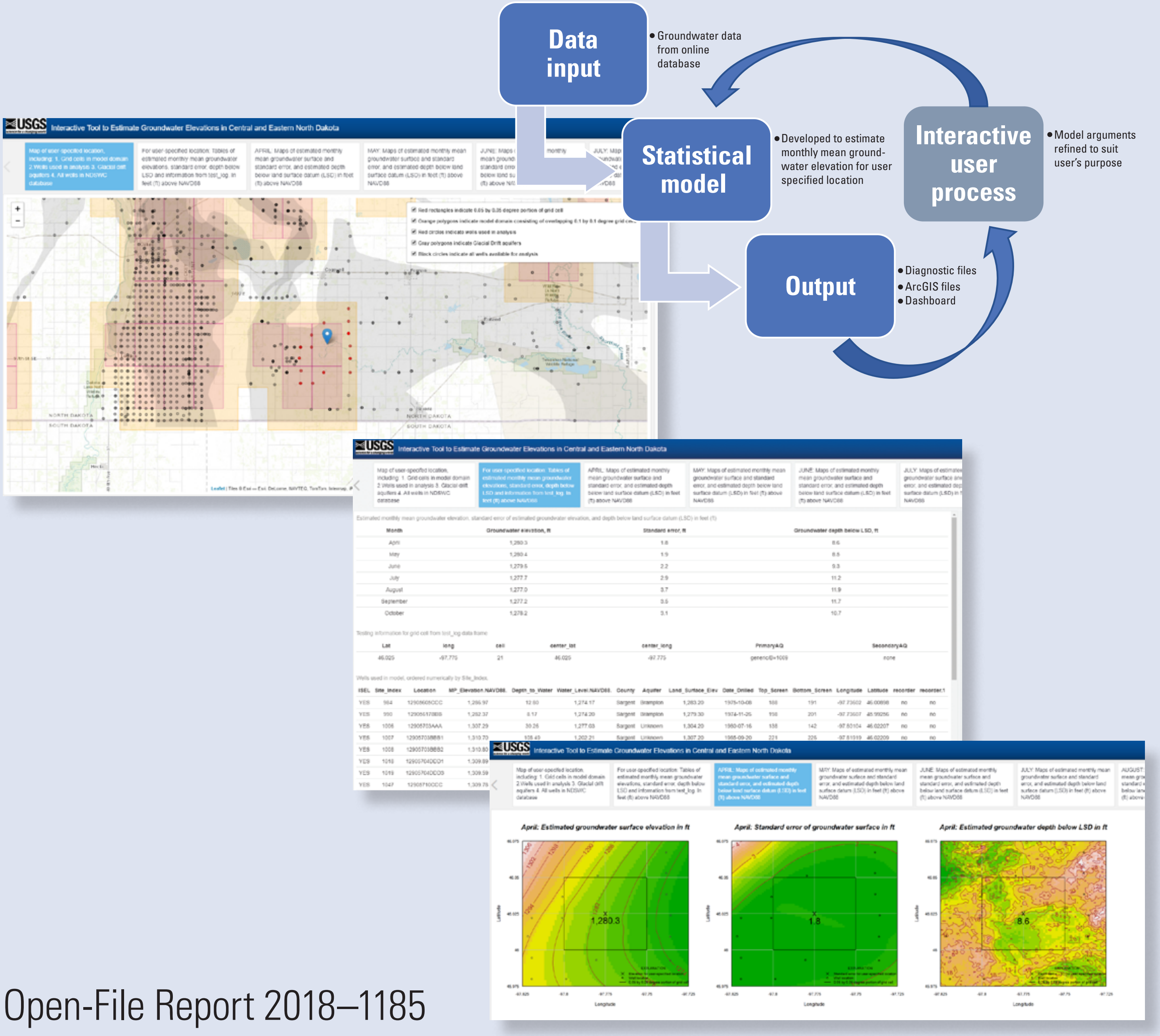
USGS
science for a changing world

Prepared in cooperation with Natural Resources Conservation Service

Interactive Tool to Estimate Groundwater Elevations in Central and Eastern North Dakota



Open-File Report 2018–1185

U.S. Department of the Interior
U.S. Geological Survey

Cover. Screenshots from the NDakGWtool of diagnostic output for a user-specified location with ample data, and diagram showing a general overview of the interactive tool, NDakGWtool.

Interactive Tool to Estimate Groundwater Elevations in Central and Eastern North Dakota

By Rochelle A. Nustad, William C. Damschen, and Aldo V. Vecchia

Prepared in cooperation with Natural Resources Conservation Service

Open-File Report 2018–1185

U.S. Department of the Interior
U.S. Geological Survey

U.S. Department of the Interior
RYAN K. ZINKE, Secretary

U.S. Geological Survey
James F. Reilly II, Director

U.S. Geological Survey, Reston, Virginia: 2018

For more information on the USGS—the Federal source for science about the Earth, its natural and living resources, natural hazards, and the environment—visit https://www.usgs.gov or call 1–888–ASK–USGS.

For an overview of USGS information products, including maps, imagery, and publications, visit https://store.usgs.gov.

Suggested citation:
Nustad, R.A., Damschen, W.C., and Vecchia, A.V., 2018, Interactive tool to estimate groundwater elevations in central and eastern North Dakota: U.S. Geological Survey Open-File Report 2018–1185, 24 p., https://doi.org/10.3133/ofr20181185.

ISSN 2331-1258 (online)

Acknowledgments

The feedback provided by Jon Petersen, Natural Resources Conservation Service hydrologist, as the tool was being developed is gratefully acknowledged. Michael Ginsbach, former North Dakota State Water Commission hydrologist, was very helpful in answering questions about data downloaded from the North Dakota State Water Commission's online database.

The authors also acknowledge Kelsey Kolars, former U.S. Geological Survey hydrologist, for her help in compiling the land-surface elevation dataset.

Contents

Figures

Tables

Conversion Factors

U.S. customary units to International System of Units

Multiply	By	To obtain
	Length	
foot (ft)	0.3048	meter (m)
mile (mi)	1.609	kilometer (km)

Datum

Vertical coordinate information is referenced to the North American Vertical Datum of 1988 (NAVD 88).

Horizontal coordinate information is referenced to the North American Datum of 1983 (NAD 83).

Abbreviations

LWR	locally weighted regression
NDSWC	North Dakota State Water Commission
NRCS	Natural Resources Conservation Service
USGS	U.S. Geological Survey

Interactive Tool to Estimate Groundwater Elevations in Central and Eastern North Dakota

By Rochelle A. Nustad, William C. Damschen, and Aldo V. Vecchia

Abstract

This report describes an interactive tool (NDakGW-tool) in which a statistical model is developed using locally weighted regression to estimate monthly mean groundwater elevations for a specified latitude and longitude, referred to as the "user-specified location." For each user-specified location, seven models are developed for each month from April through October. Localized, high spatial-resolution maps of estimated monthly mean groundwater surface elevations are produced from the models. The tool was evaluated for glacial drift aquifers of the 32-county study area in central and eastern North Dakota. Although groundwater elevations from 1960 to 2017 were available to develop the tool, groundwater elevations from 1995 to 2015 were used for model testing and development of the model domain. There are 413 grid cells of 0.1-degree latitude by 0.1-degree longitude size in the model domain, and the tool produces maps of estimated monthly mean groundwater surface elevations for the cell containing the user-specified location. Additionally, the NDakGWtool produces maps of estimated groundwater depth below land surface and ArcGIS files of estimated groundwater surface elevations and groundwater depth below land surface. The tool is composed of four main components: data input, statistical model, output, and user-interactive process.

Introduction

The Natural Resources Conservation Service (NRCS) commonly completes hydrologic analyses to determine the effects of proposed agricultural projects, for example, drain tile installation or irrigation withdrawals on shallow groundwater and surface-water resources. A criterion commonly used for their analysis is the 50-percent chance annual probability of inundation for 15 consecutive days during the wetland growing season (U.S. Department of Agriculture, 2015). As a result, in the hydrologic analysis, the NRCS commonly uses groundwater elevations in nearby observation wells to interpolate a seasonal minimum groundwater elevation for the specific location of the proposed project. In North Dakota, an online database hosted by the North Dakota State Water Commission (NDSWC) provides extensive spatial coverage of groundwater and surface-water data from various sources with about 8.5 million measurements for about 37,000 sites (fig. 1; North Dakota State Water Commission, 2017). Groundwater elevations are available from many types of wells, such as domestic, industrial, irrigation, municipal, observation, and production, but only about 10,000 groundwater sites have at least 1 groundwater measurement. Elevation is defined herein as the altitude of groundwater in feet above the North American Vertical Datum of 1988 (NAVD 88). For sites containing groundwater elevations, there is large variation in measuring periods, many wells have been discontinued or added, and measuring frequencies and seasonal coverage vary extensively. Furthermore, geologic and hydrologic conditions represented by a well network can vary substantially from well to well, and groundwater elevations are affected by spatially and seasonally variable recharge and discharge conditions and pumping. Culling pertinent groundwater elevations from millions of wells in the existing NDSWC database can be an onerous task, and, depending on the project location, availability of nearby wells, and complexity of the groundwater hydrology, it may be difficult to interpolate groundwater elevations in the location of a proposed project. To address this need, the U.S. Geological Survey (USGS), in cooperation with the NRCS, developed an interactive tool (NDakGWtool) to produce estimates of monthly mean groundwater surface elevations for a user-specified location within glacial drift aquifers for a 32-county study area in central and eastern North Dakota (fig. 1). The NDakGWtool will be useful to the NRCS and for future groundwater investigations by the USGS, NDSWC, and other agencies.

Purpose and Scope

The purpose of this report is to describe the development of an interactive tool (NDakGWtool) used to estimate monthly mean groundwater elevations in central and eastern North Dakota and to illustrate use of the tool. The tool was evaluated for glacial drift aquifers of a 32-county study area in central and eastern North Dakota and the 4 main components of the tool are data input, statistical model, output, and user-interactive process (fig. 2).

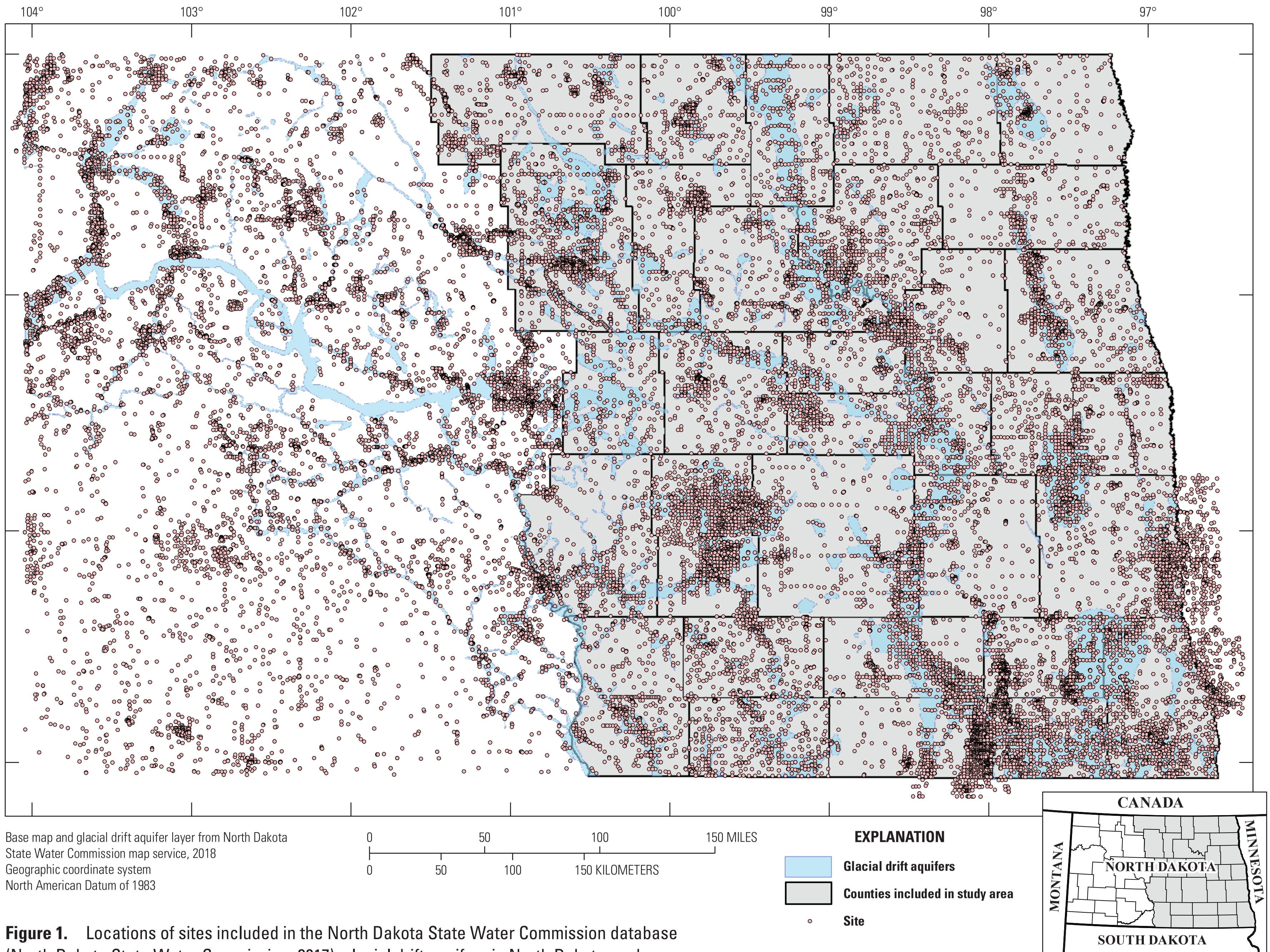


Figure 1. Locations of sites included in the North Dakota State Water Commission database (North Dakota State Water Commission, 2017), glacial drift aquifers in North Dakota, and 32 counties in central and eastern North Dakota included in the study area.

