

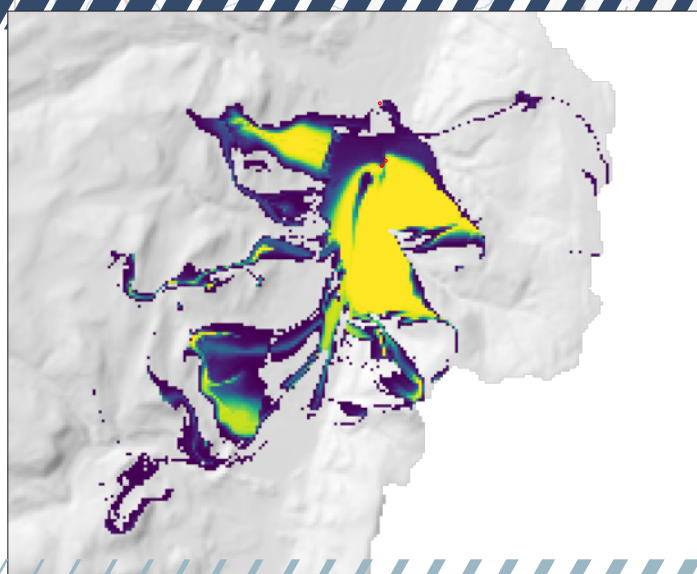
74°40'

74°30'

74°20'

Prepared in cooperation with the New York State Department of Environmental Conservation
and the New York State Department of Health

Areas Contributing Recharge to Priority Wells in Valley-fill Aquifers in the Neversink River and Rondout Creek Drainage Basins, New York



41°50'

Scientific Investigations Report 2021–5112

Areas Contributing Recharge to Priority Wells in Valley-fill Aquifers in the Neversink River and Rondout Creek Drainage Basins, New York

By Nicholas T. Corson-Dosch, Michael N. Fienen, Jason S. Finkelstein, Andrew T. Leaf, Jeremy T. White, Joshua Woda, and John H. Williams

Prepared in cooperation with the New York State Department of Environmental Conservation and the New York State Department of Health

Scientific Investigations Report 2021–5112

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

U.S. Geological Survey, Reston, Virginia: 2022

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/>.

Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Although this information product, for the most part, is in the public domain, it also may contain copyrighted materials as noted in the text. Permission to reproduce copyrighted items must be secured from the copyright owner.

Suggested citation:

Corson-Dosch, N.T., Fienen, M.N., Finkelstein, J.S., Leaf, A.T., White, J.T., Woda, J., and Williams, J.H., 2022, Areas contributing recharge to priority wells in valley-fill aquifers in the Neversink River and Rondout Creek drainage basins, New York: U.S. Geological Survey Scientific Investigations Report 2021–5112, 50 p., <https://doi.org/10.3133/sir20215112>.

Associated data:

Fienen, M.N., and Corson-Dosch, N.T., 2021, Groundwater model archive and workflow for Neversink/Rondout Basin, New York, source water delineation: U.S. Geological Survey data release, at <https://doi.org/10.5066/P9HWSOHP>.

Woda, J.C., Finkelstein, J.S., and Williams, J.H., 2022, Interpolated hydrogeologic framework and digitized datasets for upstate New York study areas: U.S. Geological survey data release, <https://doi.org/10.5066/P96R5K5R>.

ISSN 2328-0328 (online)

Acknowledgments

The authors acknowledge Paul Friesz and John Masterson (USGS) for helpful conversations related to model design and calibration. The authors also acknowledge Robert Breault (USGS) for his help coordinating this work. We also thank Daniel Feinstein (USGS) and Alden Provost (USGS) for help with MODPATH7, and Christian Langevin (USGS) for help with MODFLOW 6.

