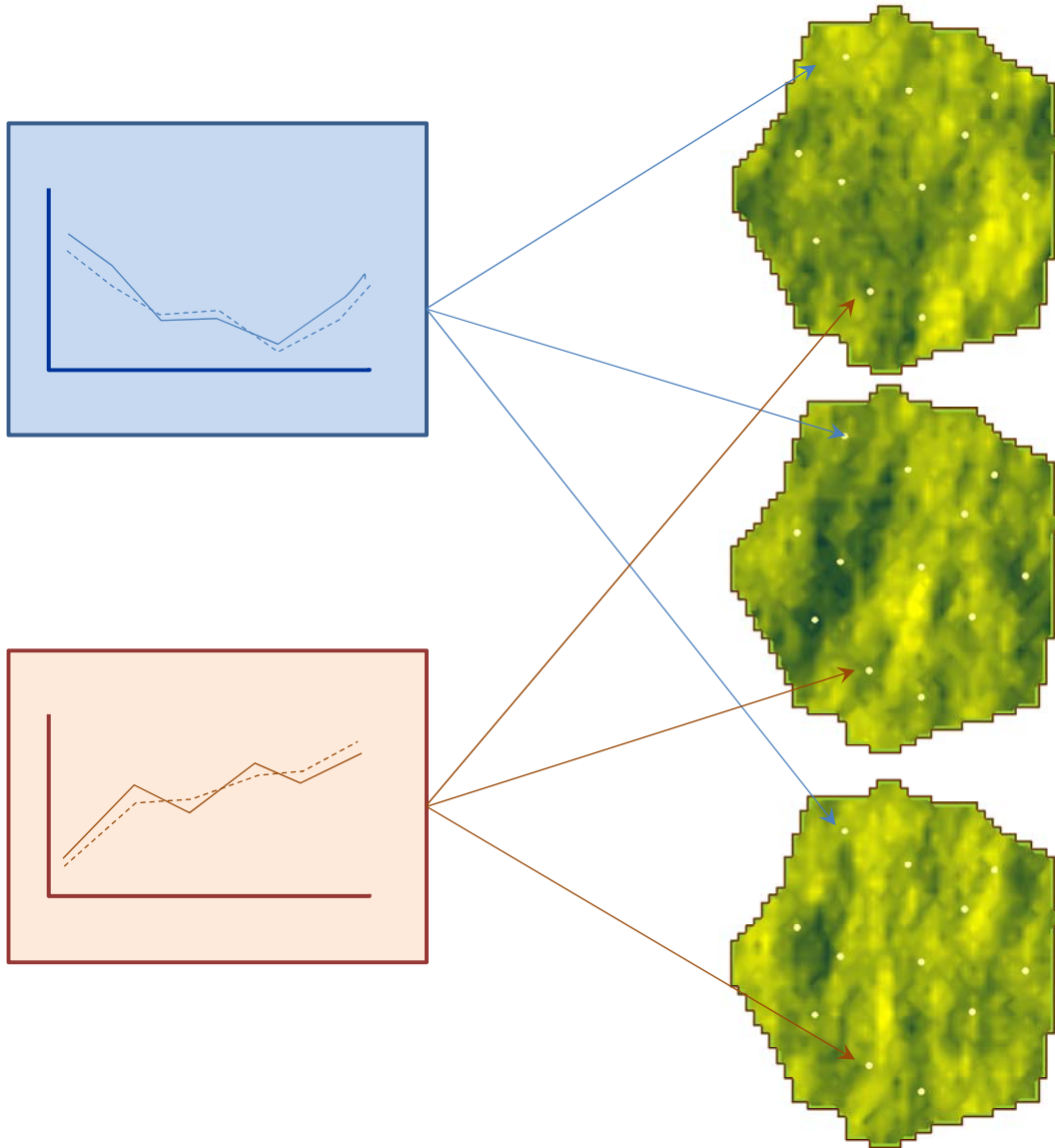


Figure D-9. The null-space Monte Carlo methodology can be used to compute many different parameter fields which all calibrate the model.

TWDB Report ##: Final – Application of PEST to Re-Calibrate
the Groundwater Availability Model for the Edwards-Trinity (Plateau) and Pecos Valley Aquifers

Appendix E File details of ETPV GAM

E.1 MODFLOW input and output files

Input files used by original model.

File	Purpose
eddt_p_tr.nam	MODFLOW name file.
bas.dat	Basic package input file.
bcf.dat	Block centered flow package input file.
wel.dat	Well package input file.
drn.dat	Drain package input file.
ghb.dat	General head boundary package input file.
rch.dat	Recharge package input file.
str1.dat	Streamflow routing package input file.
pcg2.dat	Preconditioned conjugate gradient package input file.
oc.dat	Output control package input file.

Output files generated by original model.

File	Purpose
output.dat	MODFLOW list output file.
heads.dat	Binary heads output file.
ddown.dat	Binary drawdown output file.
budget.dat	Binary cell-by-cell flow term output file.

Input files used by steady-state component of revised model.

File	Purpose
eddt_p_ss.nam	MODFLOW name file.
bas_ss.dat	Basic package input file.
bcf_ss.dat	Block centered flow package input file.
drn.dat	Drain package input file.
ghb.dat	General head boundary package input file.
rch_ss.dat	Recharge package input file.
str1.dat	Streamflow routing package input file.
pcg2_ss.dat	Preconditioned conjugate gradient package input file.
oc_ss.dat	Output control package input file.
trans[1-3].ref	Transmissivity in layers 1 to 3.
vcont[1-2].ref	Vertical conductance underlying layers 1 and 2.
top[1-3].ref	Elevation of top of layers 1 to 3.
rech_ss.ref	Recharge.

Output files generated by the steady-state component of the revised model.

File	Purpose
output_ss.dat	MODFLOW list output file.
heads_ss.dat	Binary heads output file.
ddown_ss.dat	Binary drawdown output file.
budget_ss.dat	Binary cell-by-cell flow term output file.

Input files used by transient component of revised model.

File	Purpose
eddt_p_tr.nam	MODFLOW name file.
bas_tr.dat	Basic package input file.
bcf_tr.dat	Block centered flow package input file.
wel_tr.dat	Well package input file.
drn.dat	Drain package input file.
ghb.dat	General head boundary package input file.
rch_tr.dat	Recharge package input file.
str1.dat	Streamflow routing package input file.
pcg2_tr.dat	Preconditioned conjugate gradient package input file.
oc_tr.dat	Output control package input file.
sh[1-3].ref	Initial heads for layers 1 to 3.
trans[1-3].ref	Transmissivity for layers 1 to 3.
vcont[1-2].ref	Vertical conductance beneath layers 1 and 2.
s[1-3].ref	Confined storage coefficient in layers 1 to 3.
sy[1-3].ref	Specific yield for layers 1 to 3.
top[1-3]	Elevation of top of layers 1 to 3.
rech_av_19[30-70]s.ref	10-year average recharge over nominated decade.
rech_19[80-99].ref	Recharge over nominated year.

Output files generated by the transient component of the revised model.

File	Purpose
output_tr.dat	MODFLOW list output file.
heads_tr.dat	Binary heads output file.
drown_tr.dat	Binary drawdown output file.
budget_tr.dat	Binary cell-by-cell flow term output file.

E.2 Files employed in processing model outputs used in calibration process

Observation-related files.

File	Purpose
measure_cnt_gt_6b.crd	Contains coordinates and layer numbers of wells from which yearly averaged heads were employed in the calibration process.
measure_cnt_gt_6b.smp	Yearly averaged heads used in the calibration process.
model_cnt_gt_6b.smp	MOD2OBS-generated counterpart to <i>measure_cnt_gt_6b.smp</i> . The MOD2OBS utility is run as part of the model.
model_cnt_gt_6b.ins	Contains PEST instructions to read <i>model_cnt_gt_6b.smp</i> .
measure_cnt_gt_6b_diff.smp	Temporal differences between observed flows computed using the SMPDIFF utility.
model_cnt_gt_6b_diff.smp	Model-generated equivalent of <i>measure_cnt_gt_6b_diff.smp</i> .
model_cnt_gt_6b_diff.ins	Contains PEST instructions to read the <i>model_cnt_gt_6b_diff.smp</i> .
observ_flows.smp	Yearly averaged stream flows; these can be compared with model outputs, but were not employed in the calibration process.
model_flows.smp	Model stream flow outputs at measurement points prior to temporal interpolation to measurement times. These are extracted from <i>budget_tr.dat</i> by the STRFLBUD utility.

File	Purpose
mo_flows.smp	Model-generated stream flows after temporal interpolation from data contained in <i>model_flows.smp</i> . Interpolation is carried out by the SMP2SMP utility. This file is the model-generated equivalent of <i>observ_flows.smp</i> .
mo_flows.ins	Contains PEST instructions to read the above file.

E.3 Files used in hydraulic property parameterization of revised model

Files used in transmissivity parameterization (presented roughly in order of appearance in processing sequence).

File	Purpose
kzone[1-2].inf	Integer arrays defining hydraulic conductivity zonation.
hk_lay[1-2].irc	Integer-real-correspondence files through which horizontal hydraulic conductivities are assigned to zones. PEST writes these files on the basis of templates of them prior to each model run.
hk_lay[1-2].tpl	PEST templates of the above files.
hk_lay[1-2]_zone.ref	Real-array files written by INT2REAL containing zone-based horizontal hydraulic conductivities.
hk_lay[1-2]_mul.dat	Pilot points files containing hydraulic conductivity multiplier values for layers 1 and 2. These are written by PEST on the basis of templates of these files prior to every model run.
hk_lay[1-2]_mul.tpl	PEST templates of the above files.
factors_lay[1-2].ref	Kriging factors for interpolation from pilot points to active cells comprising layers 1 and 2 of the model grid. These were written by PPK2FAC1 and are used by FAC2REAL.
hk_lay[1-2]_mul.ref	Hydraulic conductivity multiplier arrays written by FAC2REAL during each composite model run.
hk_lay[1-2].ref	Hydraulic conductivity arrays generated by multiplying <i>hk_lay[1-2]_zone.ref</i> by <i>hk_lay[1-2]_mul.ref</i> .
hk_lay[1-2].lim	Upper and lower horizontal hydraulic conductivity limits for layers 1 and 2 imposed by the CLIPARRAY1 utility.
thick[1-2].ref	Thickness of layers 1 and 2.
trans[1-2].ref	Transmissivity of layers 1 and 2 generated by multiplying <i>hk_lay[1-2].ref</i> by <i>thick[1-2].ref</i> .

Files used in primary storage coefficient parameterization (presented roughly in order of appearance in processing sequence).

File	Purpose
kzone[1-2].inf	Integer arrays defining specific storage zonation.
s_lay[1-2].irc	Integer-real-correspondence files through which specific storages are assigned to zones. PEST writes these files on the basis of templates of them prior to each model run.
s_lay[1-2].tpl	PEST templates of the above files.
s_lay[1-2]_zone.ref	Real-array files written by INT2REAL containing zone-based storage coefficients.
s_lay[1-2]_mul.dat	Pilot points files containing storage coefficient multiplier values for layers 1 and 2. These are written by PEST on the basis of templates of these files prior to every model run.
s_lay[1-2]_mul.tpl	PEST templates of the above files.

File	Purpose
factors_lay[1-2].ref	Kriging factors for interpolation from pilot points to active cells comprising layers 1 and 2 of the model grid. These were written by PPK2FAC1 and are used by FAC2REAL.
s_lay[1-2]_mul.ref	Storage coefficient multiplier arrays written by FAC2REAL during each composite model run.
s_lay[1-2].ref	Storage coefficient arrays generated by multiplying <i>s_lay[1-2]zone.ref</i> by <i>s_lay[1-2]_mul.ref</i> .
thick[1-2].ref	Thickness of layers 1 and 2.
s[1-2].ref	Primary storage factor for layers 1 and 2 generated by multiplying <i>s_lay[1-2].ref</i> by <i>thick[1-2].ref</i> .

Files used in specific yield parameterization (presented roughly in order of appearance in processing sequence).

File	Purpose
kzone[1-2].inf	Integer arrays defining specific yield zonation.
sy_lay[1-2].irc	Integer-real-correspondence files through which specific yield values are assigned to zones. PEST writes these files on the basis of templates of them prior to each model run.
sy_lay[1-2].tpl	PEST templates of the above files.
sy_lay[1-2]_zone.ref	Real-array files written by INT2REAL containing zone-based specific yield values.
sy_lay[1-2]_mul.dat	Pilot points files containing specific yield multiplier values for layers 1 and 2. These are written by PEST on the basis of templates of these files prior to every model run.
sy_lay[1-2]_mul.tpl	PEST templates of the above files.
factors_lay[1-2].ref	Kriging factors for interpolation from pilot points to active cells comprising layers 1 and 2 of the model grid. These were written by PPK2FAC1 and are used by FAC2REAL.
sy_lay[1-2]_mul.ref	Specific yield multiplier arrays written by FAC2REAL during each composite model run.
sy [1-2].ref	MODFLOW specific yield arrays generated by multiplying <i>s_lay[1-2]zone.ref</i> by <i>s_lay[1-2]_mul.ref</i> .
sy[1-2].lim	Upper and lower specific yield limits for layers 1 and 2 imposed by the CLIPARRAY1 utility.

Files used in inter-layer vertical conductance parameterization (presented roughly in order of appearance in processing sequence).

File	Purpose
vcontcalc.in	File containing keyboard response to VCONTCALC prompts. One of these responses is the global vertical anisotropy. PEST is able to adjust this parameter through use of a template of <i>vcontcalc.in</i> .
vcontcalc.tpl	PEST template of the above file.
vcont_lay1_mul.dat	Pilot points files containing vertical conductivity multiplier values. This file is written by PEST on the basis of a template of it prior to every model run.
vcont_lay1_mul.tpl	PEST template of the above file.
factors_lay1.ref	Kriging factors for interpolation from pilot points to active cells of the model grid. These were written by PPK2FAC1 and are used by FAC2REAL.

File	Purpose
vcont_lay1_mul.ref	Vertical conductance multiplier array written by FAC2REAL during each composite model run.
vcont1.ref	MODFLOW vertical conductance array between layers 1 and 2. This is generated by TWOARRAY by multiplying <i>vcont_lay1_zone.ref</i> by <i>vcont_lay1_mul.ref</i> .

Files used in recharge parameterization (presented roughly in order of appearance in processing sequence).

