Appendix A. An Introduction to the R-Package ‘wrv’

Contents

Figures

Tables

Introduction

R is a language and environment for statistical computing and graphics (R Core Team, 2014). In R, the primary mechanism for sharing with others is the *package*. Packages are collections of computer code, data, and documentation in a well-defined format. Instructions, datasets, and functions for processing and analyzing the groundwater-flow model of the Wood River Valley (WRV) aquifer system, south-central Idaho, are bundled together in an R package named **wrv**. This document is a *vignette* in the **wrv** package that describes an overview of processing steps for model construction. A package vignette is a LaTeX document with embedded R code; the code is run when the vignette is built, and all data analysis output (figures, tables, etc.) is created extemporaneously and inserted into the final document. Small chunks of stylized code are typically shown throughout a vignette and are intended to be used interactively. It is not necessary to have R-programming experience to follow the logic in these code chunks, but it may be useful for testing, development, and validation purposes. The **wrv** package includes multiple vignettes that explain and run all processing steps of model construction and analysis; the exception to this being the model-calibration process, which was not made programmatically reproducible, and executed outside of the R-programming environment. Model calibration is one aspect of model construction that was considered too arduous to implement in a reproducible manner because of its long run times.

Software

Software items needed to run the processing instructions include R, MODFLOW-USG, and PEST. If R (version $\geq$ 3.1) is not already installed on your computer, download and install the latest binary distribution from the Comprehensive R Archive Network (CRAN). Next, extend the capabilities of R by installing an assorted group of user-contributed packages available on CRAN and the Geological Survey R Archive Network (GRAN). That is, start an R session and type the following commands in your R-console window, or any other command-line interface where R is accessible (not required if the packages were previously installed):

```
repos <- c("http://owi.usgs.gov/R", getOption("repos"))
update.packages(ask = FALSE, repos = repos)
install.packages("wrv", repos = repos, dependencies = TRUE)
```

Once the packages are installed, load the **wrv** package in the current R session:

```
library("wrv")
```

Help documentation for functions and datasets in the **wrv** package (appendix B) are made accessible with the following command:

```
help(package = "wrv")
```

MODFLOW-USG is a computer program for simulating three-dimensional, steady-state and transient groundwater flow using a control volume finite-difference formulation (Panday and others, 2013). Source code and executable files for MODFLOW-USG (version 1.3) are provided in the **wrv** package.

PEST is a software suite that allows model-independent parameter estimation, sensitivity analysis, and uncertainty estimation, developed by Doherty (2005). If PEST (version $\geq$ 13.0) is not already installed on your computer, download and install the latest binary distribution and enable it to run from the command line.

Input/Output

A complete list of input-output file formats, organized by filename extension, is provided in table A1. Files that require additional clarification are described in table A2. All processing output (that is files and folders) are written to the current user-specified ‘working directory’. Specify an absolute path to the working directory below (change path as needed).

```
path <- file.path(getwd(), "SIR2016-5080")
dir.create(path, recursive = TRUE)
setwd(path)
```

Table A1. Input/output file formats.

Extension	Type	Description
.adf	binary	ArcGRID format, compressed in a ZIP file; raster graphic
.ba6	text	MODFLOW Basic Package File
.bat	text	Script file containing commands to execute
.bud	binary	MODFLOW Budget File
.csv	text	Comma-Separated Values; data table
.dis	text	MODFLOW Structured Discretization File
.drn	text	MODFLOW Drain Package
.exe	binary	MODFLOW compiled executable
.hds	binary	MODFLOW Head File
.kml	text	Keyhole Markup Language; spatial polygons
.lpf	text	MODFLOW Layer-Property Flow Package
.lst	text	MODFLOW List File
.nam	text	MODFLOW Name File
.oc	text	MODLFOW Output Control Option
.ptf	text	PEST Template File
.rda	binary	R datasets
.ref	text	Data reference file
.riv	text	MODFLOW River Package
.shp	binary	Shapefile, compressed in a ZIP file; spatial points, poly-lines, and polygons
.sms	text	MODFLOW Sparse Matrix Solver Package
.tif	binary	Geo-referenced tagged Image File Format; raster graphic
.wel	text	MODFLOW Well Package

Table A2. Files requiring additional clarification.