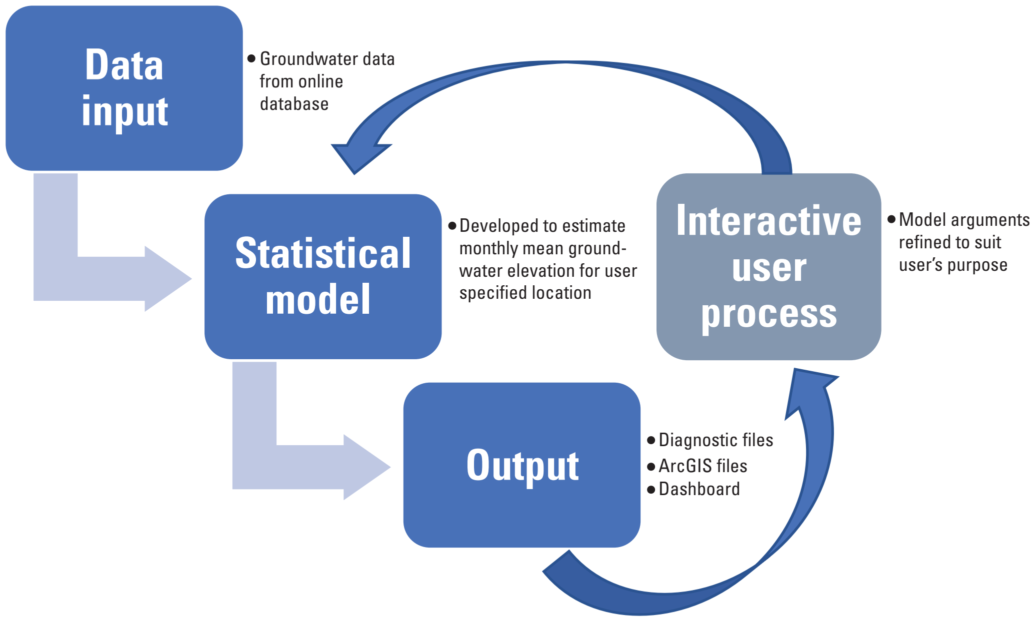


Figure 2. General overview of the interactive tool, NDakGWtool, to estimate groundwater elevations.

Overview of Interactive Tool

A detailed description of the four main components is provided in subsequent sections, but a general overview of the NDakGWtool is briefly provided here. The NDakGWtool is a statistical model that is developed using locally weighted regression (LWR) to estimate monthly mean groundwater elevations for a specified latitude and longitude, referred to as the "user-specified location."

Temporal and spatial coverages were defined during tool development. Temporal coverage of the tool consists of seven models that are developed for each month from April through October for the user-specified location. For each of these months, localized, high spatial-resolution maps of estimated monthly mean groundwater surface elevations are produced from the models. These maps can represent the water table surface of an unconfined aquifer or the potentiometric surface of a semiconfined aquifer. Consequently, maps of estimated monthly mean elevations will be referred to as "groundwater surface maps." Spatial coverage for the tool was evaluated for glacial drift aquifers of the 32-county study area. The 32-county study area was divided into 5,340 grid cells and, using the center point of each grid cell as the user-specified location, models for each grid cell were evaluated (herein referred to as "model testing"). Although groundwater elevations from 1960 to 2017 were available to develop the tool, groundwater elevations from 1995 to 2015 were used for model testing, and during model testing, the coverage of the tool was narrowed down to grid cells with sufficient data to develop a reliable model. Spatial areas of coverage, meaning output is available from the tool, will be referred to as the "model domain" (fig. 3). There are 413 grid cells in the model domain, and the tool produces maps of estimated monthly mean groundwater surface elevations for the cell containing the user-specified location. Additionally, the NDakGWtool produces maps of groundwater depth below land surface and ArcGIS (Esri, 2018) files of estimated groundwater surface elevations and groundwater depth below land surface.

Development of Interactive Tool to Estimate Groundwater Elevations

This section provides details on the components of NDakGWtool, development of those components, and use of the tool. The tool was developed in R (R Development Core Team, 2018), a programming and software environment for statistical computing. The tool contains various R packages, R scripts, R objects, files, and folders (table 1 and appendix), but to execute the tool, the primary R script model_run.R is used. Within this script, functions are called to create the groundwater elevation dataset, develop the model, and produce output (table 1). Within the model_run.R script, the code is broken into five sections, referred to as "BLOCKs" and each BLOCK corresponds to the different steps executed in the script. In this report, R functions are given in ***bold italics*** and arguments to functions or other R objects such as data frames are given in *italics*. An R data frame is an R object that holds tabular data, similar to a spreadsheet. A file with an extension ".Rdata" is an R data format that stores a collection of R objects.

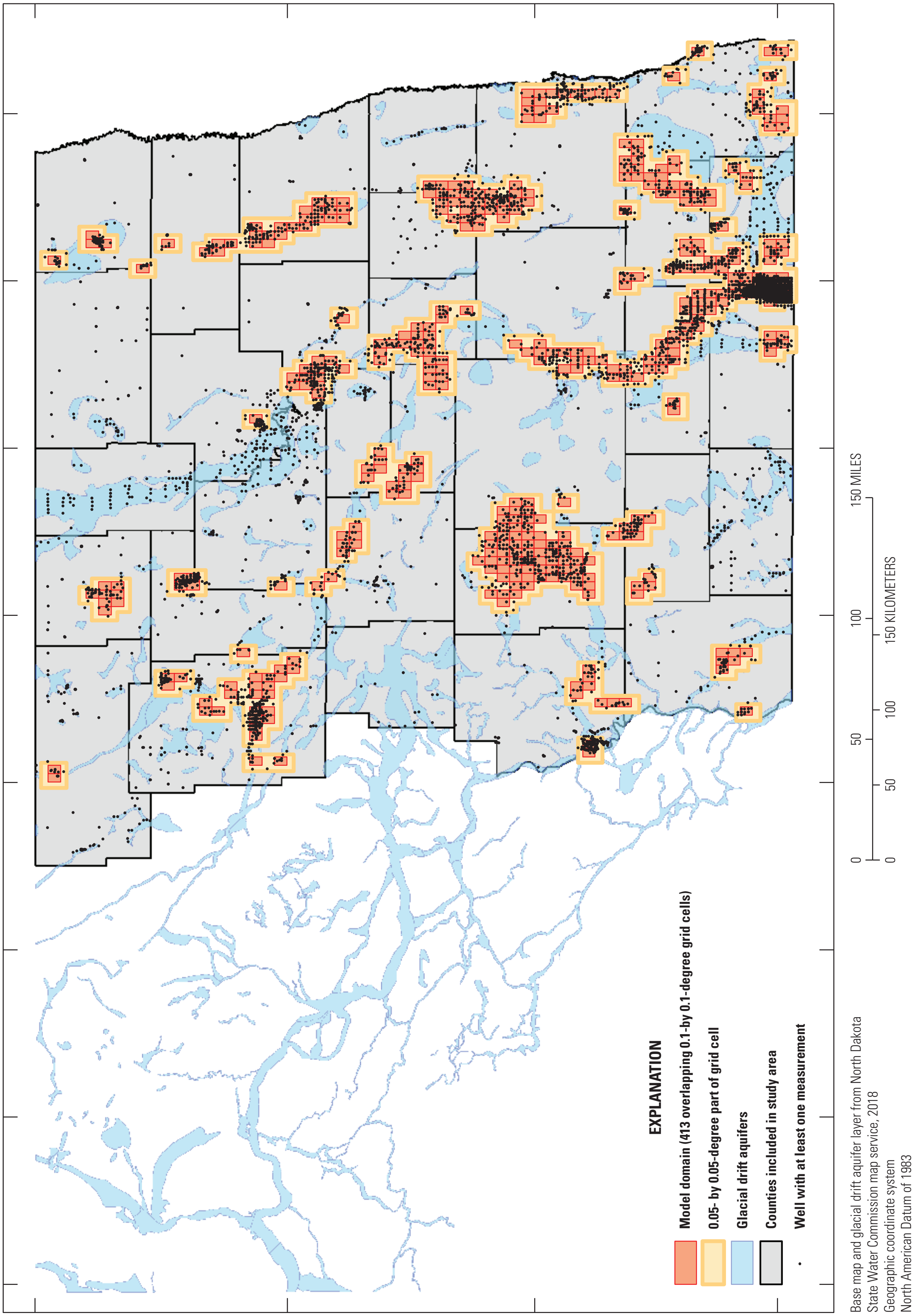


Figure 3. Wells in glacial drift aquifers of the study area with at least 1 groundwater elevation measurement from 1960 to 2017 and model domain that consists of 413 overlapping grid cells of 0.1-degree latitude by 0.1-degree longitude size. The 0.05-degree by 0.05-degree part of grid cells is highlighted.

Table 1. R packages, R scripts, R objects, files, and folders required to execute the interactive groundwater tool (NDakGWtool).

[Data are from R Development Core Team (2018). R functions are given in bold italics; arguments to functions or other R objects such as data frames are given in italics; GB, gigabyte; NDSWC, North Dakota State Water Commission; USGS, U.S. Geological Survey]

Name	Description	Source—The Comprehensive R Archive Network (CRAN)
	R packages (all descriptions directly from https://cran.r-project.org/)	
dplyr	"A fast consistent tool for working with data frame-like objects, both in memory and out of memory."	https://cran.r-project.org/web/packages/dplyr/index.html
raster	"Reading, writing, manipulating, analyzing and modeling of gridded spatial data."	https://cran.r-project.org/web/packages/raster/index.html
rgdal	"Provides bindings to the 'Geospatial' Data Abstraction Library ('GDAL') (>= 1.6.3) and access to projection/transformation operations from the 'PROJ.4' library."	https://cran.r-project.org/web/packages/rgdal/index.html
rmarkdown	"Convert R Markdown documents into a variety of formats."	https://cran.r-project.org/web/packages/rmarkdown/index.html
leaflet	"Create and customize interactive maps using the 'Leaflet' JavaScript library and the 'htmlwidgets' package."	https://cran.r-project.org/web/packages/leaflet/index.html
fields	"For curve, surface and function fitting with an emphasis on splines, spatial data and spatial statistics."	https://cran.r-project.org/web/packages/fields/index.html
flexdashboard	"Format for converting an R Markdown document to a grid oriented dashboard. The dashboard flexibly adapts the size of its components to the containing web page."	https://cran.r-project.org/web/packages/flexdashboard/index.html
knitr	"Provides a general-purpose tool for dynamic report generation in R using Literate Programming techniques."	https://cran.r-project.org/web/packages/knitr/index.html
data.table	"Fast aggregation of large data (for example, 100 GB in RAM), fast ordered joins, fast add/modify/delete of columns by group using no copies at all, list columns, friendly and fast character-separated-value read/write. Offers a natural and flexible syntax, for faster development."	https://cran.r-project.org/web/packages/data.table/index.html
readr	"The goal of 'readr' is to provide a fast and friendly way to read rectangular data (like 'csv', 'tsv', and 'fwf'). It is designed to flexibly parse many types of data found in the wild, while still cleanly failing when data unexpectedly changes."	https://cran.r-project.org/web/packages/readr/index.html
	R scripts	
model_run.R	R script used to execute the interactive tool. Calls the ***wlinputalldata***, ***wlmodelCell***, and ***gisdata*** functions to prepare the groundwater elevation dataset, develop statistical model, and produce output.	Appendix
analyses.Rmd	knitr script used to write summary report, called the dashboard, which contains a summary of output from the interactive tool.	Appendix

Table 1. R packages, R scripts, R objects, files, and folders required to execute the interactive groundwater tool (NDakGWtool).—Continued

[Data are from R Development Core Team (2018). R functions are given in bold italics; arguments to functions or other R objects such as data frames are given in italics; GB, gigabyte; NDSWC, North Dakota State Water Commission; USGS, U.S. Geological Survey]

Name	Description	Source—The Comprehensive R Archive Network (CRAN)
	R objects	
obs_wl	R object containing groundwater elevations to develop statistical models, generated from ***wlinputalldata*** function. The *obs_wl* object containing groundwater elevations from 1960 to 2017 is included in model_data.Rdata.	Appendix—model_data.Rdata
test_log	R object generated from model testing, included in model_data.Rdata.	Appendix—model_data.Rdata
lse	R object containing land-surface elevation, in feet (source, National Elevation Dataset), included in model_data.Rdata.	Appendix—model_data.Rdata
	Files	
wlinputalldata.R	File containing ***wlinputalldata*** function, which is used to manipulate and compile files from online groundwater database. This function produces the data frame *obs_wl*. An *obs_wl* data frame containing groundwater elevations from 1960 to 2017 is available in the appendix, but this function can be used if the user chooses to update groundwater elevations.	Appendix
wlmodelCell.R	File containing ***wlmodelCell***, a function used to develop a statistical model for the user-specified location and corresponding diagnostic files.	Appendix
model_data.Rdata	File containing data frames *obs_wl*, *test_log*, and *lse*.	Appendix
gisdata.R	File containing ***gisdata***, a function used to convert model output into ArcGIS file format, and executes the analyses.Rmd script.	Appendix
sites.txt	File generated from manually downloading data from NDSWC online.	Appendix—Instructions in model_run.R, BLOCK 2
water_levels.txt	File generated from manually downloading data from NDSWC online.	Appendix—Instructions in model_run.R, BLOCK 2
USGS_ID_black.png	File containing USGS logo for placement on dashboard file.	Appendix
	Folders	
GIS	Shapefiles used in the analyses.Rmd script are stored in this folder.	Appendix
OUTPUT	Model output generated from executing model_run.R is saved in this folder.	Appendix

Data Input

Groundwater elevations and land-surface elevations are required to produce output from the NDakGWtool. Groundwater elevations from 1960 to 2017 are contained in the *obs_wl* data frame, and land-surface elevations for the model domain are contained in the *lse* data frame (table 1). These two data frames are stored in the model_data.Rdata file and are available in the appendix. Instructions to load the model_data.Rdata file are in BLOCK 3 of the model_run.R script (table 1, appendix). Groundwater elevations can be updated by manually downloading the raw dataset and executing the ***wlinputalldata*** function to write a new *obs_wl* data frame. Instructions to update groundwater elevations are provided in BLOCK 2 of the model_run.R script (appendix).