File	Purpose
rainfall_ss.ref	Real array of average rainfall over model domain.
rainfall_av_19xxs.ref	Real array of ten year average rainfall recharge over nominated decade.
rainfall_19[80-99].ref	Real array of rainfall for nominated year over model domain.
rechzones.inf	Integer array used for zoned component of recharge parameterization.
rf.irc	Integer-real-correspondence file containing recharge factors pertaining to each recharge zone.
rf.tpl	PEST template of above file.
rechmul.dat	Pilot points file containing recharge multipliers.
rechmul.tpl	Template of the above file.
factors_all.dat	Kriging factors through which interpolation takes place from recharge factor pilot points to the model grid using the FAC2REAL utility.
rechmul.ref	Real array of pilot-point-based recharge multipliers.
rechmul_total.ref	Real array of total recharge multipliers (obtained by multiplying pilot point based multipliers with zoned multipliers on a cell-by-cell basis).
rech_ss.ref	Average recharge. This is used by the steady-state MODFLOW model component.
rech_av_19xxs.ref	Ten year average recharge over nominated decade. This is used by the transient MODFLOW model component.
rech_19[80-99].ref	Recharge over nominated year. This is used by the transient MODFLOW model component.

Files used in general head boundary conductance parameterization (presented roughly in order of appearance in processing sequence).

File	Purpose
ghb_lay[1-2].pts	Pilot points used for general head boundary conductance parameterization in layers 1 and 2.
ghb_lay[1-2].tpl	PEST templates of the above files.
factors_ghb_lay[1-2].dat	Kriging factors through which interpolation takes place from pilot points to real arrays representing general head boundary conductances in each of layers 1 and 2.
ghb_lay[1-2].ref	Real arrays for each of layers 1 and 2 holding general head boundary conductances.
ghb.dat.keep	General head boundary input file of the original ETPV GAM. This is read by the REAL2CND utility.
ghb.dat	General head boundary input file read by both the steady state and transient components of the ETPV GAM. This is written by the REAL2CND utility.

Files used in stream conductance parameterization (presented roughly in order of appearance in processing sequence).

File	Purpose
strcond.pts	Pilot points used for streambed conductance parameterization.
strcond.tpl	PEST template of the above file.
factors_str1.dat	Kriging factors through which interpolation takes place from pilot points to real array representing streambed conductance.
strcond.ref	Real array holding streambed conductances.
realstr2cond.in	Keyboard input file for the REALSTR2COND utility which replaces conductances in an existing streamflow routing package input file with those in a real array. Part of keyboard input is a global roughness coefficient to apply to all stream reaches. This parameter is adjustable by PEST.
realstr2cond.tpl	PEST template of the above file.
str1.dat.keep	Streamflow routing package input file of the original ETPV GAM. This is read by the REALSTR2COND utility.
str1.dat	Streamflow routing package input file read by both the steady state and transient components of the ETPV GAM. This is written by the REALSTR2COND utility.

Files used in drain conductance parameterization (presented roughly in order of appearance in processing sequence).

File	Purpose
drain_lay[1-2].inf	Integer array files used to represent drain conductance zonation in layers 1 and 2.
draincond_lay[1-2].irc	Integer-real correspondence files linking zone numbers to drain conductance values. PEST adjusts conductance values cited in these files.
draincond_lay[1-2].tpl	PEST templates of the above files.
draincond_lay[1-2].ref	Real arrays for each of layers 1 and 2 (written by the REAL2COND utility) holding drain conductances.
drn.dat.keep	MODFLOW drain conductance input file for the original ETPV GAM. Conductances within this file are replaced by those in <i>draincond_lay[1-2].ref</i> by the REAL2COND utility.
drn.dat	MODFLOW drain package input file for both the steady state and transient components of the new ETPV GAM.

Files used in well pumping parameterization (presented roughly in order of appearance in processing sequence).

File	Purpose
welmul_lay[1-2].dat	Files containing pilot point well pumping rate multipliers for layers 1 and 2. These are estimated by PEST (and are assigned a preferred value of 1.0 using Tikhonov constraints).
welmul_lay[1-2].tpl	PEST templates of these files.
factors_lay[1-2].dat	Kriging factors for implementation of interpolation from well multiplier pilot points to the MODFLOW grid.
welmul_lay[1-2].ref	Well multiplier arrays interpolated from pilot points for layers 1 and 2.
standard.wel	Well pumping rate for use during build-up stress periods after multiplication by build-up factors and pilot point based well multipliers.
buildup_zones.inf	Integer array file containing zones on which well pumping rate build-up factors are based.

File	Purpose
buildup_factors.dat	A file containing build-up factors for each zone for each build-up stress period. These build-up factors are estimated by PEST.
buildup_factors.tpl	PEST template of the above file.
post-buildup.wel	Part of a MODFLOW well package input file containing well pumping rates for all stress periods but the build-up stress periods.
wel_tr.dat	MODFLOW well package input file for the transient component of the ETPV GAM.

Files used in implementing median value regularization on estimated horizontal hydraulic conductivity arrays (presented roughly in order of appearance in processing sequence).

File	Purpose
hk_lay[1-2].ref	Hydraulic conductivity arrays used for calculation of transmissivity for MODFLOW.
median_hk_lay[1-2].ref	Arrays of median hydraulic conductivity computed from pumping test analyses within respective zones. Medians are zone-specific.
ratio_median_hk_lay[1-2].ref	Ratio of <i>hk_lay[1-2].ref</i> to <i>median_hk_lay[1-2].ref</i> computed using the TWOARRAY utility.
log_ratio_median_hk_lay[1-2].ref	Log of the above arrays on a cell-by-cell basis as computed by the LOGARRAY utility.
log_ratio_median_hk_lay[1-2].dat	Outcomes of interpolation of the above arrays to the locations of pilot points within each of layers 1 and 2. Interpolation from arrays to pilot points is undertaken using the LOGARRAY utility. These pilot-point-based log ratios are read by PEST and observed to be zero.
log_ratio_median_hk_lay[1-2].ins	Instruction files to read the above model output files.

E.4 Executable programs used by composite model

Executable programs used by composite model (presented roughly in order of their appearance in the model batch file).

File	Purpose
model_gt6b.bat	Batch file run by PEST as composite model.
int2real.exe	Builds a MODFLOW/MT3D-compatible real array based on the contents of a MODFLOW/MT3D-compatible integer array. (This is a member of the Groundwater Data Utilities suite.)
vcontcalc	Computes vertical conductance on a cell-by-cell basis between layers from layer hydraulic conductivity and thickness.
fac2real	Uses PPKFAC or PPK2FAC1-generated kriging factors to produce a MODFLOW-compatible real array on the basis of spatial interpolation from a set of pilot points.(This is a member of the Groundwater Data Utility suite.)
twoarray.exe	Combines two real arrays by addition, subtraction, multiplication, division and partial replacement. (This is a member of the Groundwater Data Utilities suite.)
cliparray1.exe	Limits values occurring within a MODFLOW or MT3D real array to within user-specified, zone-specific limits.
real2cnd	Replaces conductances cited in an existing MODFLOW drain or general head boundary input file by those recorded in layer-specific real arrays.
realstr2cnd	Replaces streambed conductances cited in an existing MODFLOW streamflow routing package input file by those recorded in layer-specific real arrays.

TWDB Report ##: Final – Application of PEST to Re-Calibrate
the Groundwater Availability Model for the Edwards-Trinity (Plateau) and Pecos Valley Aquifers

File	Purpose
wellbuild2	Builds a well package input file for the ETPV GAM using build-up factors, multiplier arrays and user-supplied standard pumping rates.
mf96_d.exe	Double precision version of USGS MODFLOW96 model.
many2one_d.exe	Double precision version of MANY2ONE program of Groundwater Data Utilities suite. Splits MODFLOW/MT3D-generated binary files comprised of multiple two-dimensional results arrays into individual ASCII/binary files.
mod2obs_d.exe	Double precision version of the MOD2OBS program of the Groundwater Data Utilities suite. Interpolates MODFLOW/MT3D-generated data to the same times and locations as those cited in a user-supplied bore sample file; writes another bore sample file.
smpdiff.exe	A member of the Groundwater Data Utilities suite. Writes a new bore sample file in which differences are taken between successive values in an existing bore sample file, or between values in an existing file and a reference value. (This is a member of the Groundwater Data Utilities suite.)
smp2smp	Interpolates data contained within one site sample file to the dates and times represented in another site sample file. (This is a member of the Groundwater Data Utilities suite.)
strflbud.exe	Extracts flows at user-specified cells from a binary MODFLOW cell-by-cell flow term file. This program was written as part of the current project.
logarray.exe	Evaluates the log (to base 10) of all elements of a real array.
arr2bore.exe	Undertakes spatial interpolation from a single array to a set of points.

Appendix F Listing of wells used for the recalibration

Listed below are the locations of the 574 wells used for the recalibration of the ETPV GAM.

Well ID	Easting (ft)	Northing (ft)	Layer	County	Number of measurements
6817302	5268857.37	19128415.98	2	Bandera	2
6817402	5244382.63	19119175.55	2	Bandera	5
6817501	5261824.23	19118429.74	2	Bandera	11
6817601	5269001.79	19113844.52	2	Bandera	3
6817709	5248464.03	19096949.53	2	Bandera	2
6817801	5254981.94	19106422.44	2	Bandera	8
6817901	5268483.56	19095016.85	2	Bandera	7
6825203	5258342.11	19086721.22	2	Bandera	10
6825210	5256157.13	19084777.54	2	Bandera	1
6825502	5251941.63	19073808.78	2	Bandera	7
6825503	5262949.84	19075128.32	2	Bandera	6
6825504	5262860.63	19075228.73	2	Bandera	6
6825505	5262864.56	19074824.04	2	Bandera	7
6825507	5252482.14	19072599.69	2	Bandera	7
6905701	5088649.59	19187511.46	1	Bandera	6
6912103	5050788.38	19184010.94	1	Bandera	3
6912302	5079061.84	19170059.3	2	Bandera	8
6912501	5055886.51	19160955.46	2	Bandera	5
6912801	5059565.61	19142247.75	2	Bandera	6
6913101	5082107.71	19176854.27	2	Bandera	7
6914501	5139099.9	19157629.74	2	Bandera	13
6914608	5157384.24	19159063.8	2	Bandera	1
6914801	5143627.73	19150978.5	2	Bandera	1
6915401	5160377.98	19158578.04	2	Bandera	13
6915402	5167335.48	19157209.3	2	Bandera	10
6915501	5175490.39	19160607.41	2	Bandera	11
6915801	5178972.86	19141505.34	2	Bandera	11
6916401	5202413.07	19159193.6	2	Bandera	2
6916801	5224147.14	19147837.23	2	Bandera	9
6916904	5233795.05	19151160.52	2	Bandera	4
6920901	5078926.37	19102356.32	2	Bandera	11
6922901	5150524.76	19105179.81	2	Bandera	16
6923301	5191507.97	19124901.33	2	Bandera	4
6923603	5186187.63	19118282.66	2	Bandera	14
6924102	5210795.12	19134669.36	2	Bandera	4
6924103	5207294.75	19132211.68	2	Bandera	6
6924202	5214774.85	19132374.77	2	Bandera	7
6924203	5214774.85	19132374.77	2	Bandera	3
6924208	5223191.7	19125969.66	2	Bandera	6
6924504	5218730.52	19122591.82	2	Bandera	4
6819208	5343572.13	19130450.8	2	Bexar	17
6819612	5346682.57	19113486.15	2	Bexar	4
6819618	5347921.58	19112995.14	2	Bexar	7
6819806	5337972.6	19097898.02	2	Bexar	9
6819902	5346467.67	19102149.11	2	Bexar	5
6820110	5357252.17	19134666.38	2	Bexar	7
6820601	5395992.09	19114321.26	2	Bexar	15
6820602	5389883.95	19116667.79	2	Bexar	9

TWDB Report ##: Final – Application of PEST to Re-Calibrate
the Groundwater Availability Model for the Edwards-Trinity (Plateau) and Pecos Valley Aquifers

Well ID	Easting (ft)	Northing (ft)	Layer	County	Number of measurements
6820603	5389883.95	19116667.79	2	Bexar	1
6827501	5336868.95	19079467.31	2	Bexar	3
5745910	5430580.86	19327069.19	2	Blanco	11
5752804	5374319.74	19291266.23	2	Blanco	8
5753305	5431997.55	19320003.45	2	Blanco	14
5753614	5427416.6	19297765.93	2	Blanco	1
5754905	5464867.01	19284055.35	2	Blanco	9
5755101	5476058.16	19313386.95	2	Blanco	8
5761101	5402591.68	19274536.8	2	Blanco	2
5761507	5412132.75	19263129	2	Blanco	8
5761803	5420191.61	19233078.02	2	Blanco	14
5762106	5446487.72	19276383.65	2	Blanco	5
5762413	5444533.32	19248820.13	2	Blanco	2
5230105	3864716.27	19468591.98	1	Brewster	6
6805605	5431294.64	19206009.11	2	Comal	7
6806701	5442464.73	19199189.75	2	Comal	8
6807407	5487000	19210813.32	2	Comal	3
6812302	5394119.07	19181296.52	2	Comal	3
6813806	5409958.02	19148619.75	2	Comal	3
6814407	5441995.72	19159809.79	2	Comal	4
6815115	5482241.95	19178852.22	2	Comal	4
6815116	5486726	19178823.04	2	Comal	2
4242710	4969843.3	19687261.88	1	Concho	7
4242904	4992620.57	19689022.81	2	Concho	13
4250102	4967868.54	19674297.41	2	Concho	9
4257301	4949188.37	19629116.55	1	Concho	1
4527901	4094360.69	19800856.71	1	Crane	2
4528701	4115382.06	19787392.75	1	Crane	16
4459810	4390158.73	19603209.83	2	Crockett	5
4461403	4464647.87	19620981.79	2	Crockett	7
4464804	4586583.81	19601663.19	1	Crockett	1
5401607	4324470.54	19576431.68	2	Crockett	12
5403505	4390378.3	19577889.87	1	Crockett	3
5405404	4454564.28	19568358	1	Crockett	6
5410806	4355195.96	19520423.22	2	Crockett	4
5411501	4388609.33	19530627.37	2	Crockett	16
5411502	4389566.59	19530612.68	2	Crockett	18
5413405	4455070.85	19528455.03	2	Crockett	3
5413707	4453105.24	19511774.18	1	Crockett	6
5415303	4558100	19535424.48	1	Crockett	1
5415304	4558445.84	19535218.18	1	Crockett	12
5416402	4574972.07	19525936.28	1	Crockett	6
5421506	4466865.64	19485365.99	1	Crockett	6
5423107	4542127.65	19491043.96	1	Crockett	14
5424206	4593381.67	19503178.22	1	Crockett	6
5424402	4577838.48	19481357.1	1	Crockett	16
5428405	4415918.77	19445568.44	2	Crockett	4
5431504	4548681.7	19435994.71	2	Crockett	3
5436903	4439748.91	19377091.14	1	Crockett	5
5437408	4455236.56	19384777.49	1	Crockett	6
5438104	4494058.9	19400680.82	1	Crockett	6
5445502	4466277.69	19346868.76	1	Crockett	1
5445603	4480516.2	19352051.34	2	Crockett	6