Name	Description
eff.csv	Irrigation efficiencies
hk1.ref, hk2.ref, hk3.ref	Hydraulic conductivity distribution in model layers 1, 2, and 3
mfusg.exe	MODFLOW-USG executable
model.rda	Multiple datasets describing the model grid, stress periods, and so forth
qa-incidental.csv	Quality assurance for incidental groundwater recharge on irrigated lands
qa-natural.csv	Quality assurance for natural groundwater recharge and discharge on non-irrigated lands
qa-pumping.csv	Quality assurance for groundwater diversions
qa-well-config.csv	Quality assurance for well configurations
RunModflow.bat	Command to run the groundwater-flow model
seep.csv	Canal seepage as a fraction of diversion
ss1.ref, ss2.ref, ss3.ref	Storage coefficient distribution in model layers 1, 2, and 3
trib.csv	Flow conditions in the major tributary canyons
UpdateBudget.bat	Command to update the water budget, requires access to R

Uncalibrated Model

Stepwise instructions for processing the uncalibrated groundwater-flow model are dependent on running R code within the following **wrv**-package vignettes: (1) appendix C, used to create **wrv**-package datasets from unprocessed data residing in a Git repository hosted on GitHub, and (2) appendix D, used to process, run, and analyze the results of the uncalibrated groundwater-flow model. Figure A1 shows a process flowchart for the interactions between these two vignettes.

Package Dataset Creation

The **wrv**-package datasets are created by running R code in the appendix C vignette (fig. A1). The resulting datasets from running this code are compared with existing package datasets and a warning given if differences are detected. These differences are likely the result of web-based data being out of synchronization with the archived datasets in this package. The following command runs the vignette's embedded R code; however, it requires an internet connection and about 10 gigabytes of memory, takes several hours to run, and has no effect on subsequent processing steps. Therefore, you may want to skip running these commands.

```
vignette("sir20165080_AppendixC")  # open appendix C
file <- system.file("doc", "sir20165080_AppendixC.R", package = "wrv")
source(file, echo = TRUE)  # or open file in a text editor and copy/paste into R console
```

Uncalibrated Model Construction

The uncalibrated model is constructed by running R code in the appendix D vignette (fig. A1). Output from this processing step is used as a template for a 'new' model archive. An archive folder named 'SIR2016-5080' is placed in the current working directory.

```
vignette("sir20165080_AppendixD")  # open appendix D
file <- system.file("doc", "sir20165080_AppendixD.R", package = "wrv")
source(file, echo = TRUE)  # or open file in a text editor and copy/paste into R console
```