Contents

Acknowledgments	iii
Abstract	1
Introduction.....	1
Purpose and Scope	2
Hydrogeologic Framework.....	4
Development of Steady-State Numerical Model	8
Model Extent and Discretization	8
Boundary Conditions	8
Recharge	9
Groundwater Flow Across Model Perimeter	9
Stream-Aquifer Interactions.....	9
Water Withdrawals	9
Hydraulic Properties.....	9
Parameter Estimation by Ensemble History Matching	11
Observation Data	21
Stepwise Estimation of Model Parameters	21
Parameterizing the Model	21
Using Uncertainty Analysis to Set Initial Values	25
History Matching with the Iterative Ensemble Smoother (IES).....	25
Base Realization.....	27
Hydraulic Conductivity.....	27
Boundary Conditions.....	30
Water Balance, Water Table, and Streamflow	31
Providing an Ensemble for Monte Carlo Analysis of Areas Contributing Recharge	35
Simulation of Areas Contributing Recharge and Prediction Uncertainty Analysis	35
Probabilistic Areas Contributing Recharge.....	35
Prediction Uncertainty Analysis.....	35
Assumptions and Limitations of Analysis	37
Model Discretization	44
Steady State Assumption	44
Data Gaps.....	44
Particle Tracking	44
Summary.....	44
References Cited.....	45
Appendix 1. Data Sources Used to Construct the Neversink-Rondout MODFLOW 6 Groundwater-flow Model.....	49

Figures

1. Map showing the location of the study area, model domain, and other key features3
2. Figure showing schematic representation of a glacially deposited valley-fill aquifer5

3.	Figure showing extent of valley-fill aquifers and till-covered bedrock uplands with priority well locations	7
4.	Figure showing model perimeter boundary conditions, streamflow routing (SFR) lines, and pumping wells in groundwater-flow model.....	10
5.	Figure showing base realization values and parameterization for horizontal hydraulic conductivity in model layer 1	13
6.	Figure showing base realization values and parameterization for horizontal hydraulic conductivity in model layer 2	14
7.	Figure showing base realization values and parameterization for horizontal hydraulic conductivity in model layer 3	15
8.	Figure showing base realization values and parameterization for horizontal hydraulic conductivity in model layer 4	16
9.	Figure showing base realization values and parameterization for vertical hydraulic conductivity in model layer 1	17
10.	Figure showing base realization values and parameterization for vertical hydraulic conductivity in model layer 2	18
11.	Figure showing base realization values and parameterization for vertical hydraulic conductivity in model layer 3	19
12.	Figure showing base realization values and parameterization for vertical hydraulic conductivity in model layer 4	20
13.	Figure showing spatial distribution of hydraulic head residuals (observed minus simulated) for the posterior parameter ensemble	22
14.	Figure showing spatial distribution of flux residuals (observed minus simulated) for the posterior parameter ensemble	23
15.	Figure showing spatial distribution of land surface inequalities (land-surface elevation minus simulated hydraulic head) for the posterior parameter ensemble	24
16.	Graph showing PHI distribution, <i>A</i> , before and, <i>B</i> , after rebalancing the objective function	26
17.	Figure showing estimated annual recharge to the groundwater system	28
18.	Figure showing histograms for selected observations showing model performance with respect to observed values for the “prior Monte Carlo”, “expanded bounds Monte Carlo”, and final “iES posterior” parameter ensembles	29
19.	Graph showing summary of ensemble and base objective function progress by iterative ensemble smoother iteration	30
20.	Figure showing comparison of posterior parameter ensembles before and after anomalously high PHI values were trimmed during rejection sampling.....	31
21.	Figure showing relations among observed and simulated flux, hydraulic head, and land-surface inequalities, and plots of residuals compared to observed values	32
22.	Figure showing summary of base realization anisotropy in the groundwater-flow model	33
23.	Figure showing base realization aquifer flux in Streamflow Routing (SFR) cells	34
24.	Figure showing base realization groundwater model mass balance	35
25.	Figure showing base realization water table and simulated flooding.	36
26.	Figure showing probabilistic areas contributing recharge to priority wells under the “Baseratio” pumping scenario	38
27.	Figure showing probabilistic areas contributing recharge to priority wells under the “Maxratio” pumping scenario.....	39
28.	Figure showing probabilistic areas contributing recharge to priority wells under the “0.25ratio” pumping scenario.....	40

29. Figure showing probabilistic areas contributing recharge to priority wells under the “0.5ratio” pumping scenario.....	41
30. Figure showing probabilistic areas contributing recharge to priority wells under the “0.75rate” pumping scenario.....	42
31. Figure showing probabilistic areas contributing recharge to priority wells under the “Maxrange” pumping scenario	43