TWDB Report ##: Final – Application of PEST to Re-Calibrate
the Groundwater Availability Model for the Edwards-Trinity (Plateau) and Pecos Valley Aquifers

Well ID	Easting (ft)	Northing (ft)	Layer	County	Number of measurements
5447205	4551785.1	19366610.79	1	Crockett	6
5509703	4615097.84	19516547.4	1	Crockett	12
2759402	4083187.06	19990042.15	2	Ector	8
2760504	4135373.78	19984346.44	2	Ector	9
2761401	4153592.64	19984540.69	2	Ector	8
2761501	4169507.84	19993302.94	2	Ector	4
2761801	4164635.2	19978824.75	2	Ector	7
2761901	4179086.66	19975571.13	2	Ector	1
2761903	4184740.38	19974739.77	2	Ector	13
2762703	4194848.42	19981310.31	2	Ector	12
2762705	4198810.82	19969578.58	2	Ector	6
4504305	4145401.49	19960516.97	2	Ector	6
4504306	4145403.78	19960617.93	2	Ector	4
4504903	4136217.02	19934897.14	2	Ector	4
4506503	4199861.16	19941904.79	2	Ector	5
4513203	4164933.66	19910549.49	2	Ector	8
4513209	4163942.47	19908545.41	2	Ector	10
4514204	4198786.54	19915390.41	2	Ector	4
4514502	4209260.63	19901598.58	2	Ector	9
5544702	4730400.87	19334298.64	1	Edwards	1
5563702	4854294.45	19235250.76	1	Edwards	7
5563803	4855612.56	19236158.83	1	Edwards	4
7006301	4836003.09	19219298.41	1	Edwards	3
7012501	4740192.84	19163188.47	1	Edwards	2
7016802	4896006.08	19153115.85	2	Edwards	8
7021301	4791161.48	19130375.57	1	Edwards	10
7024301	4911304.37	19130135.71	2	Edwards	14
7024302	4909808.32	19130642.19	2	Edwards	3
7024303	4909719.93	19129731.44	2	Edwards	13
5640701	5197875.34	19376177.07	1	Gillespie	11
5640901	5231104.82	19374432.34	2	Gillespie	10
5647406	5159324.03	19338638.32	2	Gillespie	9
5647507	5172321.22	19344196.39	1	Gillespie	11
5647603	5187406.1	19339246.3	2	Gillespie	10
5647907	5196550.69	19322513.38	2	Gillespie	9
5647908	5189292.67	19332883.17	2	Gillespie	4
5648301	5224725.96	19354128.85	1	Gillespie	11
5648503	5210231.98	19351171	2	Gillespie	11
5648802	5214244.41	19331666.63	2	Gillespie	10
5655303	5192362.23	19309522.74	2	Gillespie	4
5655603	5186227.74	19298543.45	2	Gillespie	11
5655607	5189081.28	19291985.37	2	Gillespie	2
5655608	5190412.96	19301105.98	2	Gillespie	1
5655908	5188816.18	19280848.81	2	Gillespie	4
5656202	5216608.57	19310833.25	2	Gillespie	10
5656404	5209832.4	19304298.23	2	Gillespie	11
5656405	5203516.51	19305461.18	2	Gillespie	4
5656602	5233322.59	19302374.36	2	Gillespie	4
5656805	5223199.49	19287911.26	2	Gillespie	10
5656901	5226229.73	19282066.56	2	Gillespie	4
5656902	5226523.91	19288547.65	2	Gillespie	4
5733803	5261739.81	19370263.64	1	Gillespie	8
5733903	5266359.52	19380433.38	2	Gillespie	11

TWDB Report ##: Final – Application of PEST to Re-Calibrate
the Groundwater Availability Model for the Edwards-Trinity (Plateau) and Pecos Valley Aquifers

Well ID	Easting (ft)	Northing (ft)	Layer	County	Number of measurements
5734702	5280224.34	19375005.34	1	Gillespie	10
5741403	5246283.09	19349666.38	2	Gillespie	11
5741912	5264416.63	19329188.78	2	Gillespie	10
5741914	5263849.65	19333434.74	2	Gillespie	3
5742717	5282075.53	19322787.13	2	Gillespie	7
5742718	5280505.15	19330767.76	2	Gillespie	10
5742719	5281989.29	19322684.92	2	Gillespie	4
5742722	5278292.36	19324671.32	2	Gillespie	2
5749204	5256883.67	19320713.08	2	Gillespie	11
5749503	5259412.98	19303629.8	2	Gillespie	9
5749602	5273838.83	19297597.41	2	Gillespie	4
5749701	5248885.67	19286725.66	2	Gillespie	10
5750105	5284974.05	19313503.41	2	Gillespie	4
5750116	5275952.64	19313714.46	2	Gillespie	4
5750117	5277271.77	19321725.36	2	Gillespie	4
5750409	5277372.61	19294697.87	2	Gillespie	4
5750607	5315290.92	19296319.96	2	Gillespie	4
5750901	5304924.54	19290940.74	2	Gillespie	2
5751506	5336685.46	19302237.13	2	Gillespie	4
5751508	5330215.89	19301148.32	2	Gillespie	3
5751802	5329059.21	19287671.41	2	Gillespie	9
5752118	5361552.34	19310033.37	2	Gillespie	4
4404802	4438904.96	19925856.2	2	Glasscock	10
4406702	4501871.84	19917938.62	2	Glasscock	5
4411707	4388717.26	19885278.28	2	Glasscock	14
4411808	4403975.86	19874613.44	1	Glasscock	12
4411809	4392422.39	19874181.71	1	Glasscock	14
4412302	4452950.88	19908546.19	2	Glasscock	6
4412525	4442096.43	19895631.44	1	Glasscock	14
4413401	4466479.63	19890134.8	2	Glasscock	13
4413507	4476096.82	19894565.62	2	Glasscock	2
4413904	4484705.79	19874099.57	2	Glasscock	14
4419210	4395925.63	19860961.99	1	Glasscock	13
4419307	4406369.21	19867589.27	1	Glasscock	12
4419505	4396654.53	19852241.02	1	Glasscock	18
4420239	4435208.18	19859167.97	1	Glasscock	13
4420240	4433291.01	19869829.16	1	Glasscock	13
4420305	4448920.55	19865863.33	2	Glasscock	13
4420306	4443200.95	19857435.54	2	Glasscock	11
4420402	4428389.6	19852581.48	2	Glasscock	12
4420507	4435660.05	19842350	2	Glasscock	7
4420554	4438513.95	19842917.55	1	Glasscock	7
4421102	4456703.3	19867782.79	2	Glasscock	8
4421403	4455703.95	19844908.31	2	Glasscock	10
4421604	4491259.61	19846976.91	1	Glasscock	8
5747501	5494816.85	19349630.87	2	Hays	5
5747901	5510568.34	19328734.12	2	Hays	2
5755401	5484856.75	19305429.05	2	Hays	6
5755601	5505860.79	19301421.55	2	Hays	5
5756710	5519713.16	19280396.39	2	Hays	3
5763703	5478585.87	19248233.24	2	Hays	2
5763902	5506690.19	19246667.93	2	Hays	4
5763904	5508635.4	19245789.49	2	Hays	5

TWDB Report #: Final – Application of PEST to Re-Calibrate
the Groundwater Availability Model for the Edwards-Trinity (Plateau) and Pecos Valley Aquifers

Well ID	Easting (ft)	Northing (ft)	Layer	County	Number of measurements
5764702	5516833.51	19243599.89	2	Hays	5
5764705	5516101.06	19240247.09	2	Hays	2
5764712	5513979.75	19241121.76	2	Hays	3
5849119	5554947.43	19314629.44	2	Hays	5
5849808	5566897.4	19293588.45	2	Hays	8
6808102	5523917.39	19229751.82	2	Hays	9
6808103	5514631.39	19228580.66	2	Hays	8
4349501	4627800.03	19669336.56	1	Irion	11
4351401	4691454.58	19662272.16	1	Irion	14
4357103	4613107.02	19635035.32	1	Irion	11
4358304	4678457.92	19638463.5	1	Irion	8
4447904	4558781.52	19699662.43	2	Irion	14
4455811	4553843.89	19642500	2	Irion	7
4464202	4589341.59	19635660.36	1	Irion	1
5202401	3716128.46	19591860	1	Jeff	4
5202403	3718610.39	19585694.74	1	Jeff	11
5202404	3718401.34	19584688.79	1	Jeff	3
5202407	3716591.13	19585156.96	1	Jeff	5
5757304	5268745.86	19263534.55	2	Kendall	5
5758402	5286678.59	19259364.7	2	Kendall	13
5758705	5287019	19243577.71	2	Kendall	3
5758706	5288422.24	19235292.62	2	Kendall	17
6801306	5265580.42	19228784.37	2	Kendall	2
6801312	5274257.64	19221380.6	2	Kendall	6
6801314	5270573.47	19221040.08	2	Kendall	16
6802608	5306408.28	19203905.98	2	Kendall	12
6802609	5304276.95	19206008	2	Kendall	15
6802904	5309243.19	19193815.73	2	Kendall	4
6803108	5327233.86	19226004.46	2	Kendall	3
6804312	5387339.65	19221692.73	2	Kendall	2
6804805	5376905.28	19193618.83	2	Kendall	11
6804909	5386772.5	19191724.53	2	Kendall	14
6805402	5406627.21	19211124.49	2	Kendall	4
6810615	5311877.97	19162368.56	2	Kendall	7
6810616	5314696.51	19161994.94	2	Kendall	6
6810617	5311549.13	19160340.56	2	Kendall	7
6810620	5312951.52	19160761	2	Kendall	3
6810623	5315870.67	19167068.66	2	Kendall	7
6810624	5314785.65	19161895.02	2	Kendall	2
6811103	5321876.18	19180193.3	2	Kendall	17
6811207	5337187.15	19178853.5	2	Kendall	2
6811405	5322322.81	19156514.84	2	Kendall	14
6811417	5318697.6	19158194.36	2	Kendall	1
6811708	5327615.21	19155260.15	2	Kendall	15
6811715	5329907.84	19154780.72	2	Kendall	12
6811906	5346792.12	19147998.19	2	Kendall	5
6812208	5373373.12	19174544.78	2	Kendall	13
6812409	5360169.67	19161725.48	2	Kendall	4
6818201	5299659.85	19137337.44	2	Kendall	4
5653304	5108290.89	19310185.73	2	Kerr	6
5654404	5118458.82	19292932.14	2	Kerr	5
5661601	5111754.43	19253115.1	2	Kerr	5
5662407	5126533.18	19245708.47	2	Kerr	6

TWDB Report ##: Final – Application of PEST to Re-Calibrate
the Groundwater Availability Model for the Edwards-Trinity (Plateau) and Pecos Valley Aquifers

Well ID	Easting (ft)	Northing (ft)	Layer	County	Number of measurements
5662408	5129633.12	19255544.85	2	Kerr	6
5662412	5131392.46	19254644.41	2	Kerr	4
5662414	5123742.71	19257939.61	2	Kerr	1
5662415	5128835.98	19256855.92	2	Kerr	1
5663304	5188529.18	19272546.21	2	Kerr	4
5663305	5190502.9	19266690.51	2	Kerr	1
5663408	5169099.35	19255701.19	2	Kerr	9
5663415	5158831.66	19256642.15	2	Kerr	10
5663616	5184496.96	19249234.7	2	Kerr	9
5663617	5184409.94	19249132.75	2	Kerr	4
5663916	5186620.97	19235181.45	2	Kerr	10
5663920	5195605.36	19230999.55	2	Kerr	2
5664205	5215209.27	19269015.23	2	Kerr	3
5664711	5201549.8	19233982.15	2	Kerr	1
5757703	5245656.47	19238108.37	2	Kerr	15
6801505	5259799.1	19209395.76	2	Kerr	14
6801507	5256115.48	19208854.2	2	Kerr	9
6809501	5258736.86	19155541.3	2	Kerr	6
6904502	5065793.71	19210789.23	1	Kerr	10
6907206	5177499.03	19222562.4	2	Kerr	9
6907304	5188789.87	19226999.42	2	Kerr	5
6908201	5213239.33	19221425.43	2	Kerr	11
6908510	5223857.77	19211799.26	2	Kerr	11
6908511	5213121.87	19214440.84	2	Kerr	9
6908512	5223945.65	19211800.02	2	Kerr	10
6908601	5227935.33	19207482.34	2	Kerr	4
6916201	5223827.42	19174654.12	2	Kerr	13
5540104	4891293.01	19400298.89	1	Kimble	5
5548601	4914775.68	19337117.57	1	Kimble	9
5617904	4955946.7	19464692.01	2	Kimble	12
5619402	5011436.32	19480990.7	1	Kimble	6
5620706	5047586.5	19460348.1	2	Kimble	9
5625906	4958522.61	19420249.68	2	Kimble	13
5626606	4990868.34	19439535.03	2	Kimble	5
5633804	4937594.95	19373867.54	1	Kimble	3
5635506	5023370.05	19393349.45	2	Kimble	11
5641401	4923874.99	19339040.49	1	Kimble	6
5643402	5007694.35	19338641.38	1	Kimble	3
5644203	5054136.35	19366119.59	2	Kimble	5
5645506	5097424.23	19346470.23	1	Kimble	4
7029101	4771684.33	19083094.29	2	Kinney	1
7036101	4724540.19	19037489.4	1	Kinney	11
7037402	4764266.37	19020594.67	2	Kinney	8
7038701	4808560.65	19011622.75	1	Kinney	12
7043302	4721824.16	18994200.53	2	Kinney	3
7045401	4771262.02	18983939.19	1	Kinney	17
4611702	3767381.29	19900448.16	1	Loving	6
4620403	3802801.58	19855505.28	1	Loving	14
4254201	5140150.28	19671719.86	2	Mcculloch	12
4257901	4954420.27	19602085.78	2	Menard	4
4259507	5016299.81	19613743.7	1	Menard	3
4364502	4900163.49	19624149.81	1	Menard	1
5508410	4889875.84	19571303.14	1	Menard	7