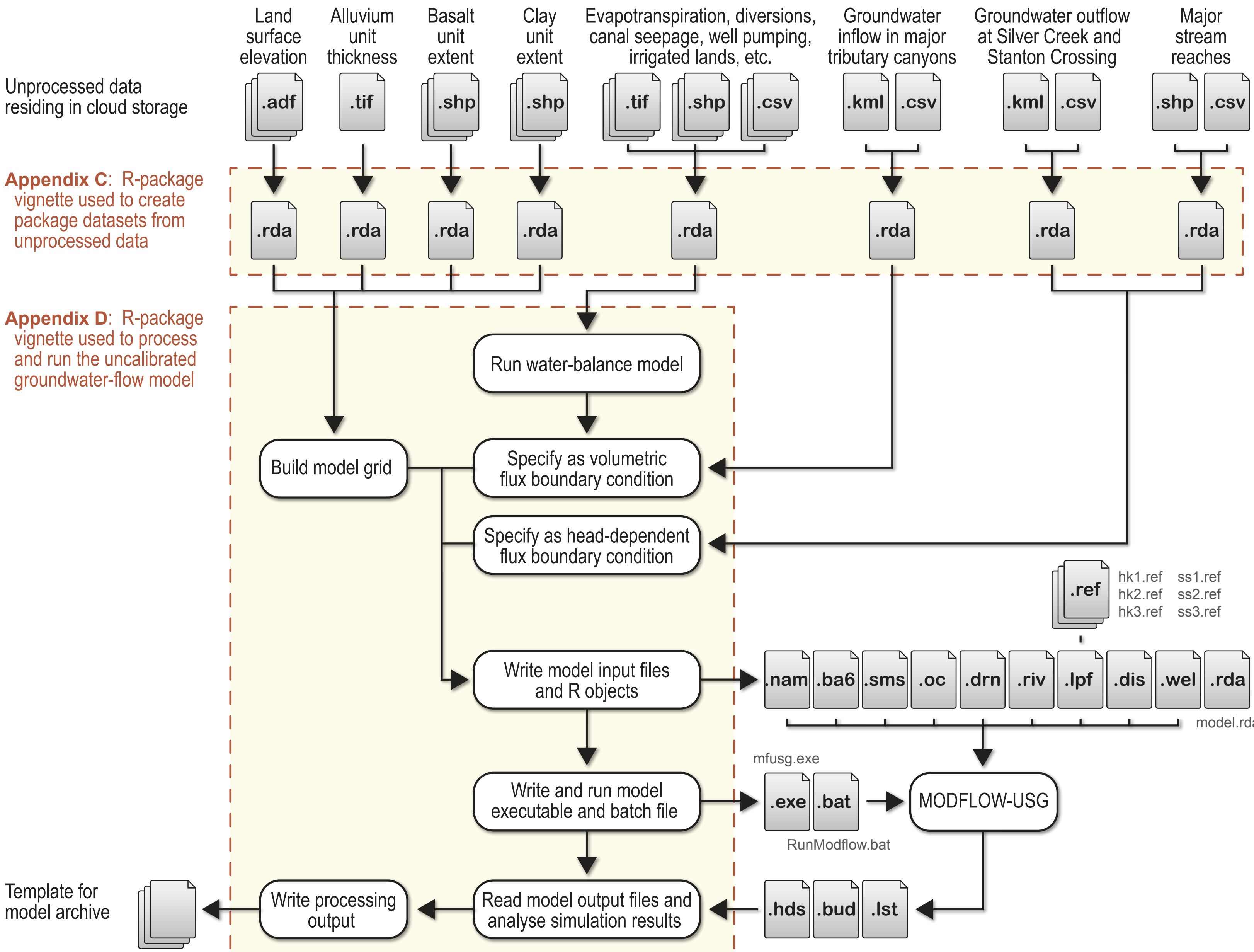

Figure A1. Procedures used to create the **wrv**-package datasets and process the uncalibrated groundwater-flow model.

Model Calibration

Most of the model-calibration processing steps were not made programmatically reproducible and are not easily documented within concise processing instructions. Nevertheless, a general description of these processing steps is provided in this vignette and thought to be adequate for understanding the model calibration workflow. Additional information regarding the model calibration setup for the WRV groundwater-flow model (such as, which parameters were adjusted through the calibration process, and the set of observations used to infer these parameters) is provided in appendix H.

Model calibration requires many individual PEST runs to finalize a set of believable model parameters that adequately minimize the model-to-measurement fit. Figure A2 shows the general processing steps for a PEST run. An iterative method is implemented by PEST to generate a sequence of improving parameter estimates. During each iteration of a PEST run, external calls are made to both the MODFLOW-USG program and an R function, specific to this study, that updates the *water budget* (`UpdateWaterBudget`). The water budget is an algorithm for calculating tributary basin underflow into the WRV aquifer system (appendix E), natural groundwater recharge and discharge on non-irrigated lands (appendix F), incidental groundwater recharge on irrigated lands (appendix G), and pumping demands (appendix G).

Many of the parameters adjusted during model calibration (such as the horizontal hydraulic conductivity) are contained within MODFLOW input files (including data reference files [‘.ref’] read by MODFLOW). At the end of each PEST iteration the newly updated parameter values are written to these model input files. The exception to this method of directly updating parameter values in the model input files occurs when the calibrated parameter values are used as input for a pre-processing program that generates a model input file(s). For example, in this study, the `UpdateWaterBudget` function (appendix B, p. B46-B48) is used to create the MODFLOW well file (‘.wel’), an input file containing specified flow boundary conditions. Input parameters for this function include, but are not limited to, the following: irrigation efficiency, tributary-underflow control parameters, and horizontal hydraulic conductivity—all of which are varied during the model-calibration process. Parameter values for irrigation efficiency and tributary-underflow are contained within the ‘eff.csv’ and ‘trib.csv’ files, respectively. And at the end of each PEST iteration the newly updated parameter values are written to these files. The horizontal hydraulic conductivity values are contained within data reference files (‘.ref’) and read by both MODFLOW-USG and the `UpdateWaterBudget` function.

The general procedure used when updating the water budget is show in figure A3. Prior to a PEST run, initialize the water-budget input files with parameter values specified for the uncalibrated model (appendix D), and output quality-assurance tables for the water-budget calculation:

```
help("UpdateWaterBudget")  # open help documentation for function call
UpdateWaterBudget("model/model1", "wrv_mfusg", qa.tables = "english")
```

The general processing steps for model calibration are shown in figure A4. Notice that the processing steps are represented in a linear workflow. This is an oversimplification of the approach taken for model calibration; in reality, the workflow was very non-linear because a new PEST run was required following any change in model conceptualization. Rather than starting the PEST run each time using the parameter set described for the uncalibrated model (appendix D), the optimized parameter set from the previous PEST run was instead used. This approach substantially reduced the overall computation time for model calibration, although it resulted in a set of model-calibration processing instructions that are not easily reproducible.

After completing the model-calibration process, manually update the model archive with the calibrated model files.