The groundwater elevation dataset for this report was downloaded in June 2017 from the NDSWC online database (North Dakota State Water Commission, 2017). All available data were downloaded within the 32-county study area. From these data, two text files were generated, sites.txt and water_levels.txt (table 1). These two files comprise the raw dataset and contain data for groundwater and surface-water sites. The ***wlinputalldata*** function combines these two files into one dataset, removes surface-water monitoring sites, selects observation wells in the study area with at least one groundwater measurement since 1960, and eliminates wells missing a measuring point elevation. Groundwater measurements from wells in the raw dataset have measuring frequencies that range from about monthly to continuous, although most wells have a monthly measuring frequency. The measuring frequency generally extends from March or April, depending on the timing of the spring thaw, through October or November. For wells with continuous recorders, the ***wlinputalldata*** function selects the last measurement in the month to make them comparable to the wells that have about monthly measurements. Finally, the ***wlinputalldata*** function writes selected measurements to the *obs_wl* data frame. The *obs_wl* data frame is the dataset used in developing models for the user-specified location.

Land-surface elevation was acquired for grid cell points in the model domain and used to produce maps of estimated depth of groundwater below land surface. These maps are produced from the ***gisdata*** function after models have been developed. Land-surface elevations (1-arc-second resolution) within the model domain were acquired from the USGS National Elevation Dataset and selected for model grid points that correspond to the maps of estimated groundwater elevation (U.S. Geological Survey, 2017). The ***gisdata*** function subtracts the estimated groundwater elevation from the land-surface elevation for each grid cell point to compute depth of groundwater below land surface.

Statistical Model for Estimating Groundwater Elevations

A statistical model is developed for each month from April through October using LWR. In the LWR, groundwater elevations in the *obs_wl* data frame are the response variable, and the latitude and longitude of the well locations are the explanatory variable. For the grid cell of the user-specified location, the ***wlmodelCell*** function (table 1) selects groundwater elevations, fits an LWR model to those elevations, screens data for outliers, and ultimately produces a localized, high spatial-resolution map of estimated monthly mean groundwater surface elevations. The ***wlmodelCell*** function is executed from BLOCK 4 of the model_run.R script (table 1, appendix). In the following paragraphs, the internal computations of ***wlmodelCell*** and external requirements, or arguments, for running the function are described.

Internal computations of the ***wlmodelCell*** function are completed on available groundwater elevations by grid cell. Grid cells are overlapping blocks of 0.1 by 0.1 degree, about 4.5 by 7 miles (mi), centered on 0.05- by 0.05-degree nodes (fig. 3). The grid cell size was chosen to balance the need for spatial detail using available elevations while considering the highly complex and spatially variable nature of groundwater flow. Each grid cell contains 6,400 grid points that are 0.00125 by 0.00125 degree or about 131 by 295 feet (ft). Based on the user-specified location, groundwater elevations for the grid cell in the model domain closest to that location are selected. If the user-specified location lies within the model domain, subsequent steps in the ***wlmodelCell*** function are completed. If the user-specified location is outside the model domain the user gets the following error message, “Specified location not in model domain.” If a cell does not contain 3 or more observation wells and has at least 30 groundwater elevations for all wells, output is not produced. These are minimum requirements, but most cells had considerably more data.

Once the ***wlmodelCell*** function has determined that sufficient data are available for the user-specified location, the next steps involve using the base R function, ***loess*** (R Development Core Team, 2018), to fit an LWR model. In the ***loess*** function, measured groundwater elevations are defined as the response variable, and the latitude and longitude of well locations are defined as locally fitted explanatory variables. A separate model is created for each month. For estimating monthly mean groundwater elevations for the end of a specified month, measured groundwater elevations for the 12-week (about 3-month) period centered on the last day of the month are used. A third explanatory variable, measurement interval, is defined in the ***loess*** function and is the difference between the measurement date and the last day of the specified month. This third explanatory variable was fitted globally using the *parametric* argument in the ***loess*** function (R Development Core Team, 2018) and was used to adjust the measurement dates to represent estimated groundwater elevations for the end of the month. For example, for estimating the monthly mean elevation for April 30, measured groundwater elevations from mid-March through mid-June are used. By globally fitting the measurement interval, adjustments are made for variable observation times and short-term trends within the 3-month window. Additional arguments, *degree* and *span*, were specified for the ***loess*** function (R Development Core Team, 2018). The *degree* argument refers to the degree of polynomials to be used or the type of model, where a *degree*=1 indicates linear and a *degree*=2 indicates quadratic. The *span* argument controls the smoothing of the data or the fraction of the data to use for each window, where a *span*=1 produces the smoothest estimates and smaller values produce more locally variable estimates. Within the ***wlmodelCell*** function, *degree* and *span* are automatically determined for the ***loess*** function based on criteria of density and distribution of wells in the grid cell. A *degree*=2 and *span*=0.75 are indicative of a model developed for a grid cell with high density and even distribution of wells, and *degree*=1 and *span*=1 are indicative of a model developed for a grid cell with low density and (or) a highly clustered well location. For grid cells with spotty or variable seasonal sampling frequencies, the *degree* and *span* can vary by month, but they will usually be the same for each month.

Further steps in the ***wlmodelCell*** function use the optional ***loess*** argument, *family*, to automatically remove outliers, and once outliers are removed, the base R function ***predict.loess*** is used to estimate groundwater elevations for each point in the model grid cell (R Development Core Team, 2018). Outliers are common in groundwater elevations and are caused by several factors: if a topographic map is used to determine well location and measuring point, accuracy of the groundwater measurement may be poor; some wells demonstrate highly localized behavior and, as such, the groundwater measurement may not be representative of a spatially coherent and relatively smooth water surface; data transcription errors for some groundwater measurements may result in incorrect values; or extreme events such as frozen or partially frozen soils, local snowmelt, recharge, or precipitation may result in measurements that are not representative of typical conditions. These types of outliers were removed to suit the needs of the NRCS for their hydrologic analysis and to improve model stability. An automated screening procedure was used to detect and remove as many outliers as possible by fitting the initial model using the *family = "symmetric"* argument in the ***loess*** function. The *family = "symmetric"* argument uses the robust M-estimation method, which is a common and effective method for removing outliers (Cleveland and others, 1992). Outliers, specifically defined as elevations with an absolute value of the residual greater than 2.5 standard errors, are removed and the model is fitted a second time with the *family* = "*symmetric"* option. Additional outliers, if there are any, are once again removed. Finally, groundwater elevations remaining after the two outlier screenings are used to fit a final model using the *family* = "*gaussian"* argument or ordinary least squares. After removing potential outliers and fitting the final model to groundwater elevations for the grid cell, the function ***predict.loess*** is used to estimate monthly mean elevations and the standard error of the estimated elevations for each grid point contained within the grid cell of the user-specified location (table 1, appendix).

As previously mentioned, the user executes the ***wlmodelCell*** function in BLOCK 4 of the model_R script. Available arguments to the ***wlmodelCell*** function include 3 required arguments and 3 optional arguments. Required arguments include *runname*, a prefix attached to output files; *latsp*, the user-specified latitude for the location of interest, in decimal degrees; and *lonsp*, the user-specified longitude for the location of interest, in decimal degrees (table 2). For user convenience, the required arguments for the ***wlmodelCell*** function can be populated in the dummy variables of BLOCK 4 in model_R.script rn for *runname* (line 89); lat for *latsp* (line 92); and lon for *latsp* (line 93).

The first optional argument, *spaqname0*, is included in the ***wlmodelCell*** function to allow the user to interactively modify the set of wells used in the models (table 2). If the user chooses to define the *spaqname0* argument, this also is in BLOCK 4 of the model_R.script, but the user populates this argument directly within ***wlmodelCell*** function, on line 100. This optional argument may be helpful to the user because grid cells can contain a mixture of wells with groundwater elevations from different aquifers, and, moreover, wells in the same aquifer can represent a mixture of unconfined or semiconfined layers, even if the wells are near one another. If *spaqname0* is omitted from the argument list or *spaqname0* = "default" is specified, the default value described later in the "Model Testing" section is used.

The argument *spaqname0* is a character variable with the following format:

spaqname0 = "aqname/spdepth/spbadwells/splevels"

where

aqname = "generic" indicates that all wells (regardless of aquifer name) should be included;

aqname = "Oakes" (or any other user-specified aquifer name) indicates that only wells in the Oakes aquifer (or any other user-specified aquifer name) should be included;

spdepth = "*D*>*d*," where *d* is a user-specified numeric constant, indicates that only wells with screen-interval depths (*D*) greater than *d* ft should be included;

spdepth = "*D*<*d*," where *d* is a user-specified numeric constant, indicates that only wells with *D* less than *d* ft should be included;

spbadwells = "*B*=*i*1, *i*2,…, *iK*," where *i*1, *i*2,…, *iK* are *K* (*K*≥1) user-specified well numbers, indicates that wells (*B*) with the user-specified well numbers should be excluded;

splevels = "*L*>*w*," where *w* is a user-specified numeric constant, indicates that only groundwater elevations (*L*) greater than *w* ft should be included; and

splevels = “*L*<*w*,” where *w* is a user-specified numeric constant, indicates that only *L* less than *w* ft should be included.

The aquifer name (aqname) must be specified, but the modifiers (terms involving *D*, *B*, or *L* separated by “/”) are optional. The first two characters of each modifier (“*D*>” or “*D*<”, “*B*=”, “*L*>” or “*L*<”) must be exactly as specified, with capital letters and no spaces, and the well indices are separated by commas (no spaces). For example,

spaqname0 = “generic” selects all groundwater elevations regardless of aquifer name;

spaqname0 = “generic/*D*>100” selects only wells with screen-interval depths greater than 100 ft;

spaqname0 = “generic/*D*>100/*B*=2417,2425” excludes wells with index number 2417 or 2425 (in addition to *D*>100);

spaqname0 = “Spiritwood” selects only wells in the Spiritwood aquifer; and

spaqname = “Spiritwood/*L*>1300” selects only wells in the Spiritwood aquifer and uses only groundwater elevations greater than 1,300 ft.

The other two optional arguments, *yrbeg* and *yrend* (table 2), can be used to override the default period of 1995–2015 used to estimate the groundwater elevations; for example, specifying *yrbeg*=2005 and *yrend*=2015 uses groundwater elevations for 2005–15 to fit the model.

Output

Three types of output are produced from the NDakGWtool and include (1) diagnostic files, in .pdf and .csv file formats; (2) ArcGIS files, in .tif and .shp file formats; and (3) a static summary file, in .html file format, called the dashboard (table 3). Diagnostic output is generated by the ***wlmodelCell*** function, and files are produced by running BLOCK 4 of the model_run.R script. The other two types of output are generated by the ***gisdata*** function (table 1, appendix) and are produced by running BLOCK 5 of the model_run.R script. The *runname* argument defined by the user in BLOCK 4 of the model_run.R script is attached as a prefix to all output file names. Output files are stored in a folder labeled

Table 2. Arguments used in the ***wlmodelCell*** function, which is used for developing a locally weighted regression model for estimating monthly mean groundwater elevations.

[Arguments to functions or other R objects such as data frames are given in italics]

Argument name	Description	Default
	Required arguments	
runname	Character variable specifying the run name for output diagnostic files and other information.	None
latsp	User-specified latitude in decimal degrees for the location of interest.	None
lonsp	User-specified longitude in decimal degrees for the location of interest.	None
	Optional arguments	
spaqname0	*spaqname0* = “aqname/spdepth/spbadwells/splevels” where aqname = “default” uses the value of spaqname0 from *test_log* data frame for the primary aquifer column; aqname = “generic” indicates that all wells (regardless of aquifer name) should be included; aqname = “Oakes” (or any other user-specified aquifer name) indicates that only wells in the Oakes aquifer (or any other user-specified aquifer name) should be included; spdepth = “*D*>*d*,” where *d* is a user-specified numeric constant, indicates that only wells with screen-interval depth (*D*) greater than *d* feet should be included; spdepth = “*D*<*d*,” where *d* is a user-specified numeric constant, specifies that only wells with *D* less than *d* feet should be included; spbadwells = “*B* = *i*1, *i*2, …, *iK*,” where *i*1, *i*2, …, *iK* are *K* (*K* ≥ 1) user-specified well numbers, indicates that wells (*B*) with the user-specified well numbers should be excluded; splevels = “*L*>*w*,” where w is a user-specified numeric constant, indicates that only measurements of groundwater elevation (*L*) greater than *w* feet should be included; and splevels = “*L*<*w*,” where *w* is a user-specified numeric constant, indicates that only *L* less than *w* feet should be included.	Defined in *test_log* but can be refined by the user
yrbeg	Beginning year for analysis; default is set but can be changed by the user.	1995
yrend	Ending year for analysis; default is set but can be changed by the user.	2015

Table 3. Output files from the NDakGWtool.

[R functions are given in ***bold italics***]

Output file name[1]	Output type	Output produced by function	Description
runname_fit.pdf	Diagnostic	***wlmodelCell***	Contains a plot of model fit, maps of groundwater surface, and standard errors for the grid cell containing the user-specified location for each month.
runname_img.pdf	Diagnostic	***wlmodelCell***	A detailed view of the maps of groundwater surface and standard errors for the grid cell containing the user-specified location for each month.
runname_near.pdf	Diagnostic	***wlmodelCell***	Contains a time-series plot of monthly mean groundwater elevations for wells used in developing the models.
runname_wells.csv	Diagnostic	***wlmodelCell***	Contains a list of all wells within the grid cell.
runname_wl_mm.tif	ArcGIS	***gisdata***	Raster image containing estimated monthly mean groundwater elevations, in feet, for each month. When the file is produced, "mm" is replaced with a number representing the month.
runname_serr_mm.tif	ArcGIS	***gisdata***	Raster image containing standard error, in feet, of estimated monthly mean groundwater elevations for each month.
runname_dblsd_mm.tif	ArcGIS	***gisdata***	Raster image containing estimated depth of groundwater below land surface, in feet, for each month.
Wells.shp	ArcGIS	***gisdata***	Point shapefile of runname_wells.csv.
specified_location.shp	ArcGIS	***gisdata***	Point shapefile for selected location containing attributes for the cell.
runname_analysis.html	Dashboard	***gisdata***	Static report containing a summary of output from the interactive tool.