TWDB Report ##: Final – Application of PEST to Re-Calibrate
the Groundwater Availability Model for the Edwards-Trinity (Plateau) and Pecos Valley Aquifers

Well ID	Easting (ft)	Northing (ft)	Layer	County	Number of measurements
5508706	4885245.92	19548728.31	1	Menard	2
5516212	4907332.23	19540511.47	1	Menard	11
5601503	4938630.05	19575343.36	1	Menard	12
5602301	4986581.95	19584310.76	1	Menard	8
5603302	5037595.83	19579382.27	1	Menard	12
5609612	4947967.16	19524825.14	1	Menard	12
5610401	4972513.87	19521410.45	1	Menard	9
5618104	4965853.26	19489407.06	2	Menard	15
5620212	5058197.38	19489040.96	2	Menard	14
4409502	4317249.19	19892032.36	2	Midland	16
4410202	4360366.5	19905783.12	2	Midland	8
4418214	4363938.31	19865922.4	2	Midland	14
4515702	4225471.96	19885667.79	2	Midland	9
4516208	4279438.89	19910544.67	2	Midland	6
4516401	4265027.3	19895219.7	2	Midland	1
4523919	4257821.46	19838740.64	2	Midland	12
2945704	4769919.74	20057073.44	2	Nolan	10
2947701	4850130.3	20058533.53	2	Nolan	14
2953104	4774850.14	20047735.93	2	Nolan	4
2955401	4848697.65	20029575.55	2	Nolan	7
4549301	4019286.12	19681694.53	1	Pecos	18
4557601	4018070.87	19628547.78	1	Pecos	3
4561601	4178868.49	19621176.41	1	Pecos	10
4563701	4221826.57	19603064.05	1	Pecos	17
4648604	3975771.3	19716692.19	1	Pecos	14
4648801	3968880.65	19702396.16	1	Pecos	16
4655602	3937259.18	19681301.36	1	Pecos	17
4655603	3943140.2	19674650.98	1	Pecos	15
4656301	3974916.5	19685516.09	1	Pecos	14
4656401	3947946	19675933.39	1	Pecos	14
4656404	3956008.99	19682089.17	1	Pecos	15
4656502	3965988.09	19676138.75	1	Pecos	1
4663302	3942339.41	19649450.64	1	Pecos	1
4663802	3930789.15	19609767.51	1	Pecos	7
4663902	3935675.46	19622694.99	1	Pecos	11
5208801	3959901.49	19563060.79	1	Pecos	14
5216303	3978086.98	19550406.41	1	Pecos	2
5216801	3959002.39	19518114.01	1	Pecos	1
5216802	3967801.74	19530835.38	1	Pecos	20
5216902	3972921.39	19523807.49	1	Pecos	1
5301402	3995381.06	19582047.86	1	Pecos	15
5301502	3997321.51	19589592.92	1	Pecos	16
5301902	4018512.36	19565535.04	1	Pecos	14
5302102	4027051.56	19596408.45	1	Pecos	9
5302403	4029705.64	19581349.6	1	Pecos	3
5302705	4023851.63	19566814.32	1	Pecos	13
5302708	4030551.7	19560159.07	1	Pecos	13
5302802	4039998.8	19569033.36	1	Pecos	9
5309102	3985355.15	19558514.59	1	Pecos	1
5309105	3981910.82	19550201.16	1	Pecos	13
5309301	4015526.39	19551129.3	1	Pecos	16
4420801	4438487.44	19828739.46	2	Reagan	15
4420854	4441355.45	19830218.53	2	Reagan	10

TWDB Report ##: Final – Application of PEST to Re-Calibrate
the Groundwater Availability Model for the Edwards-Trinity (Plateau) and Pecos Valley Aquifers

Well ID	Easting (ft)	Northing (ft)	Layer	County	Number of measurements
4420911	4453660.27	19827820.58	2	Reagan	15
4427306	4413759.18	19819373.5	2	Reagan	11
4429202	4480806.36	19818550.55	2	Reagan	8
4429701	4463691.97	19790619.93	2	Reagan	5
4429705	4466250.51	19788155.35	2	Reagan	8
4430401	4498494.03	19804150.08	2	Reagan	3
4435301	4410704.22	19770502.9	2	Reagan	9
4436804	4433839.51	19744140.5	1	Reagan	15
4437404	4462634.38	19763088.19	2	Reagan	6
4437506	4475340.9	19763022.67	2	Reagan	6
4442319	4374933.7	19727605.22	2	Reagan	5
4442325	4373243.46	19730163.75	2	Reagan	5
4445702	4457983.2	19699957.3	2	Reagan	9
4446202	4507172.57	19721610.77	1	Reagan	5
4446301	4524417.12	19723534.85	1	Reagan	4
4451304	4404875.23	19675495.1	1	Reagan	13
4454702	4500615.85	19652221.12	2	Reagan	8
4460303	4451730.43	19628546.86	1	Reagan	4
4460304	4450167.36	19634846.9	1	Reagan	4
5657601	4955902.54	19245527.67	1	Real	10
5659401	5009742.06	19255541.71	1	Real	2
6901702	4928103.66	19191868.87	2	Real	6
6902202	4984819.83	19215305.51	1	Real	8
6917101	4932606.59	19123355.98	2	Real	11
6918303	4996846.93	19134769.69	2	Real	12
6919401	5010807.33	19113045.34	2	Real	20
6920101	5042073.27	19137325.26	2	Real	10
7024601	4918520.27	19114145.36	2	Real	12
4609901	3715900.78	19902117.29	1	Reeves	8
4610701	3725256.47	19898547.28	1	Reeves	1
4618801	3734256.84	19857198.26	1	Reeves	6
4626401	3717854.25	19822507.07	1	Reeves	10
4628801	3809911.14	19800461.95	1	Reeves	4
4628802	3819080.82	19798039.42	1	Reeves	4
4633501	3696828.88	19780286.96	1	Reeves	9
4635501	3773650.5	19775714.84	1	Reeves	16
4635601	3786798.58	19765248.62	1	Reeves	16
4635803	3778979.48	19761657.23	1	Reeves	6
4635902	3789587.17	19760597.6	1	Reeves	4
4636201	3808578.43	19788651.04	1	Reeves	10
4636401	3801205.64	19771970.34	1	Reeves	13
4636707	3793059.09	19750351.99	1	Reeves	1
4636901	3820768.51	19750772.72	1	Reeves	5
4636903	3818893.29	19754378.94	1	Reeves	8
4636909	3818899.76	19754581.15	1	Reeves	4
4642810	3738239.68	19716728.81	1	Reeves	15
4642901	3745894.53	19712515.36	1	Reeves	16
4643213	3775570.91	19742418.6	1	Reeves	12
4643501	3775063.7	19729770.55	1	Reeves	18
4643902	3788147.34	19714140.95	1	Reeves	7
4644101	3801270.77	19736710.71	1	Reeves	17
4644203	3806427.82	19735530.29	1	Reeves	16
4644501	3817151.51	19729714.17	1	Reeves	19

TWDB Report #: Final – Application of PEST to Re-Calibrate
the Groundwater Availability Model for the Edwards-Trinity (Plateau) and Pecos Valley Aquifers

Well ID	Easting (ft)	Northing (ft)	Layer	County	Number of measurements
4644502	3809517.39	19721246.56	1	Reeves	3
4644701	3795844.96	19708417.94	1	Reeves	1
4644803	3811784.89	19705571.67	1	Reeves	14
4644805	3811787.71	19711042.08	1	Reeves	4
4651520	3774658.67	19683685.71	1	Reeves	9
4651903	3777194.34	19668607.17	1	Reeves	16
4651907	3788615.32	19662354.61	1	Reeves	12
4652101	3792521.44	19696875.25	1	Reeves	1
4652102	3796333.67	19699486.57	1	Reeves	2
4652104	3790777.08	19691259	1	Reeves	12
4652111	3798019.44	19698013.22	1	Reeves	12
4652201	3808608.75	19690375.65	1	Reeves	15
4652204	3807222.82	19701159.28	1	Reeves	3
4652404	3802570.82	19680642.37	1	Reeves	16
4652501	3807223.31	19679681.13	1	Reeves	17
4652607	3826273.89	19673297.1	1	Reeves	12
4652703	3793273.12	19672130.65	1	Reeves	17
4655201	3932558.67	19686703.5	1	Reeves	16
4658402	3719907.74	19630532.28	1	Reeves	4
4658403	3719824.42	19630636.21	1	Reeves	11
4659105	3757504.34	19650724.68	1	Reeves	15
4659201	3768467.04	19651471.39	1	Reeves	15
4659303	3787647.81	19643440.93	1	Reeves	10
4659304	3783614.44	19650158.88	1	Reeves	11
4659401	3756117.42	19637904.13	1	Reeves	16
4659508	3764974.74	19630109.77	1	Reeves	12
4660101	3798336.84	19654438.99	1	Reeves	10
5201302	3703622.23	19609622.77	1	Reeves	3
5203302	3784239.75	19597862.62	1	Reeves	10
5204105	3794603.56	19606438.49	1	Reeves	11
4358601	4681889.97	19620516.61	1	Schleicher	10
4360920	4752976.85	19603998.06	1	Schleicher	1
4361706	4766056.67	19596950.52	1	Schleicher	20
5502202	4665814.17	19582559	1	Schleicher	9
5502703	4656840.73	19556603.73	1	Schleicher	2
5502807	4660526.89	19561132.6	2	Schleicher	11
5504215	4744689.09	19579332.99	1	Schleicher	5
5507101	4846007.55	19581091.1	1	Schleicher	14
5509503	4621732.66	19528841.92	1	Schleicher	10
5511201	4706164.8	19546339.69	1	Schleicher	6
5511503	4700809.9	19524907.94	1	Schleicher	9
5512116	4730331.38	19544071.95	1	Schleicher	4
5512127	4734664.51	19541111.88	1	Schleicher	2
5513405	4776263.49	19524814.34	1	Schleicher	12
5514901	4832532.41	19509943.25	1	Schleicher	11
4423801	4553883.65	19824780.44	1	Sterling	9
4432402	4584369.12	19801176.63	2	Sterling	10
5520904	4762531.35	19471719.33	1	Sutton	6
5522901	4842526.55	19463650.05	1	Sutton	12
5525902	4638315.66	19417128.66	1	Sutton	14
5526105	4648420.24	19447623.32	1	Sutton	6
5527108	4695695.88	19447587.94	1	Sutton	7
5527202	4711180.79	19455286.33	1	Sutton	10

TWDB Report #: Final – Application of PEST to Re-Calibrate
the Groundwater Availability Model for the Edwards-Trinity (Plateau) and Pecos Valley Aquifers

Well ID	Easting (ft)	Northing (ft)	Layer	County	Number of measurements
5527606	4719317.92	19429927.59	1	Sutton	8
5527615	4719820.58	19441365.05	1	Sutton	10
5527616	4721788.12	19434570.52	1	Sutton	10
5527906	4712145.73	19427236.47	1	Sutton	7
5533807	4625479.81	19375727.23	1	Sutton	7
5535103	4693815.61	19400118.07	1	Sutton	13
5536502	4748871.09	19388564.63	1	Sutton	12
5537803	4783261.16	19380009	1	Sutton	9
5538402	4803577.91	19392487.43	1	Sutton	9
5538403	4803463.74	19384490	1	Sutton	13
5545307	4798549.84	19353833.53	1	Sutton	6
5353201	4153420.67	19321070.27	1	Terrell	2
5353401	4139738.11	19300618.17	1	Terrell	3
5353809	4155045.78	19291369.21	2	Terrell	1
5427802	4397348.46	19429036.07	2	Terrell	11
5435401	4375028.13	19391714.49	2	Terrell	5
4353801	4786231.36	19648101.67	2	Tom	10
4354202	4828803.52	19678136.39	1	Tom	7
4355502	4864742.48	19667426.89	2	Tom	2
4355802	4860024.71	19646170.81	1	Tom	7
5739905	5509074.11	19386114.84	2	Travis	6
5740304	5543048.5	19407056.5	2	Travis	12
5748206	5537195.39	19367566.32	2	Travis	3
5841101	5558938.59	19359756.48	2	Travis	13
5842502	5608242.87	19356846.14	2	Travis	13
5842802	5609745.13	19338347.35	2	Travis	3
5850120	5591601.15	19321490.62	2	Travis	12
4425805	4318277.1	19783543.47	2	Upton	16
4426508	4355485.13	19802660.88	2	Upton	11
4433501	4318652.66	19765509.11	2	Upton	16
4434104	4350608.97	19774180.98	2	Upton	15
4531501	4239897.98	19800603.3	2	Upton	8
4555601	4255222.41	19672689.87	1	Upton	6
5456403	4575111.98	19294293.5	1	Val	14
5456502	4591223.06	19302539.54	1	Val	5
5460502	4424191.34	19260483.16	1	Val	2
7025502	4624545.45	19068446.06	1	Val	18
7025603	4638278.15	19065600.29	1	Val	6
7026401	4650442.75	19076432.92	1	Val	10
7033501	4625453.97	19029684.06	1	Val	11
7033508	4620092.78	19032461.5	1	Val	1
7033604	4629952.97	19029342.72	1	Val	12
7033704	4604390.1	19013880.83	1	Val	15
7034501	4667542.12	19027128.96	1	Val	4
7034703	4643876.22	19003834.21	1	Val	4
7041208	4623634.69	19002481.77	1	Val	2
7041209	4622614.24	18996723.47	1	Val	18
7042106	4645269.59	19001293.63	1	Val	3
7042205	4666469.62	19000931.65	1	Val	19
7104402	4420339.69	19205877.16	1	Val	3
7104701	4419424.48	19197185.75	1	Val	3
7107601	4557371.84	19206005.87	1	Val	9
7112502	4423893.06	19165340.3	1	Val	14

TWDB Report ##: Final – Application of PEST to Re-Calibrate
the Groundwater Availability Model for the Edwards-Trinity (Plateau) and Pecos Valley Aquifers