Tables

1. Neversink-Rondout priority and nonpriority production wells with 2010–2015 withdrawal rates.....	6
2. U.S. Geological Survey streamgages and data used for streamflow routing (SFR) inflow and base flow calibration targets.....	11
3. Summary of prior and posterior base realization hydraulic conductivities aggregated by geologic zones	12
4. Summary of observation data, observation groups, and initial and rebalanced weights.....	25
5. Summaries of initial parameter values and bounds during the prior Monte Carlo, expanded Monte Carlo, and iterative ensemble smoother history matching, aggregated by parameter type	27

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)
Area		
square mile (mi ²)	259.0	hectare (ha)
square mile (mi ²)	2.590	square kilometer (km ²)
Volume		
gallon (gal)	3.785	liter (L)
gallon (gal)	0.003785	cubic meter (m ³)
gallon (gal)	3.785	cubic decimeter (dm ³)
cubic foot (ft ³)	28.32	cubic decimeter (dm ³)
cubic foot (ft ³)	0.02832	cubic meter (m ³)
Flow rate		
cubic foot per second (ft ³ /s)	0.02832	cubic meter per second (m ³ /s)
gallon per minute (gal/min)	0.06309	liter per second (L/s)
inch per year (in/yr)	25.4	millimeter per year (mm/yr)
Hydraulic conductivity		
foot per day (ft/d)	0.3048	meter per day (m/d)

International System of Units to U.S. customary units

Multiply	By	To obtain
Length		
meter (m)	3.281	foot (ft)
kilometer (km)	0.6214	mile (mi)
meter (m)	1.094	yard (yd)
Area		
hectare (ha)	0.003861	square mile (mi ²)
square kilometer (km ²)	0.3861	square mile (mi ²)
Volume		
cubic meter (m ³)	264.2	gallon (gal)
cubic meter (m ³)	35.31	cubic foot (ft ³)
cubic meter (m ³)	1.308	cubic yard (yd ³)
cubic meter (m ³)	0.0008107	acre-foot (acre-ft)
Flow rate		
cubic meter per second (m ³ /s)	35.31	cubic foot per second (ft ³ /s)
liter per second (L/s)	15.85	gallon per minute (gal/min)
millimeter per year (mm/yr)	0.03937	inch per year (in/yr)
Hydraulic conductivity		
meter per day (m/d)	3.281	foot per day (ft/d)

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).

Elevation, as used in this report, refers to distance above the vertical datum.

Abbreviations

BFI	base-flow index
CHD	MODFLOW 6 Time-Variant Specified Head Package
HUC	hydrologic unit code
iES	iterative ensemble smoother
Kh	horizontal hydraulic conductivity
Kv	vertical hydraulic conductivity
NS–RO	Neversink-Rondout
NYSDEC	New York State Department of Environmental Conservation
NYSDOH	New York State Department of Health
NWIS	National Water Inventory System
PEST++	Parameter ESTimation Software
SFR	MODFLOW 6 Streamflow Routing Package
USGS	U.S. Geological Survey

Areas Contributing Recharge to Priority Wells in Valley-fill Aquifers in the Neversink River and Rondout Creek Drainage Basins, New York

By Nicholas T. Corson-Dosch¹, Michael N. Fienen¹, Jason S. Finkelstein¹, Andrew T. Leaf¹, Jeremy T. White², Joshua Woda¹, and John H. Williams¹

Abstract

In southeastern New York, the villages of Ellenville, Wurtsboro, Woodridge, the hamlet of Mountain Dale, and surrounding communities in the Neversink River and Rondout Creek drainage basins rely on wells that pump groundwater from valley-fill glacial aquifers for public water supply. Glacial aquifers are vulnerable to contamination because they are highly permeable and have a shallow depth to water table. To protect the quality of these water resources, water managers need accurate information about the areas that contribute recharge to production wells that pump from these aquifers. The New York State Department of Environmental Conservation and the New York State Department of Health designated eight priority wells in this region for which water supply protection is of primary concern.