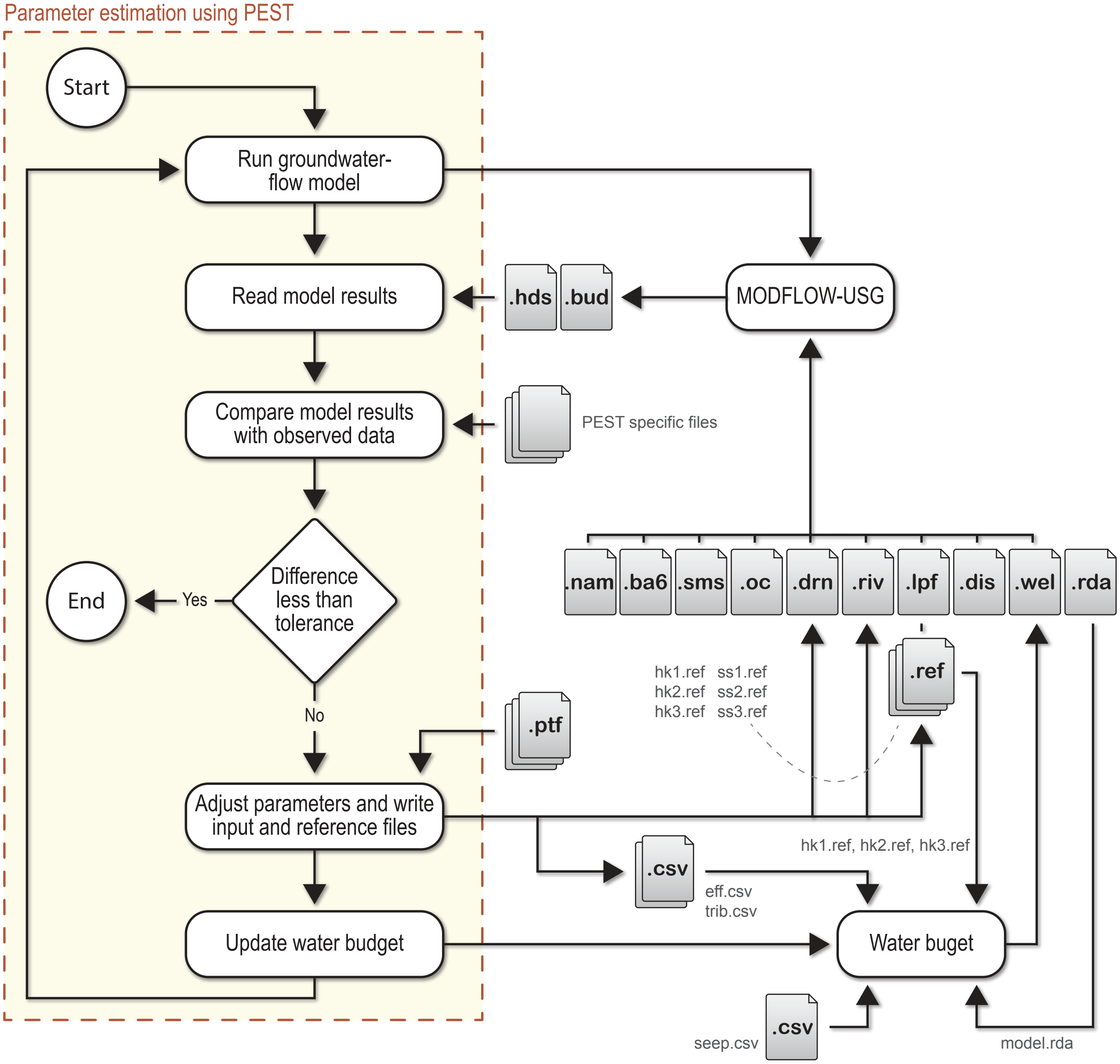

Figure A2. Procedures used in a single PEST run.