Well ID	Easting (ft)	Northing (ft)	Layer	County	Number of measurements
7123901	4558176.81	19105499.48	1	Val	14
7132401	4565612.38	19074255.04	1	Val	8
7140602	4594151.56	19033502.6	1	Val	2
4525204	4005647.66	19833594.89	1	Ward	11
4525709	3988721.2	19795048.9	1	Ward	17
4526202	4044303.02	19829750.37	1	Ward	15
4533102	3997225.09	19786715.56	1	Ward	15
4533501	4009115.28	19765733.72	1	Ward	14
4533803	4008955.1	19756418.4	1	Ward	13
4534701	4034727.28	19749465.18	1	Ward	15
4542512	4049610.96	19711910.45	1	Ward	3
4624703	3962777.17	19837701.43	1	Ward	12
4624705	3960605.73	19840294.87	1	Ward	9
4624801	3970547.82	19840930.07	1	Ward	8
4624802	3971003.42	19838688.84	1	Ward	1
4624803	3965225.8	19838849.1	1	Ward	2
4629201	3858284.72	19833278.14	1	Ward	4
4629701	3842766.2	19800936.02	1	Ward	16
4629801	3857973.52	19800867.81	1	Ward	4
4631302	3944189.2	19824447.73	1	Ward	14
4632403	3949894.96	19806355.05	1	Ward	4
4632504	3974824.94	19808192.77	1	Ward	6
4632514	3968868.59	19805216.49	1	Ward	13
4632516	3969197.04	19804599.83	1	Ward	14
4632626	3983473.11	19808563.43	1	Ward	18
4632906	3982278.36	19796439.63	1	Ward	16
4632921	3988152.19	19789999.63	1	Ward	17
4637101	3840506.73	19778515.68	1	Ward	18
4637211	3857866.21	19780608.67	1	Ward	2
4637305	3870216.09	19780127.94	1	Ward	14
4638103	3872594.88	19778738.76	1	Ward	18
4639604	3945060.91	19769921.13	1	Ward	13
4640205	3968584.28	19788712.38	1	Ward	2
4640206	3973495.83	19781890.62	1	Ward	14
4640301	3976162.8	19778271.82	1	Ward	13
4640307	3977053.81	19785541.05	1	Ward	3
4640308	3978389.27	19780641.75	1	Ward	11
4501901	4021596.34	19934777.14	1	Winkler	15
4510801	4043594.48	19883051.36	1	Winkler	5
4517901	4018459	19848147.2	1	Winkler	3
4606301	3912780.74	19968807.55	1	Winkler	9
4606901	3912612.71	19939736.98	1	Winkler	6
4607401	3917502.41	19948001.77	1	Winkler	7
4607402	3925508.78	19948173.27	1	Winkler	11
4607901	3947626.52	19939332.02	1	Winkler	14
4615402	3926416.98	19911270.95	1	Winkler	13
4616901	3988393.42	19884709.96	1	Winkler	14
4623304	3942504.13	19880416.25	1	Winkler	3
4624301	3987982.48	19872767.38	1	Winkler	15

TWDB Report ##: Final – Application of PEST to Re-Calibrate
the Groundwater Availability Model for the Edwards-Trinity (Plateau) and Pecos Valley Aquifers

Appendix G Program documentation

G.1 General

This appendix documents utility programs which were written specifically for the current project. Source code and executable versions of these programs are provided on the CD which accompanies this report.

G.2 CLIPARRAY1

CLIPARRAY1 constrains elements of a MODFLOW-specific real array by zone-specific upper and lower bounds.

CLIPARRAY prompts the user for operational details. Upon commencement of execution it asks:

```
Enter name of grid specification file:
```

Details of this file are provided in documentation for the Groundwater Data Utilities suite, many of the programs of which commence execution by reading this same file. Next it prompts:

```
Enter name of integer array file:
```

Details of this file type are also recorded in documentation of the Groundwater Data Utilities suite. Note that CLIPARRAY1 does NOT expect a number of columns, number of rows header in this file.

Its next prompt is:

```
Enter name of integer-limit-correspondence file:
```

Specifications for this file are presented below.

Finally, CLIPARRAY prompts for the names of the real array file which it must read and that which it must write (after imposition of clipping constraints); these filenames can be the same if desired. Real array files are described in documentation to the Groundwater Data Utilities. As for integer arrays, CLIPARRAY does not expect to read, nor records, a number of columns, number of rows header in these files. Prompts are:

```
Enter name of input real array file:
```

```
Enter name for output real array file:
```

After having obtained all requisite data, CLIPARRAY1 undertakes necessary processing and then terminates execution.

An example of an integer-limit-correspondence file follows.

1	-2.0	2.0
3	-10.0	40.0

Figure G-1. Example of an integer-limit-correspondence file.

Each line has three entries. The first is an integer, the second and third are real numbers. These last two numbers are the upper and lower limits enforced on real array entries within zones appearing within the integer array that are defined by the nominated integers. If an integer occurring in the integer array is missing from the integer-limit-correspondence file, bounds are not enforced in the corresponding zone. Zones nominated in the integer-limit-correspondence file that are missing from the integer array are simply ignored.

G.3 REAL2CND

REAL2CND reads a MODFLOW *drain* or *general head boundary* package input file. It replaces conductance values in that file with those recorded in real array format.

REAL2COND begins execution with the prompt:

```
Enter name of grid specification file:
```

Specifications for a grid specification file are provided in the manual for the PEST Groundwater Data Utilities. Next, REAL2CND asks for more information pertaining to the current model:

```
How many layers in model?
```

It then requests the names of NLAY files holding NLAY different MODFLOW/MT3D-compatible real arrays (where NLAY is the number of layers in the model). The prompt is:

```
Enter name of real array conductance file for layer N:
```

Note that a number-of-columns/number-of-rows header must not be present in the real array file supplied to REAL2CND.

REAL2CND then asks for the name of an existing *drain* or *general head boundary* input file for the current model:

```
Enter name of existing MODFLOW DRN/GHB file:
```

and finally for the name of the new file that it must write:

```
Enter name for new MODFLOW DRN/GHB file:
```

In writing this new file it replaces existing conductances with values assigned to corresponding cells in the conductance real arrays which it has just read.

It is important to note that REAL2CND assumes that MODFLOW reads tabular data in free field format; it also assumes that the existing *drain* or *general head boundary* file can be read using free field formatting. As explained in PEST documentation, transfer of numbers between individual programs comprising a total model must be such that all significant figures associated with each parameter-dependent number are preserved. This may not be possible if the old MODFLOW convention of fixed field formatting (with a 10 character field width) is employed.

G.4 REALSTR2CND

REALSTR2CND serves a similar purpose to REAL2CND in that it reads conductances from one or a number of MODFLOW/MT3D-compatible real arrays. It replaces conductances in an existing *streamflow routing* package input file with these new conductances.

REALSTR2CND begins execution with the prompt:

```
Enter name of grid specification file:
```

See documentation of the PEST Groundwater Data Utilities for details of this file. REALSTR2CND next asks for further information on the model, namely:

```
How many layers in model?
```

REALSTR2CND next asks for the name of a real array containing cell-by-cell stream conductance for each layer. The prompt for layer *N* is:

```
Enter name of real array conductance file for layer N:
```

It then asks for the name of an existing MODFLOW *streamflow routing* package input file. It will use this file as a basis for production of a new *streamflow routing* package input file. However, stream conductances cited in the existing file will be overwritten by those provided in the real arrays. REALSTR2CND's prompt is:

```
Enter name of existing MODFLOW STR1 file:
```

Next, it asks for the name of the *streamflow routing* package input file which it must write:

```
Enter name for new STR1 file:
```

It then checks that it can read the existing *streamflow routing* package input file using free field formatting, and write the new file using free field formatting (thereby preserving maximum numerical precision for the sake of PEST that requires this for the integrity of derivatives calculation).

```
Free field format expected on input/output files - ok? [y/n]:
```

Finally, it asks for a roughness value to employ for all stream reaches in writing this file. The prompt is:

Enter value for global roughness:

G.5 STRFLBUD

STRFLBUD reads a binary MODFLOW96 cell-by-cell flow term output file. Its task is to extract stream flow data as recorded by the STR1 package from that file and write that data in bore sample file format, ready for processing by members of the Groundwater Data Utility suite. Note that, irrespective of whether compact array storage is employed or not, *streamflow routing package* cell-by-cell flow term data is recorded using the old three-dimensional array format. The binary header to stream flow arrays is STREAM FLOW OUT.

Prior to running STRFLBUD, an input file must be prepared. An example of such a file is shown below.

5	6		NCELL	NPER	
uad_Sat	64	382	NAME	ROW	COLUMN
uad_Sp	72	373			
edern	74	335			
an_Saba	75	262			
uadCom	87	346			
1/01/1940	12:00:00		DATESTRING		
1/01/1950	00:00:00				
1/01/1960	12:00:00				
1/01/1970	00:00:00				
1/01/1980	12:00:00				
2/31/1980	18:00:00				

Figure G-2. An example STRFLBUD input file.

The first line of a STRFLBUD input file must be comprised of two integers, these being the number of cells (NCELL) followed by the number of stress periods (NPER). The meaning of NCELL is provided below. NPER must correspond to the number of stress periods for which cell-by-cell flow term output appears in the MODFLOW-generated budget file. STRFLBUD assumes that such output is generated only at the end of a stress period, and not at the end of time steps that do not mark the end of a stress period. Once it has read these integers STRFLBUD can dimension its internal arrays and read the remainder of its input file.

Following the NCELL/NPER header, the STRFLBUD input file must provide NCELL rows of data. Each row must include three entries. The first must be an ASCII string which serves as the name of a stream flow measurement point. Then must follow the row and column of that cell within the model domain. STRFLBUD reads stream flow recorded at that cell, summing stream flow over all layers of the model.

NPER lines of data must follow the above stream flow measurement cell specifications. Each such line must contain two entries, the first being a date and the second being a time. In transferring stream flows to its output bore sample file, STRFLBUD associates these date and

time strings with each stream flow output array found in the MODFLOW cell-by-cell flow term file which it reads.

Upon commencement of execution, STRFLBUD prompts the user:

```
Enter name of grid specification file:
```

The format of this file is provided in documentation of the Groundwater Data Utilities. It then asks the user for an extra model specification, namely:

```
Enter number of layers in model:
```

then for the binary cell-by-cell flow term file which it must read:

```
Enter name of MODLFLOW budget file:
```

find then for the name of its input file:

```
Enter name for STRFLBUD output file:
```

Next, it asks for the name of the bore sample file in which it will record stream flows that it extracts from nominated cells of the MODFLOW cell-by-cell flow term file:

```
Enter name for STRFLBUD output file:
```

and for a factor by which it should multiply the flows that it reads to effect units conversion:

```
Enter name for stress period list file:
```

Finally, it asks:

```
Enter name for stress period list file:
```

As it reads arrays present within the cell-by-cell flow term file, it records the stress periods and time steps corresponding to these arrays as gleaned from binary headers to these arrays. These are recorded in the above file for the user's convenience.

G.6 VCONTCALC

VCONTCALC computes the vertical conductance between two MODFLOW layers on the assumption that there is no intervening confining layer. It reads an array of horizontal hydraulic conductivity values for each of the upper and lower of these layers; vertical conductivity is assumed to be related to horizontal conductivity through multiplication by the same, uniform, vertical anisotropy value. It also reads the thickness of each layer. Within each cell, inter-layer vertical conductance is computed using the formula.

$$1/v = (t_u/k_u + t_l/k_l)/2$$

where:

- v is the vertical conductance value computed for the cell (i.e., MODFLOW VCONT value).
- t_u is the thickness of the upper layer;
- t_l is the thickness of the lower layer;
- k_u is the hydraulic conductivity of the upper layer; and
- k_l is the hydraulic conductivity of the lower layer.

Upon initiation of execution, VCONTCALC first prompts for the name of the grid specification file for the current model. The prompt is:

```
Enter name of grid specification file:
```

The format of this file type is provided in documentation of the Groundwater Data Utility suite. Next it asks for the names of four files, each of which should contain a MODFLOW-compatible real array:

```
Enter K array for upper layer:  
Enter K array for lower layer:  
Enter thickness array for upper layer:  
Enter thickness array for lower layer:
```

Note that VCONTCALC does not expect a number-of-rows, number-of-columns header in these array files. VCONTCALC then asks:

```
Enter inactive threshold for all these arrays:
```

If the absolute value of any element of any of these arrays is above the number provided, then the cell is considered inactive from the point of view of vertical conductance calculations. A dummy conductance value of 1E35 is then provided for the pertinent array element in the vertical conductance file generated by VCONTCALC.

Next, VCONTCALC asks for the name of the MODFLOW-compatible real array file which it must write:

```
Enter name for output file:
```

and for the vertical anisotropy value that it should apply to both upper and lower layer hydraulic conductivities:

```
Enter value for global vertical anisotropy ratio:
```

Armed with all of the information that it needs, VCONTCALC computes vertical conductance on a cell-by-cell basis, and records computed conductances to its output file.

G.7 WELLBUILD2

The WELLBUILD2 utility was purpose-written for the current project. Its task is to create a MODFLOW well package input file for the transient component of the ETPV GAM based on design specifications which are specific to that model. These specifications are as follows.

1. The model has two active layers.
2. The transient model runs for 25 stress periods.
3. MODFLOW *well* package data is provided for the last 20 of these stress periods.
4. For the first five stress periods, pumping is obtained by multiplying a standard well package input dataset by stress-period-specific build-up factors.
5. *Well* extraction over all stress periods is subject to domain-wide, layer-specific, cell-by-cell alteration; the means through which this alteration is effected is through layer-specific multiplier arrays.

WELLBUILD2 begins execution with the prompt:

```
Enter name of grid specification file:
```

See documentation of the PEST Groundwater Data Utilities for specifications of this file type. WELLBUILD2 next asks for more model specifications:

```
How many layers in the model?
```

WELLBUILD2 then asks for a multiplier array file for every layer. For each layer, it prompts:

```
Enter multiplier array file for layer N:
```

where N is the pertinent layer number. When writing the well package input file for the model, the pumping rate in every cell is multiplied by the factor associated with that cell. The array (which must have the same number of rows and columns as the model) must be supplied in a standard array file, without number of columns, number of rows header.

WELLBUILD2 next asks:

```
Enter name of standard WELL stress period data file:
```

The expected format for this file is shown below. Its format is identical to that provided for each stress period to MODFLOW in its well package input data file. An integer header provides the number of data lines to follow. Each ensuing line contains four entries, these being layer number, row number, column number and cell injection rate (which is negative for extraction).

16				
	1	58	244	-116.5
	1	58	245	-116.6
	1	58	253	-116.5
	1	58	254	-117
	1	58	255	-117.3
	1	58	261	-113.8
	1	58	262	-13.6
	1	58	263	-11.7
	1	58	267	-13.6
	1	58	268	-13.6
	1	59	242	-40.9
	1	59	243	-107.9
	1	59	244	-116.7
	1	59	245	-119.2
	1	59	253	-113.8
	1	59	254	-106.2

Figure G-3. An example of a well stress period data file.