[1]"runname" is the user specified name for the model run, and "mm" is a number representing the month.

with the *runname* prefix in a subdirectory of the "OUTPUT" folder.

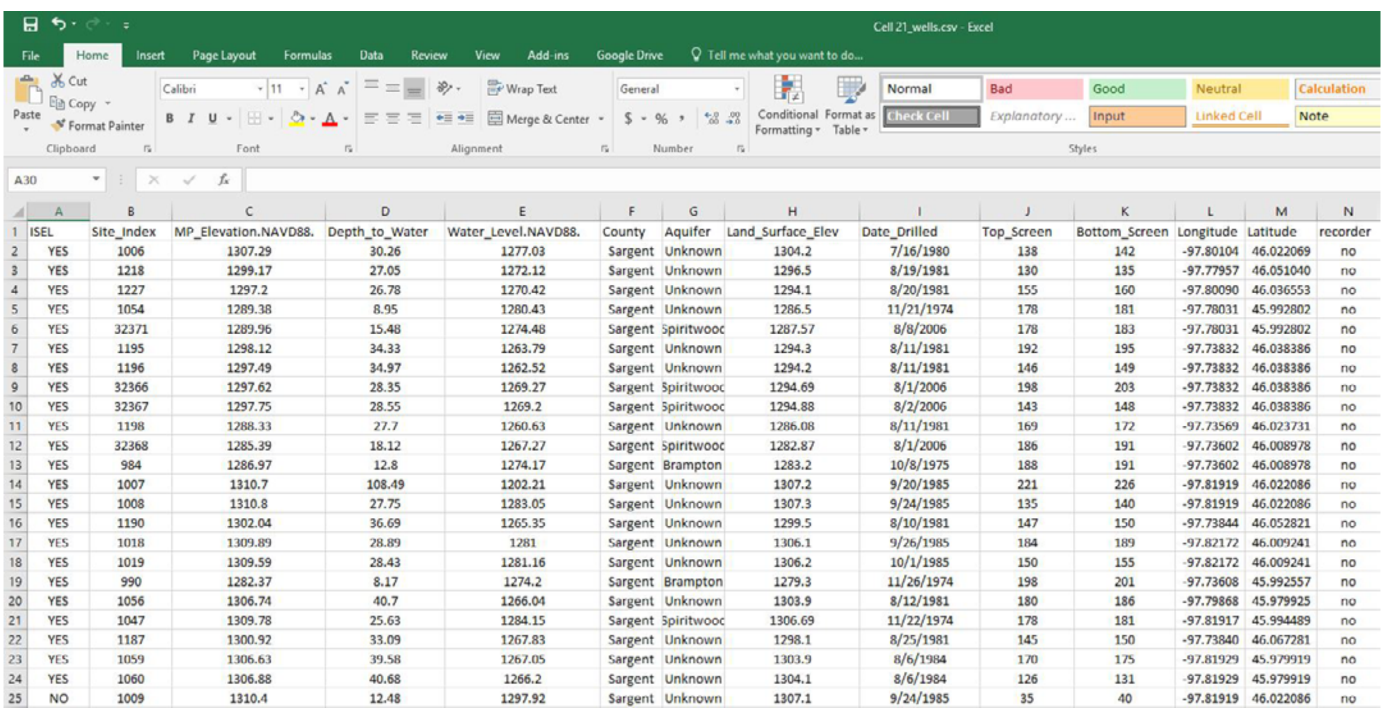
Output for the center point of Cell 21, the user-specified location of *latsp*=46.025, *lonsp*=−97.775, a location with ample data, will be used to illustrate output from the tool (fig. 4). Grid cell numbers were given to the 413 grid cells in the model domain and are included in the *test_log* data frame (appendix). In this example, the cell number was used for the *runname* argument, but any name can be specified. Specifically, in BLOCK 4 of the model_run.R script, dummy variables rn, lat, and lon in lines 89, 92, and 93, respectively, were used to populate arguments for the ***wlmodelCell*** function of *runname, latsp,* and *lonsp* as follows: rn = "Cell 21"; lat=46.025; lon=−97.775 (appendix). To produce output for this example, BLOCK 4 and BLOCK 5 in the model_run.R script were executed.

Diagnostic files include (1) runname_wells.csv, a file containing a table of the wells in the user-specified grid cell available for developing the model (fig. 4*A*); (2) runname_near.pdf, a file containing a time-series plot of monthly mean groundwater elevations for wells used in developing the model (fig. 4*B*); (3) runname_fit.pdf, a file containing a plot of model fit, maps of estimated monthly mean groundwater surface and standard errors for the grid cell containing the user-specified location for each month (fig. 4*C*); and (4) runname_img.pdf, a detailed view of the map of the estimated monthly mean groundwater surface and standard errors (fig. 4*D*). These files are described in detail below. In figures 4*C* and 4*D*, April is used for an example of the plots, but plots for each month are produced in the files.

The runname_wells.csv file contains a list of all wells within the user-specified grid cell. Wells used in the model are listed first and are ordered from closest to farthest from the user-specified location (*latsp, lonsp*). Wells not used in the model are listed last and are indicated by "NO" in the column heading titled "ISEL" (fig. 4*A*, table 4).

The runname_near.pdf file (fig. 4*B*) includes two plots per well for wells used in developing the model (each well for which ISEL = "yes") and a hypothetical "well" at the user-specified location (*latsp, lonsp*). For figure 4*B*, the first four wells used in the analysis are shown. User-specified arguments *runname*, *latsp,* and *lonsp* are identified in the title of the plots, and the well order is the same as in the runname_wells.csv file. For each well, the first or left-hand plot shows the estimated monthly means for the end of each month. The value of the highest and lowest monthly mean groundwater elevation is explicitly labeled. The 98th and 2d percentile of the estimated groundwater elevations also are identified, and most of the measurements should lie within these upper and lower prediction limits. For the hypothetical well, only estimated groundwater elevations are plotted. Note that groundwater elevations removed through the outlier screening are not shown in these plots. The second (right-hand) plot contains the estimated monthly mean groundwater elevations and prediction limits

A



	ISEL	Site_Index	MP_Elevation.NAVD88.	Depth_to_Water	Water_Level.NAVD88.	County	Aquifer	Land_Surface_Elev	Date_Drilled	Top_Screen	Bottom_Screen	Longitude	Latitude	recorder
2	YES	1006	1307.29	30.26	1277.03	Sargent	Unknown	1304.2	7/16/1980	138	142	-97.80104	46.022069	no
3	YES	1218	1299.17	27.05	1272.12	Sargent	Unknown	1296.5	8/19/1981	130	135	-97.77957	46.051040	no
4	YES	1227	1297.2	26.78	1270.42	Sargent	Unknown	1294.1	8/20/1981	155	160	-97.80090	46.036553	no
5	YES	1054	1289.38	8.95	1280.43	Sargent	Unknown	1286.5	11/21/1974	178	181	-97.78031	45.992802	no
6	YES	32371	1289.96	15.48	1274.48	Sargent	Spiritwooc	1287.57	8/8/2006	178	183	-97.78031	45.992802	no
7	YES	1195	1298.12	34.33	1263.79	Sargent	Unknown	1294.3	8/11/1981	192	195	-97.73832	46.038386	no
8	YES	1196	1297.49	34.97	1262.52	Sargent	Unknown	1294.2	8/11/1981	146	149	-97.73832	46.038386	no
9	YES	32366	1297.62	28.35	1269.27	Sargent	Spiritwooc	1294.69	8/1/2006	198	203	-97.73832	46.038386	no
10	YES	32367	1297.75	28.55	1269.2	Sargent	Spiritwooc	1294.88	8/2/2006	143	148	-97.73832	46.038386	no
11	YES	1198	1288.33	27.7	1260.63	Sargent	Unknown	1286.08	8/11/1981	169	172	-97.73569	46.023731	no
12	YES	32368	1285.39	18.12	1267.27	Sargent	Spiritwooc	1282.87	8/1/2006	186	191	-97.73602	46.008978	no
13	YES	984	1286.97	12.8	1274.17	Sargent	Brampton	1283.2	10/8/1975	188	191	-97.73602	46.008978	no
14	YES	1007	1310.7	108.49	1202.21	Sargent	Unknown	1307.2	9/20/1985	221	226	-97.81919	46.022086	no
15	YES	1008	1310.8	27.75	1283.05	Sargent	Unknown	1307.3	9/24/1985	135	140	-97.81919	46.022086	no
16	YES	1190	1302.04	36.69	1265.35	Sargent	Unknown	1299.5	8/10/1981	147	150	-97.73844	46.052821	no
17	YES	1018	1309.89	28.89	1281	Sargent	Unknown	1306.1	9/26/1985	184	189	-97.82172	46.009241	no
18	YES	1019	1309.59	28.43	1281.16	Sargent	Unknown	1306.2	10/1/1985	150	155	-97.82172	46.009241	no
19	YES	990	1282.37	8.17	1274.2	Sargent	Brampton	1279.3	11/26/1974	198	201	-97.73608	45.992557	no
20	YES	1056	1306.74	40.7	1266.04	Sargent	Unknown	1303.9	8/12/1981	180	186	-97.79868	45.979925	no
21	YES	1047	1309.78	25.63	1284.15	Sargent	Spiritwooc	1306.69	11/22/1974	178	181	-97.81917	45.994489	no
22	YES	1187	1300.92	33.09	1267.83	Sargent	Unknown	1298.1	8/25/1981	145	150	-97.73840	46.067281	no
23	YES	1059	1306.63	39.58	1267.05	Sargent	Unknown	1303.9	8/6/1984	170	175	-97.81929	45.979919	no
24	YES	1060	1306.88	40.68	1266.2	Sargent	Unknown	1304.1	8/6/1984	126	131	-97.81929	45.979919	no
25	NO	1009	1310.4	12.48	1297.92	Sargent	Unknown	1307.1	9/24/1985	35	40	-97.81919	46.022086	no

Figure 4. Screenshots from the NDakGWtool of diagnostic output for a user-specified location with ample data. Arguments specified in ***wlmodelcell*** function: *latsp*=46.025, *lonsp*=−97.775, *runname* = Cell 21, and *spaqname0* = "default." *A*, runname_wells.csv, a file containing a table of the wells in the user-specified grid cell available for developing the model.

B

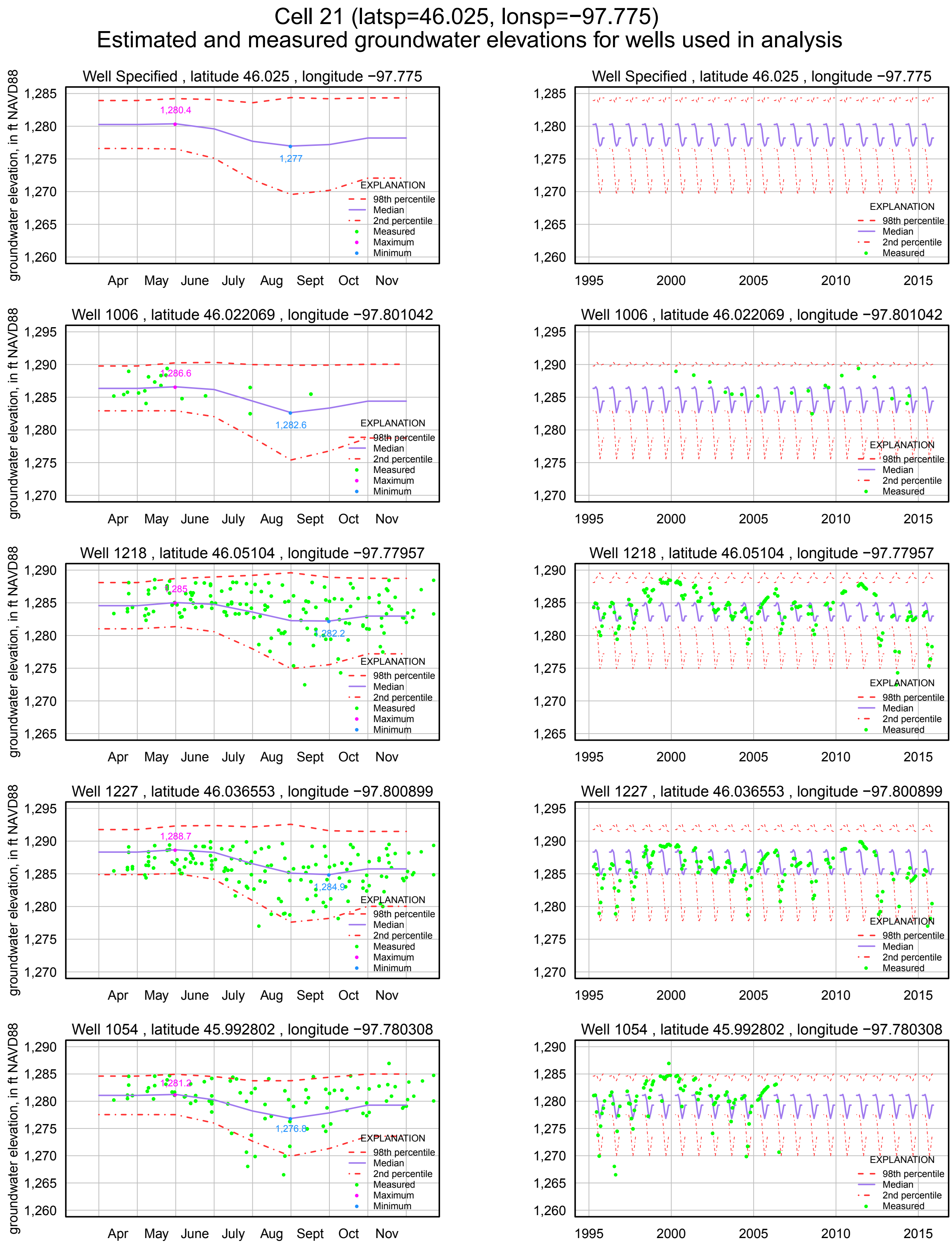


Figure 4. Screenshots from the NDakGWtool of diagnostic output for a user-specified location with ample data. Arguments specified in ***wlmodelcell*** function: *latsp*=46.025, *lonsp*=–97.775, *runname* = Cell 21, and *spaqname0* = “default.—Continued. *B*, runname_near.pdf, a file containing a time-series plot of monthly mean groundwater elevations for wells used in developing the model (a subset of wells is presented here).

from the left-hand plot, repeated for each year used for the water-level model, along with the same measured groundwater elevations used to fit the model to produce the monthly averages. These plots can be examined to identify any potential long-term trends or problems from nonoverlapping or missing periods of record for different wells.