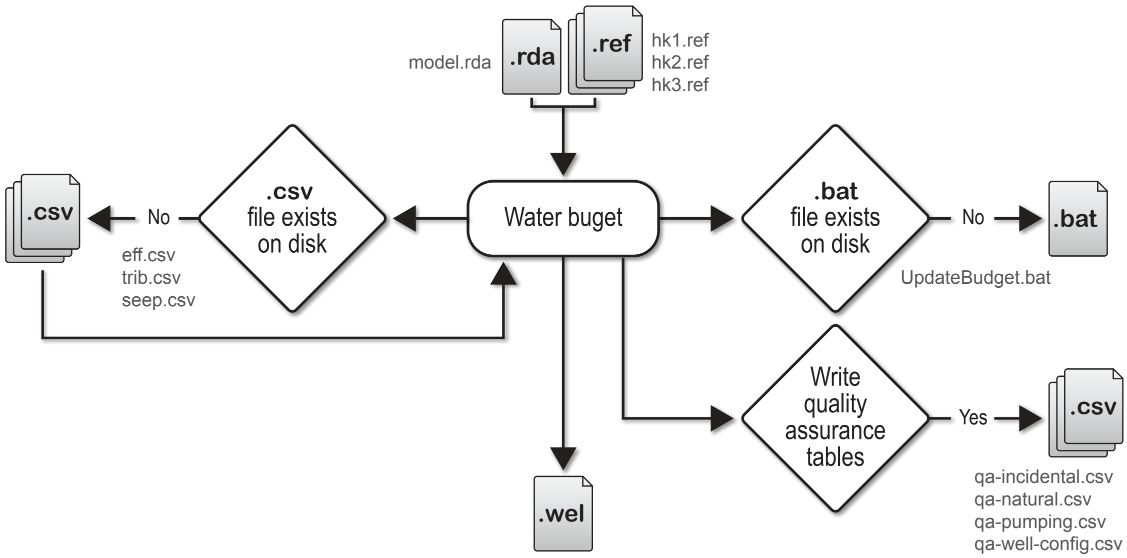

Figure A3. Procedures used when updating the water budget.

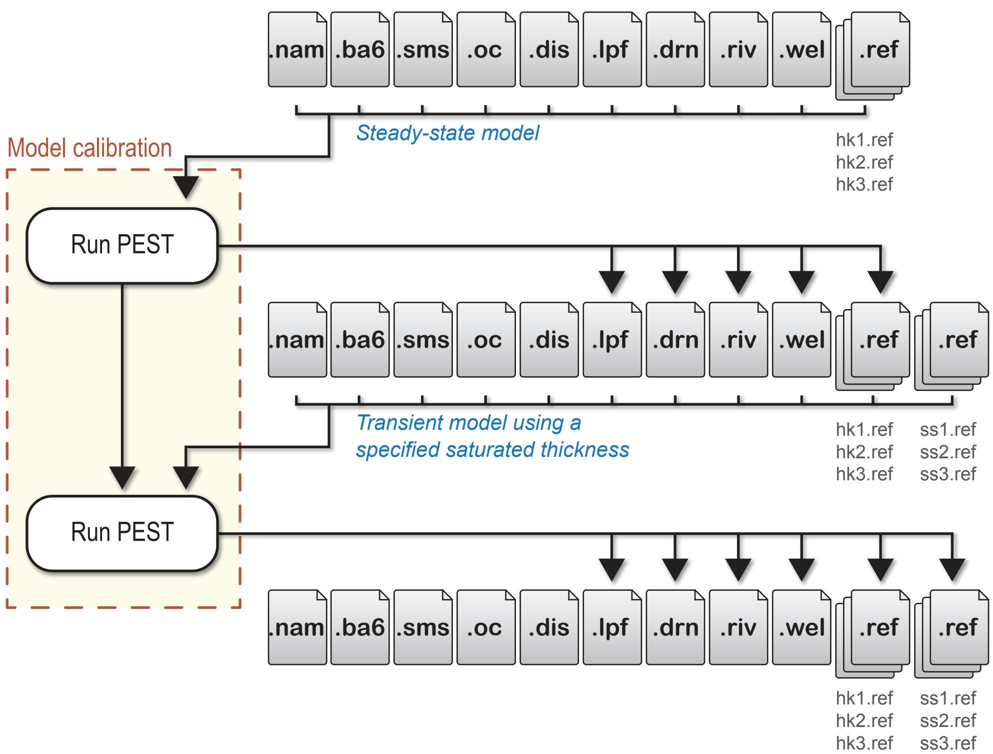

Figure A4. Procedures used in the model-calibration process.

References Cited

Doherty, J.E., 2005, PEST, model-independent parameter estimation–user manual, 5th ed., with slight additions, Watermark Numerical Computing, Brisbane, Australia.

Panday, Sorab, Langevin, C.D., Niswonger, R.G., Ibaraki, Motomu, and Hughes, J.D., 2013, MODFLOW-USG version 1: An unstructured grid version of MODFLOW for simulating groundwater flow and tightly coupled processes using a control volume finite-difference formulation, in U.S. Geological Survey Techniques and Methods 6-A45, accessed March 9, 2016 at http://pubs.usgs.gov/tm/06/a45.

R Core Team, 2014, R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, accessed March 9, 2016 at http://www.R-project.org/.