WELLBUILD2 then prompts:

Enter number of build-up periods:

For the ETPV GAM, the response to this prompt should be five. WELLBUILD will then write *well* package input data for the first five stress periods by multiplying standard pumping rates supplied in the standard well stress period data file by build up factors (and then by the multiplier arrays discussed above). Build-up factors are zone-based; WELLBUILD2 prompts for the name of a file containing an integer array defining these zones. The integer array must be compatible with the current model; no number of columns/number of rows header is expected in this file. The prompt is:

Enter name of build-up integer zonation file:

WELLBUILD2's next prompt is:

Enter name of build-up factor file:

An example of a build-up factor file appears below. The first line must contain a single integer, this representing the number of lines to follow. Ideally, this should be equal to the number of zones represented in the build-up integer array zonation file. Then, on each subsequent data line, a single integer followed by N real numbers must be provide, where N is the number of build-up periods. The integer identifies a zone number; this must correspond to an integer represented in the build-up integer zonation file. The five ensuing numbers provide build-up factors for the nominated zone. All pumping in the standard well stress period data file for cells which lie within this zone are multiplied by the provided factor during the stress period to which that factor pertains. Pumping rates for all build-up stress periods are then multiplied by cell-specific multipliers provided in the multiplier array files before being written to a MODFLOW *well* package input file.

15					
1	1.9966E-01	3.9779E-01	5.9851E-01	7.9798E-01	9.9350E-01
2	1.0447E-01	1.8801E+00	3.0996E+00	5.2092E+00	3.5293E+00
3	9.3001E-02	6.2014E-01	6.0000E+00	5.3063E+00	9.6409E-01
4	2.7273E+00	4.4255E+00	1.5348E+00	1.9337E+00	1.0910E+00
5	2.3631E-01	5.8942E-01	1.1130E+00	1.5139E+00	8.5822E-01
6	1.9994E-01	4.0003E-01	5.9944E-01	7.9211E-01	9.7344E-01
7	6.8396E-01	2.9346E+00	1.9052E+00	1.0088E+00	1.1481E+00
8	1.0125E-01	4.4360E-01	5.1840E-01	1.1014E+00	9.8423E-01
9	2.0560E-01	4.2661E-01	6.5984E-01	9.3955E-01	1.2913E+00
10	2.0358E-01	4.2353E-01	6.6999E-01	1.0037E+00	1.9741E+00
11	2.0609E-01	4.3605E-01	7.5499E-01	1.4717E+00	2.0000E+00
12	1.6625E-01	3.2001E-01	5.6912E-01	2.0000E+00	2.0000E+00
13	1.9910E-01	3.9609E-01	5.9895E-01	8.2675E-01	1.1418E+00
14	1.9998E-01	3.9998E-01	5.9983E-01	7.9995E-01	9.9983E-01
15	1.2390E+01	2.0000E+01	2.0000E+01	8.3028E+00	1.5240E+00

Figure G-4. Example of a build-up factor file.

Next WELLBUILD2 asks:

Enter name of post-build-up WELL package input file:

This file should contain *well* data for as many stress periods as there are in the model following the build-up stress periods. The format is identical to that of the standard stress period data file, except that data pertaining to different stress periods are listed one set after the other. These data are transferred to the MODFLOW *well* package input file that WELLBUILD2 writes, after multiplication by multiplier arrays as discussed above. The name of this new MODFLOW *well* package input file must be supplied in response to WELLBUILD2's next prompt, which is:

Enter name for new WELL package input file:

WELLBUILD2's final prompt is:

Enter MXWELL:

As discussed in MODFLOW documentation, MXWELL is the maximum number of wells for which pumping data is supplied in any one stress period. This number is written to the top of the well package input file written by WELLBUILD2, together with other information required in this file.

G.8 PTINGRID

PTINGRID is a member of the Groundwater Data Utility suite. It is comprehensively documented in the manual for that suite. It performs the following tasks.

1. It reads the coordinates of a list of named points.
2. It establishes the row and column number of the model cell in which each point lies.
3. Optionally it reads a MODFLOW-compatible real or integer array and reports the value of the element of that array that pertains to each point-occupied cell.

G.9 NONLINOBS

NONLINOBS was written specifically for the purpose of assisting in calibration of the Edwards-Trinity model. It enforces an objective function penalty if the head calculated within a particular model cell falls below a certain value. This measure was used in calibration of the Edwards Trinity model in order to avoid estimation of a set of parameters that leads to MODFLOW convergence problems; these express themselves in unrealistically low heads in certain parts of the model domain, and consequential local mass imbalance.

NONLINOBS is hardwired to read an input file named *badcell.dat*. This is expected to be a PTINGRID output file containing the row and column number of a single cell, together with the model-calculated head for that cell. It writes an output file named *nonlinobs.out*. This contains a single number. That number is zero if the head is above a value of 3900. It is equal to 3900 minus the head if the head is below 3900. PEST reads this value as a model output for which the observed counterpart is zero. Thus PEST enforces an objective function penalty (the size of the penalty depending on the weight assigned to this observation) if the model-calculated head in the cell falls below a certain value.

Appendix H The PEST control file

The total length of the PEST control file is 14,911 lines. This listing provides the contents for each input section up to a maximum of 20 lines.

```

pcf
* control data
restart regularisation
    2485    9832    20    2485    28
    25      6
single point 1 0 0
10.0 -3.0 0.3 0.03 10 999
10.0 10.0 0.001
0.1 noaui
    0 0.005 4 4 0.005 4
0 0 0
* singular value decomposition
0
2410 5e-7
0
* lsqr
1
1e-10 1e-10 5e3 50000
0
* parameter groups
hkzone    relative    0.015 0.0    switch 2.0 parabolic
szone     relative    0.015 1.0e-7 switch 2.0 parabolic
syzone    relative    0.015 1.0e-4 switch 2.0 parabolic
rfzone    relative    0.015 1.0e-4 switch 2.0 parabolic
vertanis  relative    0.015 0.0    switch 2.0 parabolic
welfac    relative    0.015 0.002 switch 2.0 parabolic
drain     relative    0.015 0.0    switch 2.0 parabolic
ghb       relative    0.015 0.0    switch 2.0 parabolic
hkmul1    relative    0.015 0.0    switch 2.0 parabolic
hkmul2    relative    0.015 0.0    switch 2.0 parabolic
rechmul   relative    0.015 0.0    switch 2.0 parabolic
smul1     relative    0.015 0.0    switch 2.0 parabolic
smul2     relative    0.015 0.0    switch 2.0 parabolic
symul1    relative    0.015 0.0    switch 2.0 parabolic
symul2    relative    0.015 0.0    switch 2.0 parabolic
vcontmul  relative    0.015 0.0    switch 2.0 parabolic
welmul1   relative    0.015 0.0    switch 2.0 parabolic
welmul2   relative    0.015 0.0    switch 2.0 parabolic
strcond   relative    0.015 0.0    switch 2.0 parabolic
str_rough relative    0.015 0.0    switch 2.0 parabolic
* parameter data
welfac1_1 log factor 0.2 0.1 2.0 welfac 1.0 0.0 1
welfac2_1 log factor 0.1 0.1 2.0 welfac 1.0 0.0 1
welfac3_1 log factor 0.1 0.1 2.0 welfac 1.0 0.0 1
welfac4_1 log factor 4.0 0.1 4.0 welfac 1.0 0.0 1
welfac5_1 log factor 0.2 0.1 2.0 welfac 1.0 0.0 1
welfac6_1 log factor 0.2 0.1 2.0 welfac 1.0 0.0 1
welfac7_1 log factor 0.7 0.1 2.0 welfac 1.0 0.0 1
welfac8_1 log factor 0.1 0.1 2.0 welfac 1.0 0.0 1
welfac9_1 log factor 0.2 0.1 2.0 welfac 1.0 0.0 1
welfac10_1 log factor 0.2 0.1 2.0 welfac 1.0 0.0 1

```

TWDB Report ##: Final – Application of PEST to Re-Calibrate
the Groundwater Availability Model for the Edwards-Trinity (Plateau) and Pecos Valley Aquifers

welfac11_1	log	factor	0.2	0.1	2.0	welfac	1.0	0.0	1
welfac12_1	log	factor	0.2	0.1	2.0	welfac	1.0	0.0	1
welfac13_1	log	factor	0.2	0.1	2.0	welfac	1.0	0.0	1
welfac14_1	log	factor	0.2	0.1	2.0	welfac	1.0	0.0	1
welfac15_1	log	factor	9.1	0.1	15.0	welfac	1.0	0.0	1
welfac16_1	log	factor	0.2	0.1	2.0	welfac	1.0	0.0	1
welfac1_2	log	factor	0.4	0.1	2.0	welfac	1.0	0.0	1
welfac2_2	log	factor	0.87	0.1	3.0	welfac	1.0	0.0	1
welfac3_2	log	factor	0.49	0.1	3.0	welfac	1.0	0.0	1
welfac4_2	log	factor	18.0	0.1	18.0	welfac	1.0	0.0	1
welfac5_2	log	factor	0.4	0.1	2.0	welfac	1.0	0.0	1
----- lines missing -----									
* observation groups									
dummy_head									
flows									
heads_lay1									
heads_lay2									
diffs_lay1									
diffs_lay2									
regul_mhk1									
regul_mhk2									
r_hkzone									
r_szone									
r_syzone									
r_rfzone									
r_vertan									
regul_welfac									
regul_drain									
regul_ghb									
regul_hkmul1									
regul_hkmul2									
regul_rechmu									
regul_smul1									
regul_smul2									
* observation data									
dummy_head			0.00		0.005	dummy_head			
f_s08128000_01			35.91250		0.00000	flows			
f_s08128000_02			29.26667		0.00000	flows			
f_s08128000_03			16.50000		0.00000	flows			
f_s08128000_04			9.505000		0.00000	flows			
f_s08128000_05			5.685000		0.00000	flows			
f_s08128000_06			12.78333		0.00000	flows			
f_s08128000_07			32.10500		0.00000	flows			
f_s08128000_08			49.40000		0.00000	flows			
f_s08128000_09			21.71667		0.00000	flows			
f_s08128000_10			13.74333		0.00000	flows			
f_s08128000_11			37.14583		0.00000	flows			
f_s08128000_12			29.84167		0.00000	flows			
f_s08128000_13			50.16667		0.00000	flows			
f_s08128000_14			23.93333		0.00000	flows			
f_s08128000_15			17.56667		0.00000	flows			
f_s08128000_16			15.62222		0.00000	flows			
f_s08144500_01			112.2450		0.00000	flows			
f_s08144500_02			36.72500		0.00000	flows			
f_s08144500_03			25.08333		0.00000	flows			
f_s08144500_04			15.72417		0.00000	flows			
f_s08144500_05			23.49667		0.00000	flows			

TWDB Report ##: Final – Application of PEST to Re-Calibrate
the Groundwater Availability Model for the Edwards-Trinity (Plateau) and Pecos Valley Aquifers

```

* model command line
  model_gt6b_jd.bat
* model input/output
  buildup_factors.tpl          buildup_factors.dat
  hk_lay1.tpl                  hk_lay1.irc
  hk_lay2.tpl                  hk_lay2.irc
  s_lay1.tpl                   s_lay1.irc
  s_lay2.tpl                   s_lay2.irc
  sy_lay1.tpl                  sy_lay1.irc
  sy_lay2.tpl                  sy_lay2.irc
  vcontcalc.tpl               vcontcalc.in
  rf.tpl                       rf.irc
  hk_lay1_mul.tpl             hk_lay1_mul.dat
  hk_lay2_mul.tpl             hk_lay2_mul.dat
  s_lay1_mul.tpl              s_lay1_mul.dat
  s_lay2_mul.tpl              s_lay2_mul.dat
  sy_lay1_mul.tpl             sy_lay1_mul.dat
  sy_lay2_mul.tpl             sy_lay2_mul.dat
  vcont_lay1_mul.tpl          vcont_lay1_mul.dat
  rechmul.tpl                 rechmul.dat
  welmul_lay1.tpl             welmul_lay1.dat
  welmul_lay2.tpl             welmul_lay2.dat
  draincond_lay1.tpl          draincond_lay1.irc
  draincond_lay2.tpl          draincond_lay2.irc
  ghb_lay1.tpl                ghb_lay1.dat
* prior information
welfac1_1  1.0 * log(welfac1_1) = -0.6989700    1.0 regul_welfac
welfac2_1  1.0 * log(welfac2_1) = -1.0000000    1.0 regul_welfac
welfac3_1  1.0 * log(welfac3_1) = -1.0000000    1.0 regul_welfac
welfac4_1  1.0 * log(welfac4_1) =  0.6020600    1.0 regul_welfac
welfac5_1  1.0 * log(welfac5_1) = -0.6989700    1.0 regul_welfac
welfac6_1  1.0 * log(welfac6_1) = -0.6989700    1.0 regul_welfac
welfac7_1  1.0 * log(welfac7_1) = -0.1549020    1.0 regul_welfac
welfac8_1  1.0 * log(welfac8_1) = -1.0000000    1.0 regul_welfac
welfac9_1  1.0 * log(welfac9_1) = -0.6989700    1.0 regul_welfac
welfac10_1 1.0 * log(welfac10_1) = -0.6989700    1.0 regul_welfac
welfac11_1 1.0 * log(welfac11_1) = -0.6989700    1.0 regul_welfac
welfac12_1 1.0 * log(welfac12_1) = -0.6989700    1.0 regul_welfac
welfac13_1 1.0 * log(welfac13_1) = -0.6989700    1.0 regul_welfac
welfac14_1 1.0 * log(welfac14_1) = -0.6989700    1.0 regul_welfac
welfac15_1 1.0 * log(welfac15_1) =  0.9590414    1.0 regul_welfac
welfac16_1 1.0 * log(welfac16_1) = -0.6989700    1.0 regul_welfac
welfac1_2  1.0 * log(welfac1_2) = -0.3979400    1.0 regul_welfac
welfac2_2  1.0 * log(welfac2_2) = -6.0480E-02    1.0 regul_welfac
welfac3_2  1.0 * log(welfac3_2) = -0.3098039    1.0 regul_welfac
welfac4_2  1.0 * log(welfac4_2) =  1.255273     1.0 regul_welfac
welfac5_2  1.0 * log(welfac5_2) = -0.3979400    1.0 regul_welfac
welfac6_2  1.0 * log(welfac6_2) = -0.3979400    1.0 regul_welfac
* regularisation
  300.0 309.0          0.1000000
  1.0  1.0e-10  1.0e10  continue
  1.3  1.0e-2   1

```

TWDB Report ##: Final – Application of PEST to Re-Calibrate
the Groundwater Availability Model for the Edwards-Trinity (Plateau) and Pecos Valley Aquifers