The runname_fit.pdf file provides an overview of the model with a set of three plots for each month: a plot of model fit; a map of the estimated monthly mean groundwater surface; and a map of the standard errors of the estimated groundwater surface (fig. 4*C*). In the left-hand plot, estimated versus measured groundwater elevations, along with the line of equality plus or minus one standard error, are plotted to show model fit. From this plot, irregularities such as clustering of points above or below the line of equality; highly nonsymmetrical distribution of points, for example, many more points below the line than above; or other irregularities can be identified. If the user identifies irregularities, the user can refine the process to correct these irregularities. This is illustrated below in the “Use of the Interactive Tool” section. Additional information about the model is provided in the text above the three plots and includes the number of outliers removed through the outlier screening, the number of measurements left after outlier screening, the number of wells used in the model, and the span and degree of the LWR model. The middle plot is a map of the estimated monthly mean groundwater surface elevation for the 0.1- by 0.1-degree cell and a black box highlights the center 0.05- by 0.05-degree part of the cell. The X identifies the user-specified location for the model run (*latsp*, *lonsp*), and circles represent the location of the wells included in the analysis. This map allows a cursory examination of the shape of the groundwater surface and the number and location of wells that had groundwater elevations for a given monthly window. Note that the number and locations of the wells shown in these plots can differ among months, especially if wells have spotty seasonal coverage. Also note that the range in groundwater surface elevation is indicated above each plot and differs from month to month. The right-hand plot is a map of the standard errors for estimated groundwater elevations provided in the groundwater surface map.

A more detailed view of the estimated monthly mean groundwater surface map and standard errors are shown in the runname_img.pdf file (fig. 4*D*). In this file, all wells used in the model, regardless of month, are shown and the well index numbers are given. Also, groundwater elevation ranges and colors are the same for every month, so groundwater surface elevations in different months can be compared.

ArcGIS files include raster files of estimated monthly mean groundwater surface elevations, standard error (in feet) of estimated groundwater surface, and estimated depth of groundwater below land surface (in feet) for each month for the grid cell of the user-specified location (fig. 5). Note, only output for estimated groundwater surface elevation for the month of April is shown on figure 5. There are 21 .tif files produced for each model run; a .tif file is produced for each of the 7 months, April–October, for 3 parameters: (1) estimated groundwater surface elevations, (2) standard error of estimated groundwater surface, and (3) estimated groundwater depth below land surface. In addition, two shapefiles (.shp) that

C

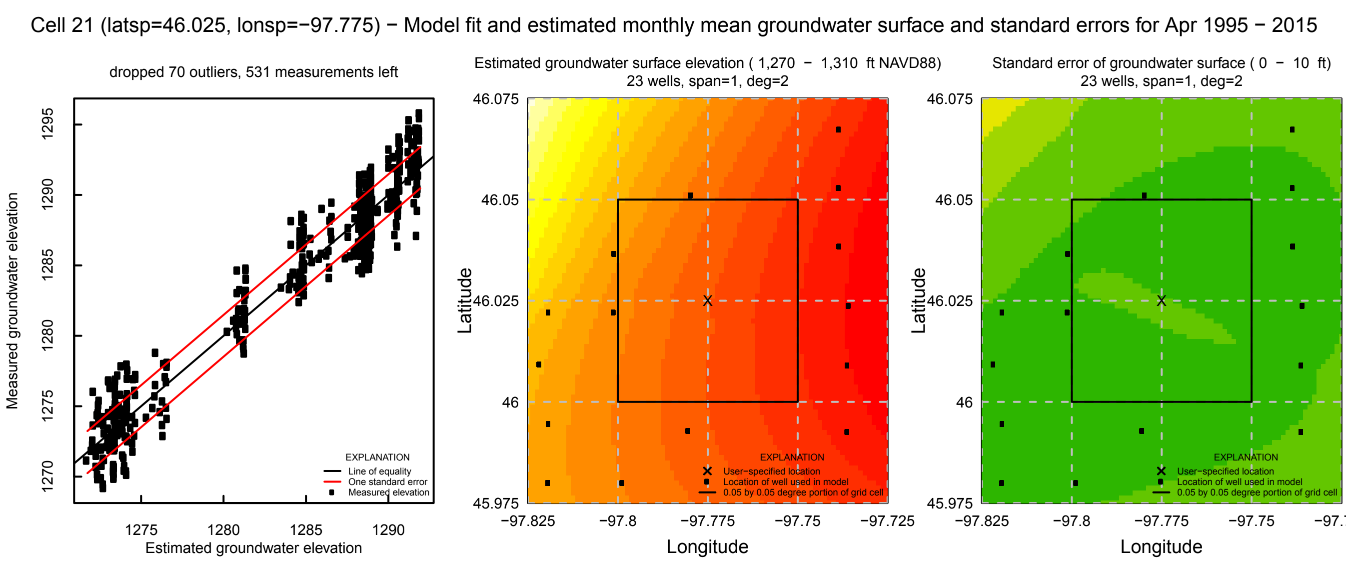


Figure 4. Screenshots from the NDakGWtool of diagnostic output for a user-specified location with ample data.—Continued. Arguments specified in ***wlmodelcell*** function: *latsp*=46.025, *lonsp*=–97.775, *runname* = Cell 21, and *spaqname0* = “default. *C*, runname_fit.pdf, a file containing a plot of model fit, maps of estimated monthly mean groundwater surface and standard errors for the grid cell containing the user-specified location for each month (only April is shown here).

D

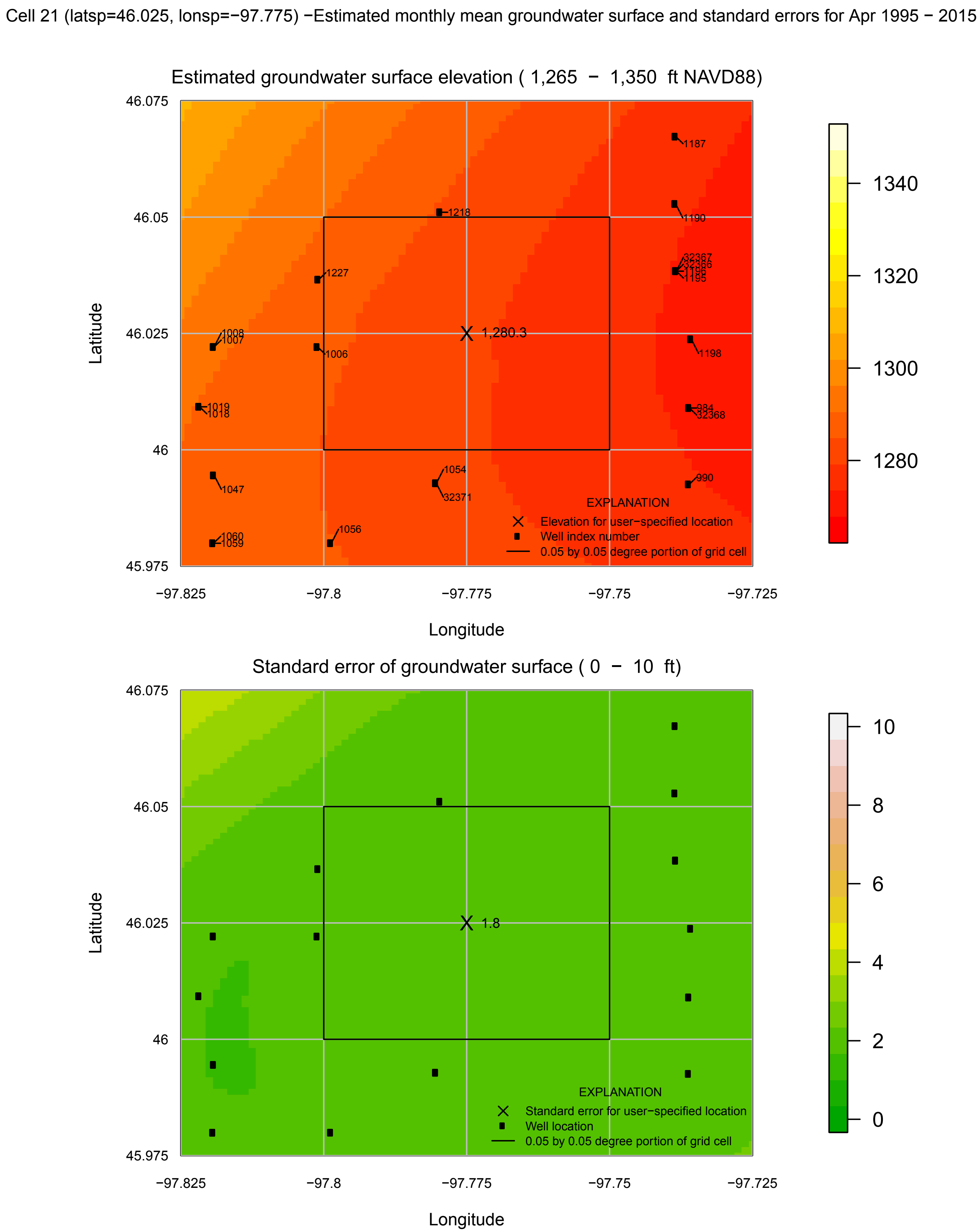


Figure 4. Screenshots from the NDakGWtool of diagnostic output for a user-specified location with ample data.—Continued. Arguments specified in ***wlmodelcell*** function: *latsp*=46.025, *lonsp*=–97.775, *runname* = Cell 21, and *spaqname0* = "default. *D*, runname_img.pdf, a file containing a more detailed view of the map of estimated monthly mean groundwater surface and standard errors (only April is shown here).

Table 4. Information provided in diagnostic file, runname_wells.csv.

Column name	Description
ISEL	Identifies if the well is used in the development of the water level model.
Site_Index	Well name.
MP_Elevation.NAVD88.[1]	Measuring point elevation for the well.
Depth_to_Water	Depth to water calculated for the first observation for the well by subtracting Water_Level.NAVD88. from Land_Surface_Elev.
Water_Level.NAVD88.	Groundwater elevation from the first observation for the well.
County	County where the well is.
Aquifer	Aquifer where the well is.
Land_Surface_Elev	Land surface elevation where the well is.
Date_Drilled	Date the well was drilled.
Top_Screen	Top-screen interval depth.
Bottom_Screen	Bottom-screen interval depth.
Longitude	Longitude where the well is.
Latitude	Latitude where the well is.
Recorder	Identifies if the well is a continuous recorder well.

[1]All elevations are in feet and use the North American Vertical Datum of 1988.

identify the user-specified location of the model run and the wells used in the analysis are produced.

The summary file runname_analysis.html, called the dashboard, is a static report that contains a summary of output from the NDakGWtool for the user-specified location. In the ***gisdata*** function, the analyses.Rmd script is called to write the dashboard (table 1, fig. 6, appendix). The analyses.Rmd script uses the flexdashboard R package (table 1) and, for a given set of user-specified arguments, a static dashboard containing nine tabs is generated. The first tab is a map showing different layers: model domain, glacial drift aquifers, all wells in the NDSWC database, and wells used in the analysis (fig. 6*A*). The second tab is a summary table of the estimated monthly mean groundwater elevations, standard error of estimated groundwater elevations, and estimated depth of groundwater below land-surface datum for the user-specified location (fig. 6*B*). The remaining tabs, one for each of the 7 months, contain three maps of the grid cell containing the user-specified location: estimated monthly mean groundwater surface and standard error, and depth of groundwater below land-surface datum (fig. 6*C*). Note, only the month of April is shown on figure 6*C*. Similar to the maps produced in the