Appendix I The model batch file

The batch file which is run repeatedly by PEST as the model is listed below.

```
REM *****  
REM Here is a list of executables that are run through this batch file.  
  
REM int2real  
REM vcontcalc  
REM fac2real  
REM twoarray  
REM real2cnd  
REM realstr2cnd  
REM welbuild2  
REM cliparray1  
REM mf96_d  
REM many2one_d  
REM mod2obs_d  
REM smpdiff  
REM strflbud_d  
REM smp2smp  
REM logarray  
REM arr2bore  
  
REM *****  
REM Intermediate files are deleted in the first part of the batch file.  
  
del hk_lay1.ref  
del hk_lay2.ref  
del s_lay1.ref  
del s_lay2.ref  
del rechmul_total.ref  
del draincond_lay1.ref  
del draincond_lay2.ref  
del ghb_lay1.ref  
del ghb_lay2.ref  
del strcond.ref  
del wel_tr.dat  
  
REM Zone based BCF real array files are deleted.  
  
del hk_lay1_zone.ref  
del hk_lay2_zone.ref  
del s_lay1_zone.ref  
del s_lay2_zone.ref  
del sy_lay1_zone.ref  
del sy_lay2_zone.ref  
del vcont_lay1_zone.ref  
del rf.ref  
  
REM Pilot point intermediate multiplier array files are deleted.  
  
del rechmul.ref  
del hk_lay1_mul.ref  
del hk_lay2_mul.ref
```

TWDB Report ##: Final – Application of PEST to Re-Calibrate
the Groundwater Availability Model for the Edwards-Trinity (Plateau) and Pecos Valley Aquifers

```
del s_lay1_mul.ref
del s_lay2_mul.ref
del sy_lay1_mul.ref
del sy_lay2_mul.ref
del vcont_lay1_mul.ref
del welmul_lay1.ref
del welmul_lay2.ref

REM MODFLOW text output files are deleted so that we can see if anything
REM is going wrong with MODFLOW.

del output_ss.dat
del output_tr.dat

REM Files passed between the steady state and transient model are deleted.

del heads_ss.dat
del sh1.ref
del sh2.ref
del sh3.ref

REM Transient model output files are deleted.

del heads_tr.dat
del budget_tr.dat

REM Arrays actually used by MODFLOW are deleted.

del trans1.ref
del trans2.ref
del sy1.ref
del sy2.ref
del vcont1.ref
del s1.ref
del s2.ref

del rech_ss.ref
del rech_av_1930s.ref
del rech_av_1940s.ref
del rech_av_1950s.ref
del rech_av_1960s.ref
del rech_av_1970s.ref
del rech_1980.ref
del rech_1981.ref
del rech_1982.ref
del rech_1983.ref
del rech_1984.ref
del rech_1985.ref
del rech_1986.ref
del rech_1987.ref
del rech_1988.ref
del rech_1989.ref
del rech_1990.ref
del rech_1991.ref
del rech_1992.ref
del rech_1993.ref
del rech_1994.ref
del rech_1995.ref
```


TWDB Report ##: Final – Application of PEST to Re-Calibrate
the Groundwater Availability Model for the Edwards-Trinity (Plateau) and Pecos Valley Aquifers

```
del rech_1996.ref
del rech_1997.ref
del rech_1998.ref
del rech_1999.ref

REM Other MODFLOW input files are deleted.

del drn.dat
del ghb.dat
del str1.dat

REM Files used in generation of regularization observations are deleted.

del ratio_median_hk_lay1.ref
del ratio_median_hk_lay2.ref
del log_ratio_median_hk_lay1.ref
del log_ratio_median_hk_lay2.ref

REM The zone-based component of BCF parameters is handled.

int2real.exe < int2real_hk_lay1.in
int2real.exe < int2real_hk_lay2.in
int2real.exe < int2real_s_lay1.in
int2real.exe < int2real_s_lay2.in
int2real.exe < int2real_sy_lay1.in
int2real.exe < int2real_sy_lay2.in

REM Pilot point multiplier parameters are handled.

fac2real.exe < fac2real_hk_lay1.in
fac2real.exe < fac2real_hk_lay2.in
fac2real.exe < fac2real_sy_lay1.in
fac2real.exe < fac2real_sy_lay2.in
fac2real.exe < fac2real_s_lay1.in
fac2real.exe < fac2real_s_lay2.in
fac2real.exe < fac2real_vcont_lay1.in
fac2real.exe < fac2real_rechmul.in

REM Zone-based BCF parameters are multiplied by pilot point multipliers

twoarray.exe < twoarray_hk_lay1.in
twoarray.exe < twoarray_hk_lay2.in
twoarray.exe < twoarray_s_lay1.in
twoarray.exe < twoarray_s_lay2.in
twoarray.exe < twoarray_sy_lay1.in
twoarray.exe < twoarray_sy_lay2.in

cliparray1.exe < cliparray1_hk_lay1.in
cliparray1.exe < cliparray1_hk_lay2.in

vcontcalc.exe < vcontcalc.in
twoarray.exe < twoarray_vcont_lay1.in
cliparray1.exe < cliparray1_vcont1.in

REM The pilot-point recharge multiplier array is built.

fac2real.exe < fac2real_rechmul.in
```

TWDB Report ##: Final – Application of PEST to Re-Calibrate
the Groundwater Availability Model for the Edwards-Trinity (Plateau) and Pecos Valley Aquifers

REM The recharge fraction multiplier array is built.

```
int2real.exe < int2real_rf.in
```

REM The two are combined.

```
twoarray.exe < twoarray_rechmul_total.in
```

REM Yearly recharge arrays are built.

```
twoarray.exe < twoarray_ss.in  
twoarray.exe < twoarray_av_1930s.in  
twoarray.exe < twoarray_av_1940s.in  
twoarray.exe < twoarray_av_1950s.in  
twoarray.exe < twoarray_av_1960s.in  
twoarray.exe < twoarray_av_1970s.in  
twoarray.exe < twoarray_1980.in  
twoarray.exe < twoarray_1981.in  
twoarray.exe < twoarray_1982.in  
twoarray.exe < twoarray_1983.in  
twoarray.exe < twoarray_1984.in  
twoarray.exe < twoarray_1985.in  
twoarray.exe < twoarray_1986.in  
twoarray.exe < twoarray_1987.in  
twoarray.exe < twoarray_1988.in  
twoarray.exe < twoarray_1989.in  
twoarray.exe < twoarray_1990.in  
twoarray.exe < twoarray_1991.in  
twoarray.exe < twoarray_1992.in  
twoarray.exe < twoarray_1993.in  
twoarray.exe < twoarray_1994.in  
twoarray.exe < twoarray_1995.in  
twoarray.exe < twoarray_1996.in  
twoarray.exe < twoarray_1997.in  
twoarray.exe < twoarray_1998.in  
twoarray.exe < twoarray_1999.in
```

REM Drain parameterization takes place.

```
int2real.exe < int2real_draincond_lay1.in  
int2real.exe < int2real_draincond_lay2.in  
real2cnd.exe < real2cnd_drain.in
```

REM Pilot point parameterization of GHB conductances takes place

```
fac2real.exe < fac2real_ghb_lay1.in  
fac2real.exe < fac2real_ghb_lay2.in  
real2cnd.exe < real2cnd_ghb.in
```

REM Pilot point parameterization of STR1 conductances takes place

```
fac2real.exe < fac2real_strcond.in  
realstr2cnd.exe < realstr2cnd.in
```

REM Composite parameter arrays for the use of the model are built.

```
twoarray.exe < twoarray_trans1.in  
twoarray.exe < twoarray_trans2.in
```

TWDB Report ##: Final – Application of PEST to Re-Calibrate
the Groundwater Availability Model for the Edwards-Trinity (Plateau) and Pecos Valley Aquifers

```
twoarray.exe < twoarray_s1.in
twoarray.exe < twoarray_s2.in

REM The transient MODFLOW well package input file is built
REM on the basis of pumping multipliers

fac2real.exe < fac2real_welmul_lay1.in
fac2real.exe < fac2real_welmul_lay2.in
welbuild2.exe < welbuild2.in

REM Bounds are enforced on some MODFLOW arrays.

cliparray1.exe < cliparray1_sy1.in
cliparray1.exe < cliparray1_sy2.in

REM The model is run.

mf96_d.exe < mf96_ss.in
many2one_d.exe < many2one.in
mf96_d.exe < mf96_tr.in

REM MOD2OBS is run to obtain model-generated heads.
mod2obs_d.exe < mod2obs_gt6b.in

REM SMPDIFF is run to obtain temporal head differences.
smpdiff.exe < smpdiff_gt6b.in

REM STRFLBUD is run in order to extract streamflow.
strflbud_d.exe < strflbud.in

REM The model-generated equivalent of observed streamflow is obtained.
smp2smp.exe < smp2smp.in

REM Processing is undertaken for assertion of Kh constraints at pilot point
REM locations in the upper two layers.

twoarray.exe < twoarray_regul_lay1.in
twoarray.exe < twoarray_regul_lay2.in
logarray.exe < logarray_hk_lay1.in
logarray.exe < logarray_hk_lay2.in
arr2bore.exe < arr2bore_regul_hk_lay1.in
arr2bore.exe < arr2bore_regul_hk_lay2.in

REM The following postprocessing is undertaken to try to prevent bad heads.

many2one_d.exe < many2one_d_ss.in
ptingrid.exe < ptingrid_ss.in
nonlinobs.exe
```

TWDB Report ##: Final – Application of PEST to Re-Calibrate
the Groundwater Availability Model for the Edwards-Trinity (Plateau) and Pecos Valley Aquifers

Appendix J The HPC cluster

J.1 General hardware description

An HPC cluster was constructed for the purpose of running PEST in parallel, allowing an order of magnitude increase in the speed at which optimization could be achieved. The cluster consisted of six off-the-shelf Dell Optiplex desktops with Core 2 Duo processors, interconnected by an 8-port switch, and connected to the WAN via a router. Table J-1 provides the basic hardware information. Figure H-1 illustrates the cluster as constructed.

Table J-1. Basic hardware description of cluster components

Item	Additional Description
6 Dell Optiplex 330N Desktops	
CPU	Pentium E7200 Core 2 Duo
Hard Drive	160 GB SATA
RAM	1.0 GB Non-ECC
Preinstalled OS	none
Flat Panel Monitor	
D-Link EBR-2310 Router	4 Port 10/100
D-Link DES-1108 Switch	8 Port 10/100
Small Steel Shelving Unit	
Various Cabling	Cat-5, Power, USB, Cable Ties

Because the processors are dual-core, they allow simultaneous execution of two forward simulations, with virtually no performance penalty. Thus, the cluster consists of 12 nodes. The total hardware cost per node was approximately \$300.

J.2 General software description

J.2.1 Operating system

Part of keeping the expenses to a minimum on the cluster included ordering computers that lacked an operating system, thus avoiding associated licensing fees. A Linux distribution, Fedora Release 9 was downloaded and installed on each of the nodes. A kickstart disc was used for an unattended initial installation on all of the nodes. Additional software installation and updates was performed using *yum* and looping through each node with a bash script.

J.2.2 Basic system configuration

The machines are named *twdb1-twdb6*, numbered from left to right in Figure H-1. The first machine, *twdb1*, has the only windowing system installed, for those users who prefer a point-and-click style desktop interface. This interface has not been rigorously tested, since local access to a cluster is often inconvenient and to be avoided unless a critical failure occurs. Typical remote access occurs through *ssh*.

A single user account was created on each machine with the following credentials:

```
Login:      pest
Password:   pestpass1
```

It is recommended that this password should be changed by TWDB staff upon receipt.

The root account is as follows:

```
Login:      root
Password:   rootpass1
```

Again, it is recommended that the root password be changed upon receipt.

RSA authentication files have been configured on each of the nodes so that *ssh* between nodes requires no password for these two accounts.

This cluster was designed to reside inside a secure internal network or LAN. That is, it does not have strict firewall settings appropriate for direct outward facing WAN connection (or DMZ connection). Although the only access possible is through *ssh* on port 22 (no other accessible services should be running, and the router only allows port 22 as an inward connection), good practice would require strict *iptables* settings, if outward facing WAN connection were desired.

Sometimes these strict security measures can interfere with simple communication between the nodes and make cluster administration more difficult. There is always a balance between security and convenience, and this cluster was designed for the latter, assuming, as noted previously, that it would sit inside a reasonably secure LAN.

J.2.3 WINE and native bytecode

WINE (a recursive acronym for WINE is not an emulator) is a set of compatibility libraries that allows most Windows executables to be run directly under Linux, without the need for recompilation or other changes. Basic FORTRAN executables compiled for Windows (or DOS) have been found to run flawlessly and without performance penalty under WINE.

WINE is installed on all of the cluster machines. Many of the PEST utilities and executables that run as part of the Edwards-Trinity optimization are Windows executables that run under WINE.

Of course, MODFLOW (and PEST) have versions compiled for Linux that may be used on this cluster, as well. A basic Fortran 95 compiler *gfortran* is installed on the first machine to allow Fortran compilation for those cases where it is necessary or convenient to do so.

J.2.4 File sharing

Parallel PEST requires a single *master* node to have access to a common directory on each *slave* node. Therefore, each *slave* node must have a shared directory that is visible to the master. On Linux systems, this directory sharing can be accomplished through either Samba or venerable NFS. If a cluster contains a mixture of Linux and Windows operating systems, then Samba is a

good choice since SMB is supported natively by Windows. In our case, because Linux is exclusive; thus, we chose NFS for its slight performance edge.

The directory structure is configured such that each node has a directory in which all of its tasks are carried out. Since two nodes exist on each computer, each computer has two directories with each directory being directly accessed by one of the two slave processes that are running on that computer. In addition, each of these directories is shared to the master node, such that the master node can communicate with the slave processes through these directories. Table J-2 shows a list of the directories and on which machines they reside.

Table J-2. Basic directory structure for running Parallel PEST

Computer	Node	Directory
twdb1	master	/pdata/data0/eddt
twdb1	1	/pdata/data1/eddt
twdb1	2	/pdata/data2/eddt
twdb2	3	/pdata/data3/eddt
twdb2	4	/pdata/data4/eddt
twdb3	5	/pdata/data5/eddt
twdb3	6	/pdata/data6/eddt
twdb4	7	/pdata/data7/eddt
twdb4	8	/pdata/data8/eddt
twdb5	9	/pdata/data9/eddt
twdb5	10	/pdata/data10/eddt
twdb6	11	/pdata/data11/eddt
twdb6	12	/pdata/data12/eddt

In this case, the *eddt* directory contains the files for the Edwards-Trinity optimization runs. If a different model were to be run with PEST, a new directory could be created at the same level (i.e., below the *data#* directories). Table I.2 shows that the first computer contains directories for each of the two nodes that reside on that computer, plus a directory *data0* that represents the master directory. Most of the modification and configuration should take place in that *data0* directory, and this is where the input and output from Parallel PEST resides. The node directories, *data1-data10* should be exact replicas, with the exception of the small changes made to model inputs and outputs by PEST during the course of the optimization run.