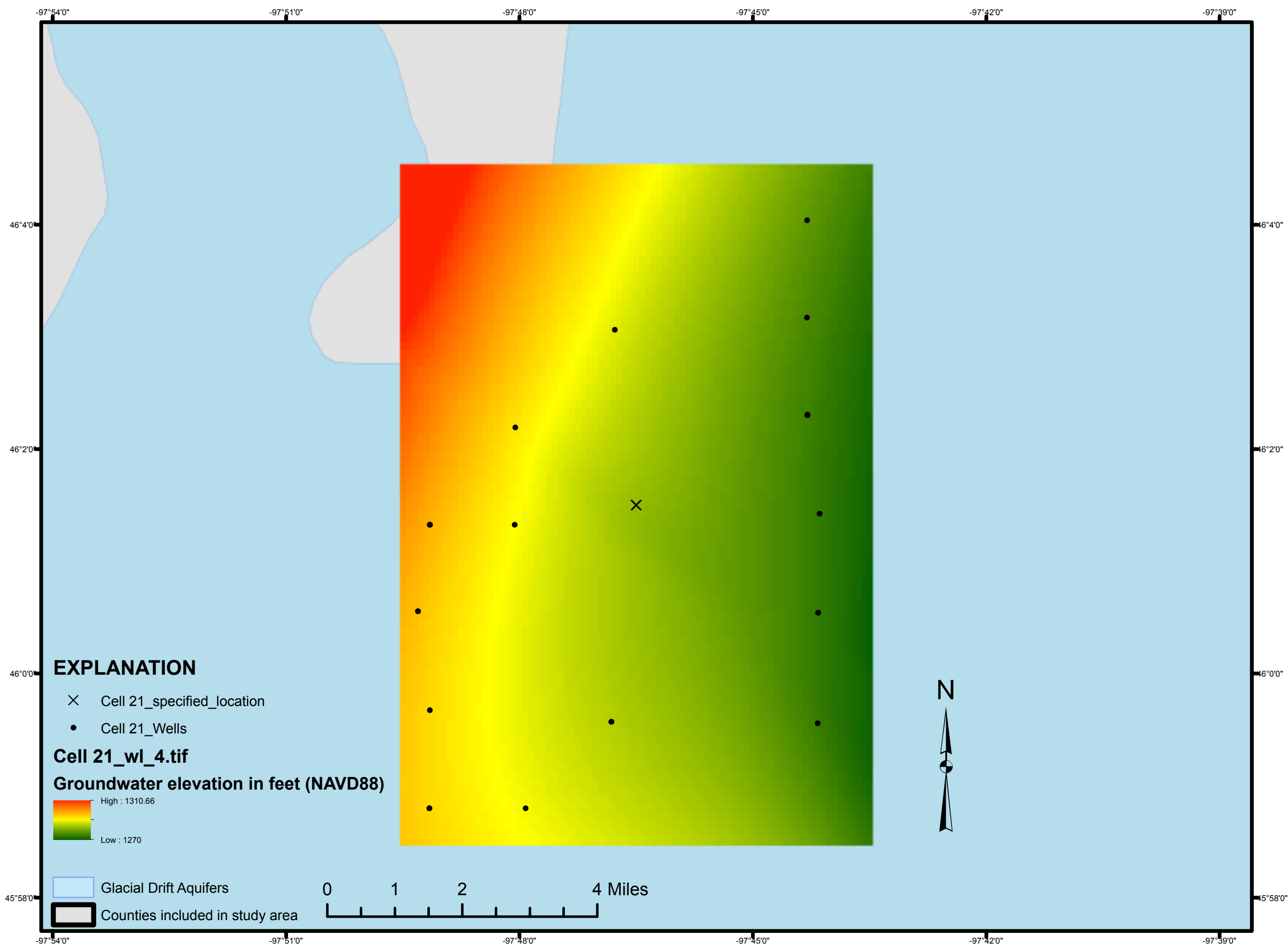


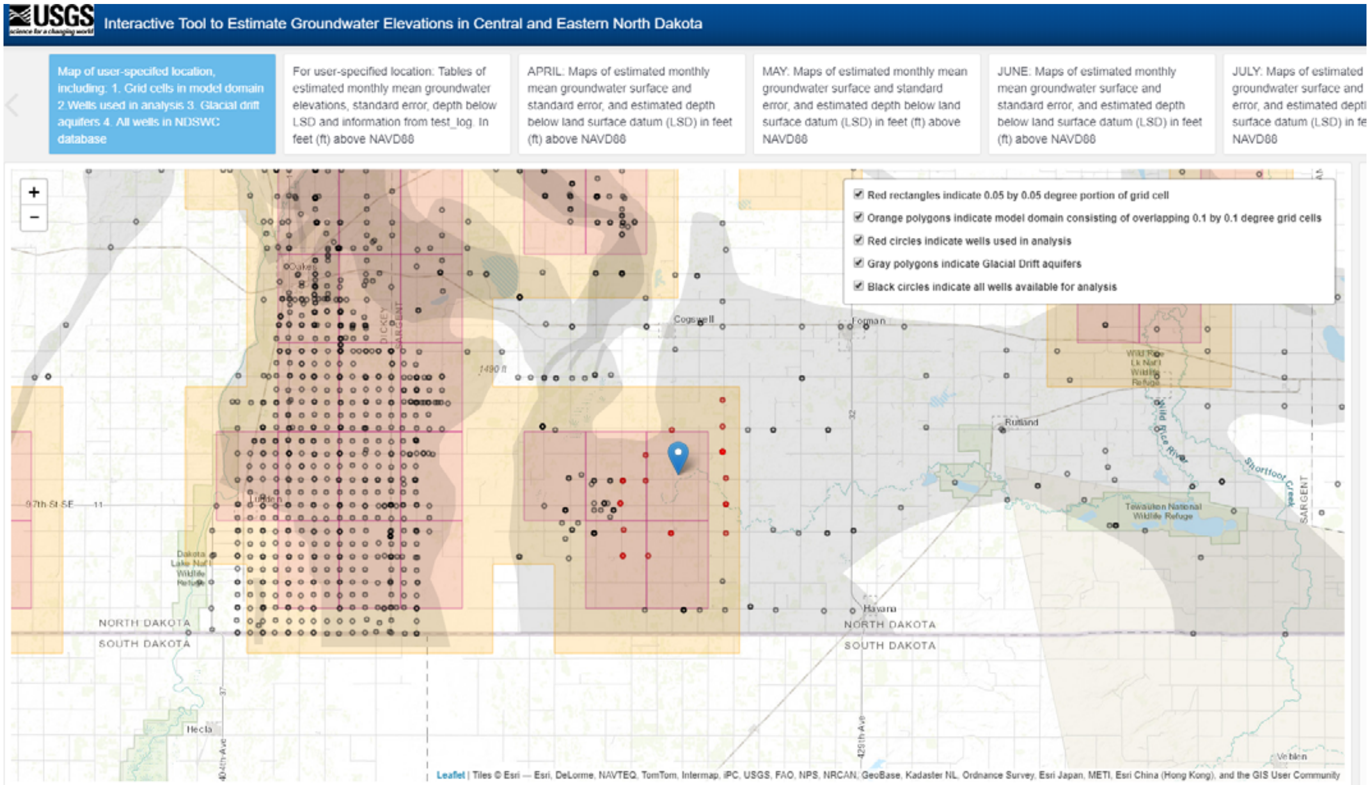
Figure 5. Example output from the NDakGWtool of ArcGIS files for Cell 21: Cell 21_specified_location, shape file layer that identifies the user-specified location; Cell 21_Wells, shape file layer that identifies wells used in the analysis, and Cell 21_wl_4.tif is the raster image of estimated groundwater surface for April.

diagnostic files, output is for the 0.1- by 0.1-degree cell and the X indicates the user-specified location for the model run (*latsp*, *lonsp*) and circles representing the locations of wells included in the analysis.

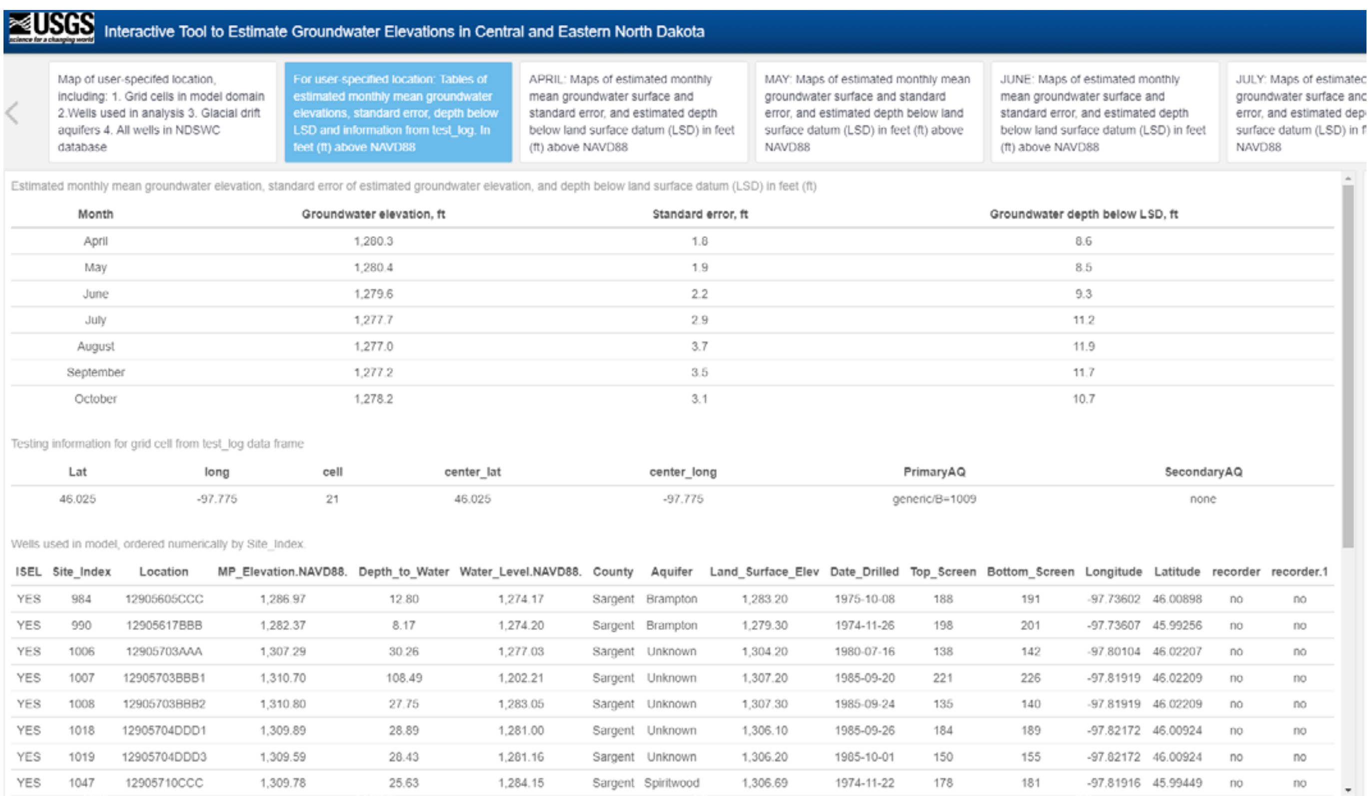
Model Testing

Model testing was performed by evaluating diagnostic output of the model for the center point of each grid cell, and through model testing, the model domain was established. During model testing, the ***wlmodelCell*** function could be executed for any location in the study area, but once the model domain was established, results were only available for the model domain. The ***wlmodelCell*** function was executed for the center point of each of the 5,340 grid cells for the 32-county study area using groundwater elevations from 1995 to 2015. The more recent (1995–2015) period represents a period of wetter conditions and generally higher groundwater elevations in central and eastern North Dakota compared to earlier years. The coverage of the tool was narrowed down through 2 main steps to include 413 grid cells in the model domain. First, most grid cells were eliminated because they did not meet the minimum requirements of sufficient data as determined in the ***wlmodelcell*** function (data requirements are described in the “Statistical Model for Estimating Groundwater Elevations” section). Many of these grid cells had no data because they were not overlying areas of glacial drift aquifers. Grid cells that were not in designated areas of glacial drift aquifers were initially included to address the need of the NRCS to include as many shallow groundwater areas as possible. Other grid cells in designated areas of glacial drift aquifers that may have had sufficient observation wells had measurement frequencies that were too sparse, a period of record that was too short, or well coverage that was too sparse and inconsistent to produce

A



B



Estimated monthly mean groundwater elevation, standard error of estimated groundwater elevation, and depth below land surface datum (LSD) in feet (ft)

Month	Groundwater elevation, ft	Standard error, ft	Groundwater depth below LSD, ft
April	1,280.3	1.8	8.6
May	1,280.4	1.9	8.5
June	1,279.6	2.2	9.3
July	1,277.7	2.9	11.2
August	1,277.0	3.7	11.9
September	1,277.2	3.5	11.7
October	1,278.2	3.1	10.7

Testing information for grid cell from test_log data frame

Lat	long	cell	center_lat	center_long	PrimaryAQ	SecondaryAQ
46.025	-97.775	21	46.025	-97.775	generic/B=1009	none

Wells used in model, ordered numerically by Site_Index.

ISEL	Site_Index	Location	MP_Elevation.NAVD88.	Depth_to_Water	Water_Level.NAVD88.	County	Aquifer	Land_Surface_Elev	Date_Drilled	Top_Screen	Bottom_Screen	Longitude	Latitude	recorder	recorder.1
YES	984	12905605CCC	1,286.97	12.80	1,274.17	Sargent	Brampton	1,283.20	1975-10-08	188	191	-97.73602	46.00898	no	no
YES	990	12905617BBB	1,282.37	8.17	1,274.20	Sargent	Brampton	1,279.30	1974-11-26	198	201	-97.73607	45.99256	no	no
YES	1006	12905703AAA	1,307.29	30.26	1,277.03	Sargent	Unknown	1,304.20	1980-07-16	138	142	-97.80104	46.02207	no	no
YES	1007	12905703BBB1	1,310.70	108.49	1,202.21	Sargent	Unknown	1,307.20	1985-09-20	221	226	-97.81919	46.02209	no	no
YES	1008	12905703BBB2	1,310.80	27.75	1,283.05	Sargent	Unknown	1,307.30	1985-09-24	135	140	-97.81919	46.02209	no	no
YES	1018	12905704DDD1	1,309.89	28.89	1,281.00	Sargent	Unknown	1,306.10	1985-09-26	184	189	-97.82172	46.00924	no	no
YES	1019	12905704DDD3	1,309.59	28.43	1,281.16	Sargent	Unknown	1,306.20	1985-10-01	150	155	-97.82172	46.00924	no	no
YES	1047	12905710CCC	1,309.78	25.63	1,284.15	Sargent	Spiritwood	1,306.69	1974-11-22	178	181	-97.81916	45.99449	no	no

Figure 6. Screenshots from the NDakGWtool of the static summary file runname_analysis.html, called the dashboard, for Cell 21. *A*, The first tab of the dashboard identifies user-specified location and wells used in analysis. *B*, The second tab provides a summary table for each month.

C

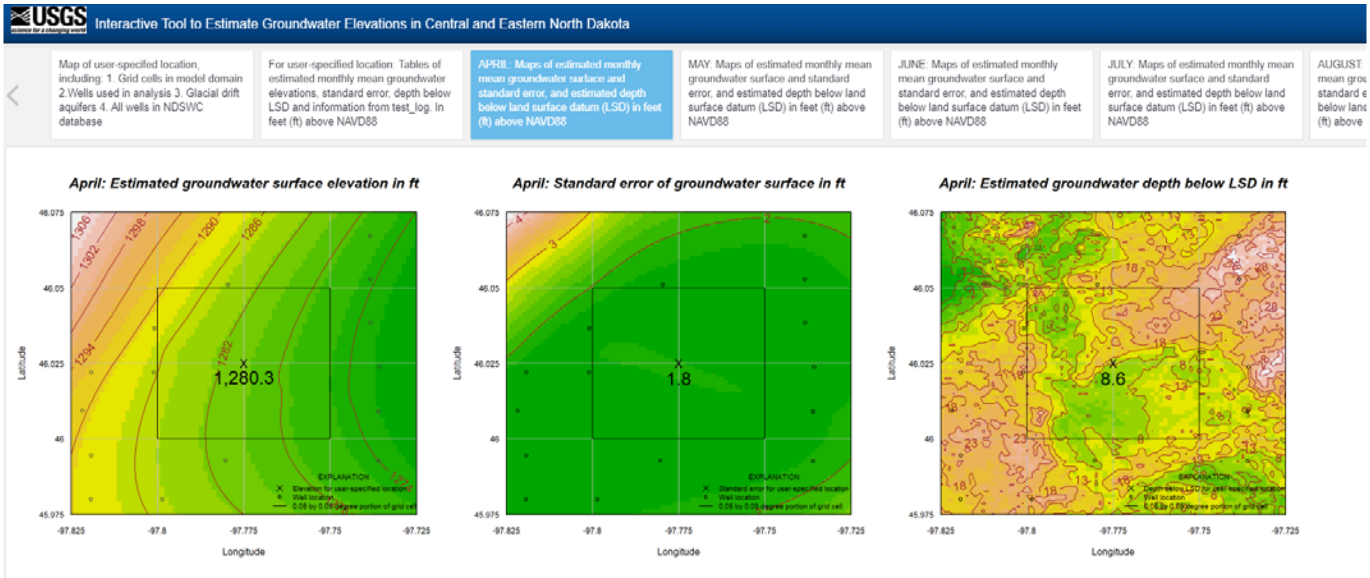


Figure 6. Screenshots from the NDakGWtool of the static summary file runname_analysis.html, called the dashboard, for Cell 21.—Continued. *C*, The third tab shows summary plots of estimated monthly mean groundwater surface and standard error and depth of groundwater below land surface, in feet, for April. Subsequent tabs are the same as the third tab but specific to each month.

reliable models. As a result, only 660 grid cells of the initial 5,340 cells contained sufficient data to produce reliable model results. Of those 660 cells, 413 were identified as having reliable output and were used to define the model domain (fig. 3). The model was deemed reliable through manual examination of diagnostic output to verify groundwater elevations generally were within the prediction limits, standard errors were relatively low (generally less than 10 ft), and the estimated maps seemed to represent a spatially homogeneous groundwater system. Results for the 413 cells in the model domain were manually examined, and modified as needed, to optimize accuracy and reliability based on diagnostic output to determine an appropriate default for the *spaqname0* argument. In some cases, one or more alternate choices were appropriate for the *spaqname0* argument depending on the location or application being considered. The default and alternate values, if defined, for *spaqname0* are provided in the *test_log* data frame (table 1, appendix). In the *test_log* data frame, the default value is referred to as PrimaryAQ, and the alternate value is referred to as SecondaryAQ. If the *spaqname0* argument is not specified by the user, the default value from *test_log* is used. The default results can be examined, and the *spaqname0* argument can be modified for the user's purpose. In some cases, the estimated groundwater surface may be highly uncertain, and there may be no apparent way to modify the *spaqname0* argument to obtain more accurate results with the available data. Although the model domain was determined using the period 1995–2015, for locations with sufficient data availability, it could be used to estimate groundwater elevations for the earlier (1960–94) period, or for any user-specified period during 1960–2017. It is important to note that any variations in the arguments beyond the testing conditions may result in a model that is not reliable; thus, the user should closely assess the model output when specifying different values for arguments.

Use of the Interactive Tool

Depending upon the groundwater elevation data available for developing the model, uncertainty of the estimated groundwater elevations from the NDakGWtool can vary for the user-specified location. In this section, an example of results for a location with ample data and a location with sparse data available for developing a model are illustrated. Also, two examples are illustrated on how the user can use the *spaqname0* argument to refine results depending on the user's purpose.

An example of a location with ample data available for developing the model is Cell 21 (fig. 4). From the *test_log* data frame, the default argument is defined as *spaqname0* = "generic/*B*=1009" (appendix). The modifier "generic" indicates wells with any aquifer name are being used and well 1009 is being removed using the B (spbadwells; table 2) modifier within the *spaqname0* argument. Well 1009 was removed from the analysis for this cell because all groundwater elevations were outside the prediction limits.

Investigation of the diagnostic files provides the user with a wealth of information about the data available to fit the model, the resulting model itself, and results from the model; for example, from runname_wells.csv (fig. 4*A*, table 4), wells used in the model for Cell 21, column ISEL=YES, are in a mix of aquifers identified as Unknown, Spiritwood, or Brampton aquifers. Aquifers designated as "Unknown," "Undefined," and "Unnamed" are present throughout the dataset and indicate the well is in an aquifer that has not been identified for various reasons (Michael Ginsbach, NDSWC, oral commun., 2017). Based on the "Latitude" and "Longitude" columns, wells 1195, 1196, 32366, and 32367 can be identified as being colocated. The "Date_Drilled" column identifies wells 1195 and 1196 were drilled in 1981, but the aquifer name was not identified, and wells 32366 and 32367 were drilled in 2006 and identified to be in the Spiritwood aquifer. From inspection of the runname_near.pdf file (not shown), groundwater elevations from wells 1195 and 1196 are no longer available after 2006, but wells 32366 and 32367 have groundwater elevations from 2006 to 2015. From these two diagnostic files, because the wells are of the same depth, it can be determined that wells 32366 and 32367 are a continuation of wells 1195 and 1196, respectively. Because wells fail over time, wells are discontinued and new wells are drilled in about the same location and depth (Michael Ginsbach, NDSWC, oral commun., 2017). Valuable information for developing the model is included by using wells in the "Unknown" aquifer in the analysis.

Details about the model developed for Cell 21 are provided in the runname_fit.pdf file (fig. 4*C*). As mentioned in the "Statistical Model for Estimating Groundwater Elevations" section, the LWR model can be different for each month, depending on the availability of groundwater elevations. For the April model of Cell 21, 70 groundwater elevations were identified as outliers, but 531 groundwater elevations for 23 wells remained to develop the model (fig. 4*C*). The LWR model has a *degree*=2, *span*=1, indicating the density and distribution of wells are fairly evenly distributed and a relatively smooth groundwater surface is estimated. The runname_img.pdf file provides a more detailed view of the results and shows that, although the user-specified location is in an area where the closest wells are roughly 5 mi (0.05 degrees) away, the estimated groundwater elevation for the user-specified location has a small standard error, less than 2 ft (fig. 4*D*). Also, the colocated wells identified in the runname_wells.csv file are easily detected in the runname_img.pdf file. This is an example of a grid cell for which using the default argument produces reliable results and would probably not require further modification.

There also are many cells in the model domain that have few wells and few groundwater elevations for developing the LWR model; for example, Cell 33 has only 5 wells and 103 groundwater elevations available to develop a model (fig. 7*A*). The resulting model is the simplest LWR form of *degree*=1 and *span*=1. Output is available for this grid cell, but standard errors are larger than what was seen in Cell 21 (fig. 7*B*).

The default argument for *spaqname0* specified in the *test_log* data frame can be used, but the user may have detailed knowledge of aquifer behavior and may want to modify the optional arguments to better match the conceptual understanding of the groundwater surface. A couple of examples are illustrated below. The first example is a situation in which two aquifers overlay one another within the same grid cell. In Cell 1, the Guelph aquifer overlays the Ellendale aquifer (fig. 8). By using the default argument from the *test_log*, *spaqname0* = "default," wells in the Ellendale aquifer

A

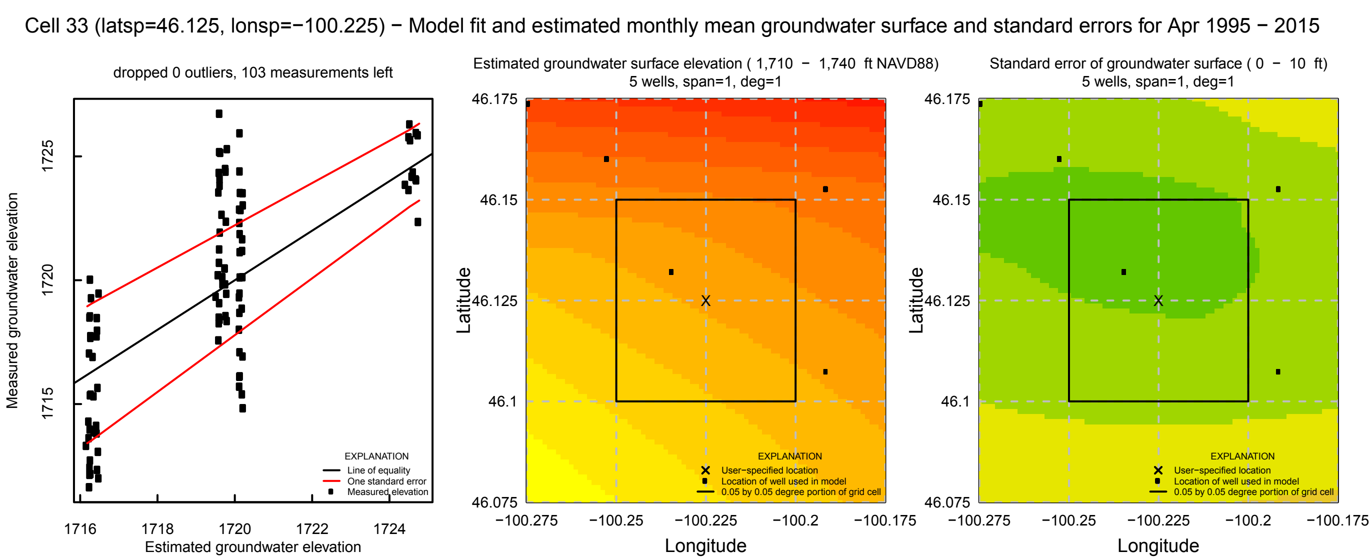


Figure 7. Screenshot from the NDakGWtool of diagnostic output for a user-specified location with sparse data. Arguments specified in ***wlmodelcell*** function include *latsp*=46.125, *lonsp*=–100.225, *runname* = Cell 33, and *spaqname0* = "default." *A*, runname_fit.pdf.

B

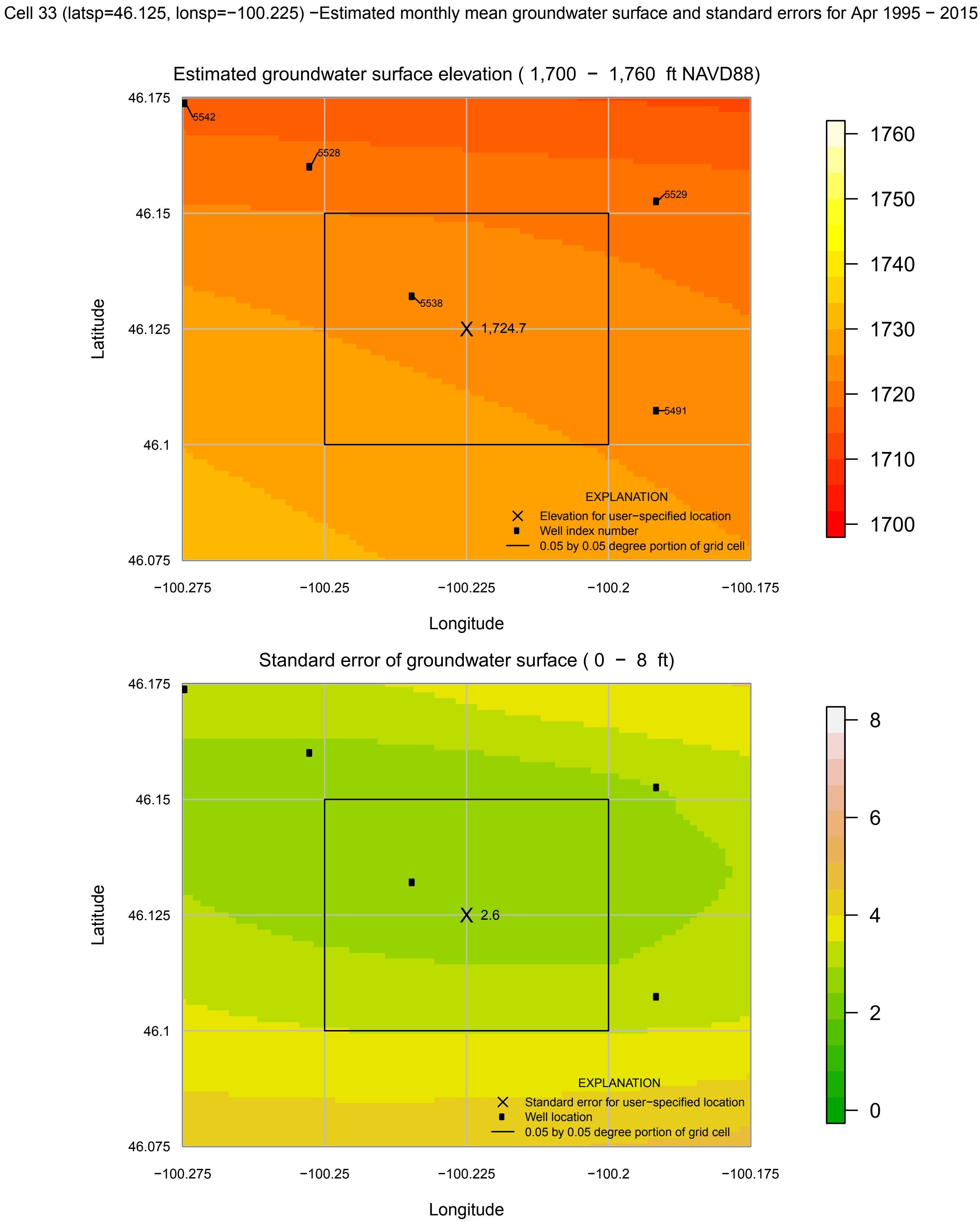


Figure 7. Screenshot from the NDakGWtool of diagnostic output for a user-specified location with sparse data.—Continued. *B*, runname_img.pdf.