J.3 Running Parallel PEST on the cluster

J.3.1 Check the configuration

The first requirement for running Parallel PEST is to have a valid PEST configuration with all of the associated files and scripts in place. If possible, it is a good idea to test the PEST configuration by running a single iteration (in series mode, using *pest.exe*) in the *data0* directory and checking for errors from PEST or other programs in the execution chain. Once a valid PEST configuration is available, an *rmf* file must be added to allow parallel execution.

J.3.2 Mirror essential files to the node directories

All of the files that are an essential part of the execution chain must be copied from the *data0* directory to the node directories (i.e., *data1-data12*). This includes the model executables, the *tpl* and *ins* files, and the model data files that are not modified by PEST during the course of the optimization. Any PEST utilities that are needed as part of the execution chain should already be installed and in a directory accessible through the PATH environmental variable. If this is not the case, these utilities can simply be copied to each of the node directories for local execution.

J.3.3 Launch PSLAVE on all of the Nodes

PSLAVE is the slave process (*pslave.exe*) that runs on each node, pre- and post-processing necessary files on that node and launching the execution chain when triggered by PPEST from the *master* computer. This process can be launched on each node manually, or it can be automated using *ssh* as the fundamental communication protocol. An example script *launchSlaves* is provided that will automate this process.

A second script, *killSlaves*, will automatically locate and terminate slave processes. This can become necessary when a Parallel PEST run has gone awry and not finished normally (Parallel PEST will terminate the slaves automatically if the run finishes normally).

J.3.4 Running a test case

A test case resides in the *ppestex* directory in */pdata/data0*. The pest configuration file is called *test.pst*. To launch the test case (after *pslave* has been started on all of the nodes), enter:

```
ppest test
```

Successful execution indicates that the cluster is functioning correctly.

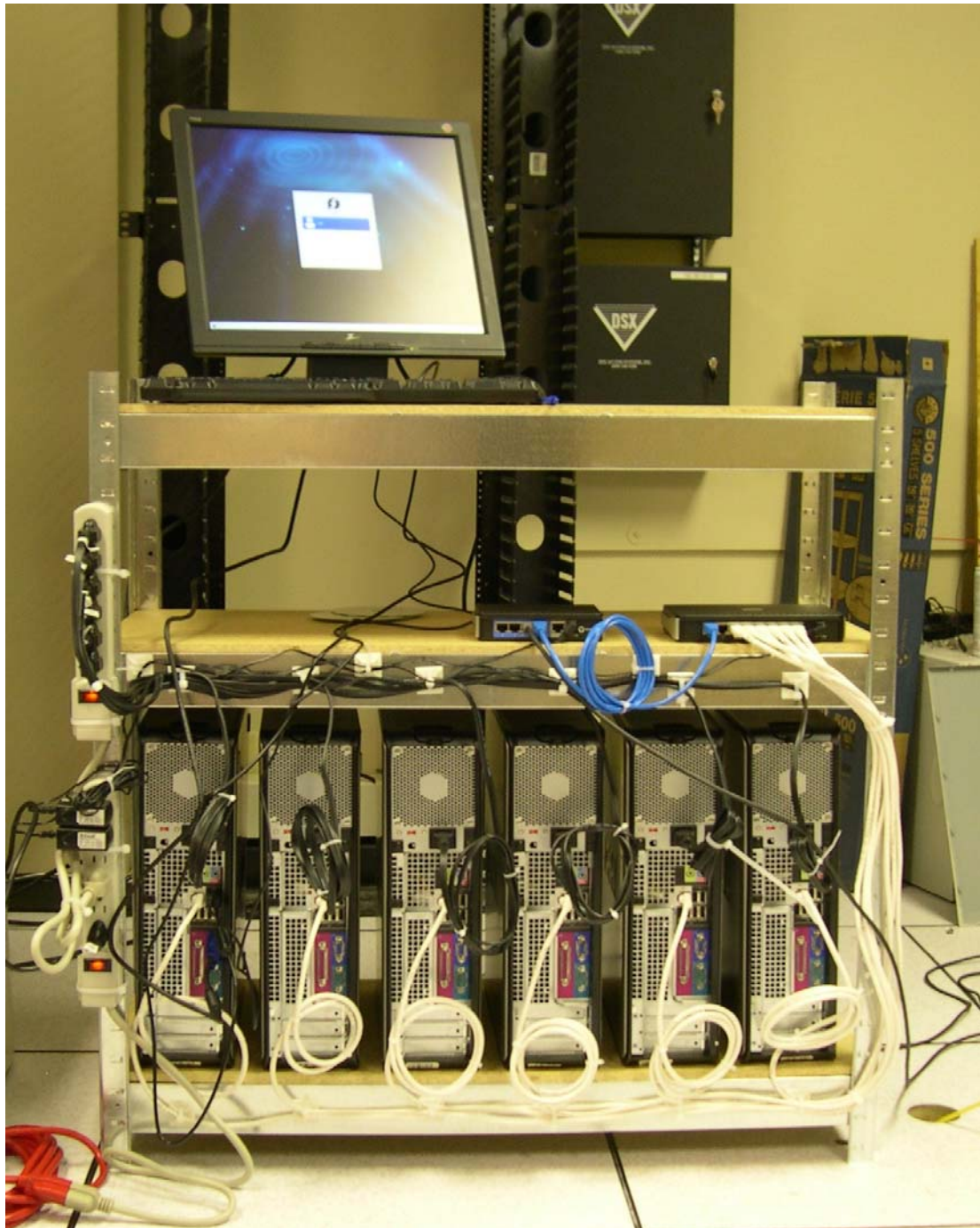


Figure J-1. The cluster and its basic cabling.

TWDB Report ##: Final – Application of PEST to Re-Calibrate
the Groundwater Availability Model for the Edwards-Trinity (Plateau) and Pecos Valley Aquifers

Appendix K TWDB comments on draft PEST application and recalibration report and responses

K.1 General comments

This is a well written, comprehensive report, with adequate supporting documentation. It provides explanations for the re-calibration process used. It was difficult to find significant problems with the re-calibration approach or the content of the report. Staff however did identify some minor corrections that are outlined below.

Response: No response required.

K.2 Specific Comments

1. Executive summary, paragraph 2: Please indicate that comparison is between error values of combined layers 1 and 2.

Response: Paragraph 2 has been modified to indicate that the water level measurements and corresponding error values are from model layers 1 and 2.

2. Section 2.1, page 3, paragraph 1: Please indicate that some of the 11 counties are in New Mexico.

Response: Paragraph 1 has been modified to indicate that some of the 11 counties are in New Mexico.

3. Section 2.2, page 3, paragraph 3: Please cite Anaya and Jones (2009) at the end of the first sentence.

Response: Paragraph 3 has been modified to include a citation for Anaya and Jones (2009).

4. Section 2.4, page 4, paragraph 5: Please indicate where the 500 hydrographs are included for the recalibration of the transient model using PEST are located.

Response: Paragraph 5 has been modified to indicate where the 500 hydrographs are included as part of the recalibration of the transient model using PEST.

5. Section 2.4, page 5, paragraph 2: Please correct grammar for last sentence of this paragraph “account for the how changes”.

Response: Paragraph 2 has been appropriately reworded.

6. Figure 2-1: Please update caption to reference the Edwards-Trinity (Plateau) and Pecos Valley aquifers instead of the “Edwards-Trinity Pecos Valley” aquifers.

Response: Figure 2-1 has been modified per the comment’s recommendation.

7. Figure 2-2: Please revise this figure to include the portion of the Pecos Valley Aquifer that overlies the Edwards-Trinity (Plateau) Aquifer within Reeves and Pecos counties.

Response: The figure has been revised to show the portion of the Pecos Valley Aquifer that overlies the Edwards-Trinity (Plateau) Aquifer.

8. Section 3.1.1, page 9: Please introduce the full name of acronyms such as Shuffled Complex Evolution (SCE), Covariance Matrix Adaptation Evolution Strategy (CMA-ES), singular value decomposition (SVD).

Response: Page 9 has been modified per the comment suggestion. The modification also introduce the full name to least-squares using QR factorization (LSQR).

9. Please move Chapters 3 and 4 to Appendix section(s).

Response: Chapters 3 and 4 have been moved to the Appendix Section.

10. Section 5.2.2, page 31, paragraph 5: 365,250 days is 1,000 years. Please revise the text to reflect this.

Response: The report was revised per the comment.

11. Section 5.2.2, page 31, paragraph 6: Please define or explain “type 2” status.

Response: Paragraph 6 has been modified to explain the meaning of a “Type 2” model layer.

12. Figures 5-2: Please indicate active cells in these figures. Only the study area boundaries and constant-head cells are shown.

Response: Figure 5-2 has been has been modified to show active cells.

13. Figures 5-2 and 5-3: Please indicate active cells in a darker shade for these figures to make it easier to visualize the “active” cells.

Response: The active cells in Figures 5-2 and 5-3 have been marked with a darker shade.

14. Section 5.2.2, page 31, paragraph 7: The shades in Figure 5-4 are not green. Please revise the text to reflect that

Response: Paragraph 7 has been modified to correctly reference the colors in Figure 5-4.

15. Figures 5-5 and 5-6: In the legend, please reverse the sorting of the intervals and flip the symbols such that highest pumping rates are indicated by the darkest shades

Response: Figures 5-5 and 5-6 have been changed per the comment.

16. Section 5.2.2, page 32, paragraph 1: Please change “5-5 and 5-6” to “5-7 and 5-8.”

Response: The reference to Figures 5-5 and 5-6 has been changed to Figures 5-7 and 5-8.

17. Section 5.3.1, page 33, paragraph 2: Please change “hundred-year” to “thousand-year”.

Response: Change from “hundred-year” to “thousand-year” has been effected.

18. Section 5.3.1, page 33, paragraph 2: Please provide table showing stress periods and corresponding time periods for re-calibrated model(s).

Response: Report was modified to include a table showing stress periods and corresponding time periods for the re-calibrated model(s).

19. Section 5.3.2, page 33, paragraph 3: Please clearly state the location of Tables B-3 to B-6.

Response: The report incorrectly references Tables B-3 to B-6. The report has been modified to say that the MODFLOW input and output files are listed in Appendix C.

20. Figure 6-3: This figure shows the extent of Layer 1, but data is applicable to both layers 1 and 2. Please add the outline of Layer 2 to the figure.

Response: Figure 6-3 has been modified to accommodate Layer 2 boundaries.

21. Section 6.1.2, page 47, paragraph 2: The sentence “The restriction of five wells ...” does not make sense. Please revise sentence to reflect a minimum of five well measurements rather than five wells.

Response: The sentence was revised to reflect a minimum of five well measurements rather than five wells.

22. Section 6.1.4, page 49, paragraph 1: The first sentence indicates three reasons for not using streamflow data for PEST calibration but only two reasons appear in the text. Please add the third reason or revise text to clearly show the three reasons.

Response: The text was changed to indicate there is two reasons and not three reasons for not including the streamflow data in the calibration criteria.

23. Section 6.2.1, page 51, paragraph 2: Please change reference from “Doherty et al. (2009)” to “(2009c)” and update in reference section

Response: The recommendation was made.

24. Section 6.5.1, page 57, paragraph 3, bullet 2: Please provide reference(s) for US Agricultural Survey reports in references section.

Response: The report has been modified to indicate that the agricultural reports are prepared by the United States Bureau of the Census. The referenced years are 1930, 1935, 1940, 1945, 1950, 1954, 1959, and 1964. This information provides is sufficient to locate the reference reports in any major library.

25. Section 6.5.1, page 57, paragraph 3, bullet 5: Please remove extra space between “200,” and “our estimated

Response: Extra space has been removed.

26. Section 8.5.1, page 97, Table 8-9: Please provide statistics per layer. Please add or revise table to provide statistics by model layer and also for the end of the transient simulation.

Response: The report has been modified to include two additional tables to provide the model calibration statistics per layer.

27. Figure 8-24, page 113: Please clarify and adjust legend in the figures to show Root Mean Square Error (RMSE) to be consistent with Figure 8-25.

Response: The legend on Figure 8-24 has been modified to show Root Mean Square Error (RMSE).

28. Figure 8-26, page 115: Please adjust caption to agree with figure. The legend in the figure suggests this is a comparison between RMSE of the original model compared with [RMSE?] of the re-calibrated model. Please confirm if this is a comparison of RMSE between the models. In addition, please provide a figure showing the residuals from the re-calibrated model with the location of the targets.

Response: The caption has been revised per the comment. The figure does provide a comparison of RMSE between models. A figure has been added to showing the RMSE of the residuals from the recalibrated model at the target well locations.

29. Figure 8-27, page 116: Please use a darker grey line to represent groundwater contribution to streamflow from recalibrated ETPV GAM.

Response: The grey line that represents groundwater contribution to streamflow in Figure 6-27 have been slightly darken.

30. Appendix C, C.3, page 166: Please clarify and adjust file name, as applicable, from “stardard.wel” to “standard.wel”.

Response: The change in spelling has been made.

31. Section 9.1, page 121, paragraph 2: Please remove the extra parenthesis in “(2009))”.

Response: The extra parenthesis has been removed.

32. Page 128, paragraph 1: Please change “Doherty, J., Fienen, M.N. and Hunt, R.J., 2009” to “Doherty, J., Fienen, M.N. and Hunt, R.J., 2009c

Response: The change has been made.

K.3 Comments on Data and Model Files

None

Response: No response required

K.4 Suggestions

1. Suggest adding several hydrograph samples from the 500 hydrographs of simulated versus measured water levels into the report.

Response: We consider this suggestion but decided against adding several hydrograph samples because of the problem of selecting a few “representative” examples and concerns over how some readers may interpret the significance of the comparison in regard to the overall model calibration goals. Readers who are interested in such comparison can plot the information from the model output files.

2. Suggest providing several hydrographs demonstrating the sensitivity of water-level fluctuations to changes in important hydrologic properties of the model.

Response: We did not follow-up on this suggestion. To adequately demonstrate and discuss this sensitivity would be a major undertaking.

3. Suggest adding a map figure of residuals for each of the two model layers.

Response: The report was modified to include figure showing the root-mean square error for the residuals at well locations in the two model layers.