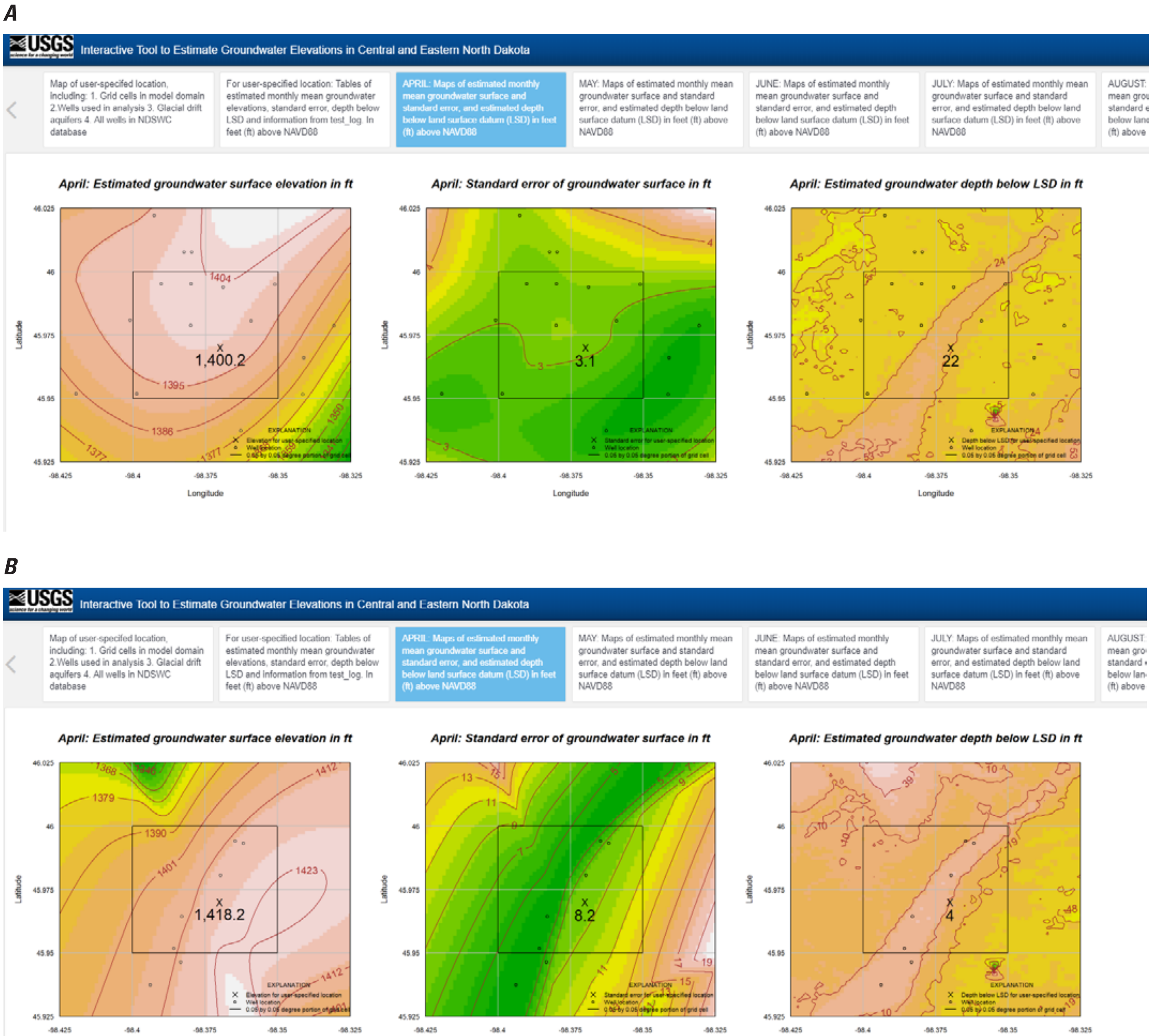
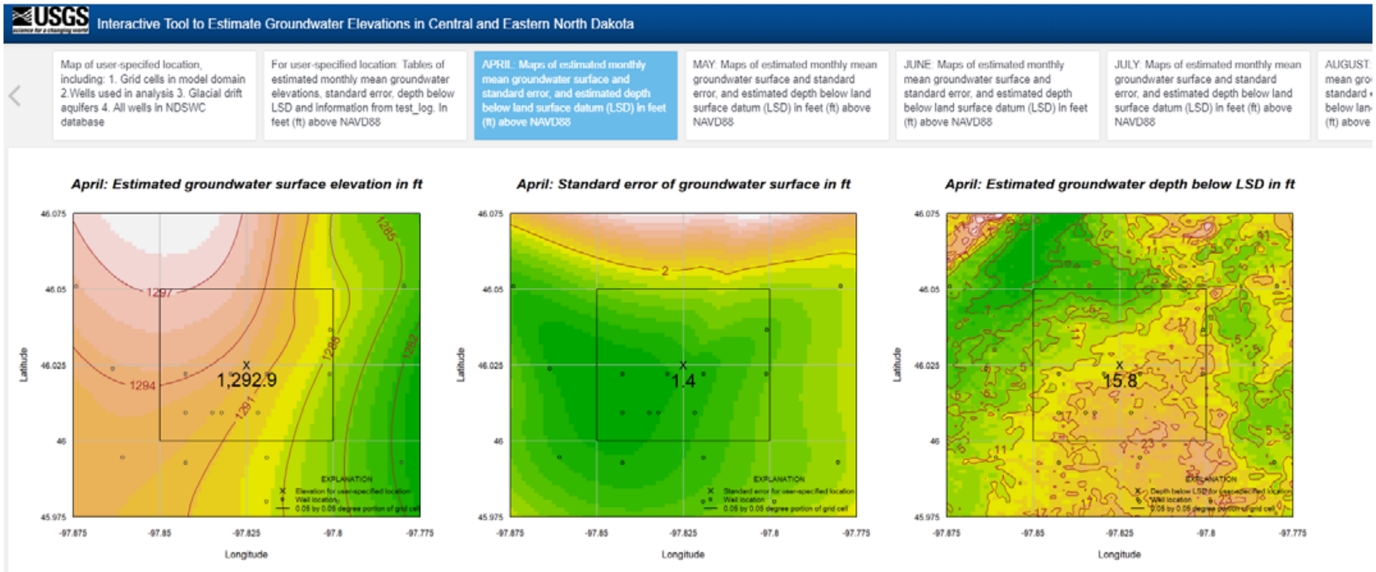


Figure 8. Screenshots from NDakGWtool to illustrate the modification of optional argument *spaqname0*. Arguments specified in the ***wlmodelcell*** function include *latsp*=45.97, *lonsp*=–98.37, *runname* = Cell 1. *A*, *spaqname0* = "default," default argument = "Ellendale." *B*, *spaqname0* = "Guelph."

are selected, and the April monthly mean groundwater elevation is 1,400.2 ft, standard error is 3.1 ft, and the depth below land surface is 22 ft (fig. 8*A*). If the user is interested only in wells in the Guelph aquifer, the aqname modifier can be specified as "Guelph" in the *spaqname0* argument (*spaqname0* = "Guelph"), and the results are different (fig. 8*B*). Using *spaqname0* = "Guelph," the April monthly mean groundwater elevation is higher at 1,418.2 ft, and depth below land surface is 4 ft; however, the standard error of 8.2 ft is larger because the Guelph aquifer has fewer wells with groundwater elevations (indicated by open circles) and a narrow, smaller spatial distribution of those wells in the Guelph aquifer.

The second example for Cell 20 demonstrates how the user can interactively refine output based on the well screen-interval depth (fig. 9). For Cell 20, in *test_log*, the default argument for *spaqname0* for Cell 20 is specified as "generic/*D*>100." Using the default argument, wells with a deeper screen-interval depth are used for the model, and the groundwater surface slopes from northwest to southeast across the cell. The monthly mean groundwater elevation for April

A



B

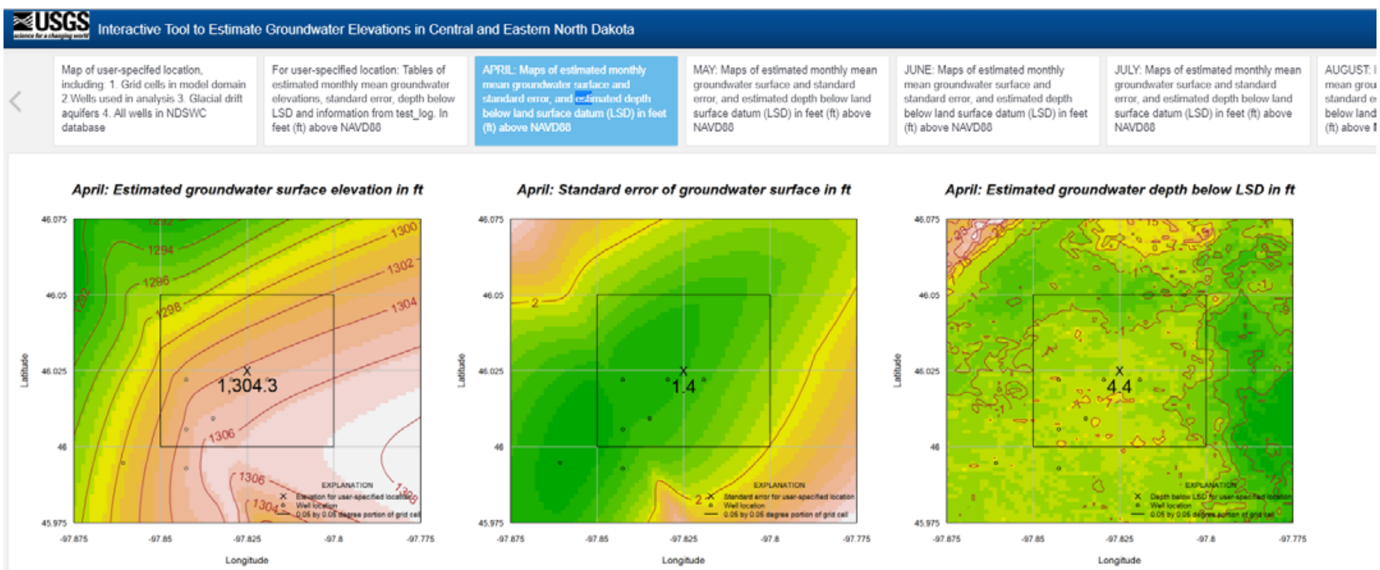


Figure 9. Example of output from the interactive groundwater tool to illustrate modification of an optional argument. Arguments specified in the ***wlmodelcell*** function include *latsp*=46.025, *lonsp*=–97.825, and *runname* = Cell 20. *A*, *spaqname0* = default; default argument = “generic/D>100.” *B*, *spaqname0* = “generic/D<100.”

is 1,292.9 ft (fig. 9*A*). If the user is only interested in wells open to the aquifer at a shallower depth, the modifier within the *spaqname0* argument for screen-interval depth can be specified as *D*<100 (spdepth; table 2). Model results are much different, with the water surface sloping from southeast to northwest across the cell, and the monthly mean groundwater elevation for April is more than 11 ft higher at 1,304.3 ft (fig. 9*B*). For Cell 20, groundwater elevation results are different when wells are selected based on screen-interval depth.

References Cited

Cleveland, W.S., Grosse, E., and Shyu, W.M., 1992, Local regression models, chap. 8 *of* Chambers, J.M., and Hastie, T.J., eds., Statistical models in S: Pacific Grove, Calif., Wadsworth and Brooks/Cole Advanced Books and Software, p. 309–376.

Esri, 2018, ArcGIS desktop: Esri software, accessed August 17, 2015, at https://desktop.arcgis.com/en/desktop/.

North Dakota State Water Commission, 2017, Ground/surface water data: North Dakota State Water Commission web page, accessed July 21, 2017, at http://www.swc.nd.gov/info_edu/map_data_resources/groundsurfacewater/.

R Development Core Team, 2018, R—A language and environment for statistical computing: Vienna, Austria, R Foundation for Statistical Computing, accessed March 23, 2018, at https://www.r-project.org.

U.S. Department of Agriculture, 2015, Hydrology tools for wetland identification and analysis, chap. 19 *of* Engineering field handbook: Washington D.C., U.S. Department of Agriculture, Natural Resources Conservation Service 113 p.

U.S. Geological Survey, 2017, National Elevation Dataset: U.S. Geological Survey web page, accessed July 24, 2017, at https://nationalmap.gov/elevation.html.

Appendix. R Documentation

A zipped folder, NDakGWtool.zip, contains the following:

- Rscripts—model_run.R, analyses.Rmd;
- R objects—obs_wl, test_log, lse;
- files—wlinputalldata.R, wlmodelCell.R, model_data.Rdata, gisdata.R, sites.txt, water_levels.txt, USGS_ID_black.png, readme.txt; and
- folders—GIS, OUTPUT.

The folder can be downloaded at https://doi.org/10.3133/ofr20181185.

For more information about this publication, contact:
Director, USGS Dakota Water Science Center
821 East Interstate Avenue
Bismarck, ND 58503
701–250–7400

For additional information, visit: https://www.usgs.gov/centers/dakota-water

Publishing support provided by the
Rolla Publishing Service Center

Nustad and others—**Interactive Tool to Estimate Groundwater Elevations in Central and Eastern North Dakota**—Open-File Report 2018–1185

ISSN 2331-1258 (online)
https://doi.org/10.3133/ofr20181185