The U.S. Geological Survey, in cooperation with the New York State Department of Environmental Conservation and the New York State Department of Health, began an investigation in 2019 with the general objectives of (1) improving understanding of regional groundwater-flow system, (2) delineating areas contributing recharge to eight priority production wells, and (3) quantifying the uncertainty of these contributing areas in a probabilistic way that can be used to inform decision-making related to priority well source-water protection. To complete these objectives, a MODFLOW 6 groundwater model was created encompassing the eight priority wells and the surrounding flow system, which includes parts of the Neversink River and Rondout Creek Basins in Sullivan County and Ulster County, New York. The model was built using Python tools (such as [flopy](#), [modflow-setup](#), and [sfr-maker](#)) that facilitate transparent and repeatable model development using existing datasets. The model parameters were estimated with a stepwise approach using an iterative ensemble smoother implementation of the Parameter ESTimation

software PEST++ (version 5.0.0). We evaluated initial “best guess” parameter bounds with a prior Monte Carlo analysis. Results of the first prior Monte Carlo analysis were used to make informed adjustments to model parameter bounds (typically resulting in expanded bounds), and a second prior Monte Carlo analysis was run to identify improved ranges for model parameters during history matching.

The history matching effort produced an ensemble of parameter values for the groundwater-flow model that spans the range of values within prior uncertainty bounds. The ensemble is informed by the historical observation data, within a reasonable range of uncertainty on those observations. This history-matched ensemble was used in a particle tracking Monte Carlo analysis to delineate the areas contributing recharge to priority wells. The groundwater-flow and particle tracking (MODPATH7) models were run once for each ensemble member. Deterministic contributing areas computed for each ensemble member were aggregated to produce maps showing the probability that a location contributes recharge to priority wells. Finally, the particle tracking Monte Carlo analysis was repeated for six pumping scenarios, representing a wide range of possible pumping levels, to incorporate uncertainty in future pumping rates related to population growth or other management decisions. Increasing pumping rates generally led to larger contributing recharge areas and larger areas of high probability that a location contributes recharge to priority wells. These maps show the overall uncertainty of the areas contributing recharge to priority wells in the study area and provide a tool for risk-based decision making for protection of well source water.

Introduction

Glacial aquifers are the principal source of water for many communities in upstate New York. The high permeability and the shallow depth to the water table in these aquifers make them highly susceptible to contamination. Sources of potential contamination are often related to land use; for example, leachate from septic systems in residential areas,

¹U.S. Geological Survey

²INTERA Geoscience and Engineering Solutions

stormwater runoff in developed areas, chemical spills or leaky petroleum-product storage tanks at commercial and industrial facilities, and pesticides and fertilizers from agricultural operations. Protecting these aquifers from contamination is critical in areas where groundwater use is high and alternative sources of drinking water are not readily available (Miller and others, 1998). Water managers need accurate information about the size and location of the land surface area that contributes recharge to production wells to make land-use planning decisions to reduce the risk of aquifer contamination.

The New York State Department of Environmental Conservation (NYSDEC) and the New York State Department of Health (NYSDOH) are working to reduce the risk of glacial aquifer contamination by delineating areas that contribute recharge to production wells and to encourage land-use planning within these zones. In southeastern New York, the villages of Ellenville, Wurtsboro, Woodridge, the hamlet of Mountain Dale, and surrounding communities rely on groundwater pumped from shallow valley-fill glacial aquifers for public water supply. In this region, which encompasses adjacent parts of the Neversink River and Rondout Creek drainage basins ([fig. 1](#)), the NYSDEC and NYSDOH have designated eight “priority” production wells, for which water supply protection is of primary concern. In 2019, in cooperation with the NYSDEC and the NYSDOH, the U.S. Geological Survey (USGS) began a study to characterize the regional groundwater-flow system and to delineate the contributing areas for these eight priority wells for the purpose of source water protection.

Purpose and Scope

This report describes the construction and history matching of a groundwater-flow model used for the purpose of delineating the areas contributing recharge to eight priority wells in the Neversink-Rondout (NS-RO) study area ([fig. 1](#)). The specific objectives of this study were to (1) construct a MODFLOW 6 groundwater-flow model using a Python-scripted workflow for repeatable and transparent model building; (2) do history matching using an iterative ensemble smoother (iES) to produce a posterior ensemble of model parameters, conditioned on available observations, to facilitate uncertainty analysis; (3) delineate probability-based areas contributing recharge to priority wells using a Monte Carlo analysis; and (4) produce maps of results to inform risk-based land management decisions that can be used to help protect source waters for priority wells.

This report includes (1) a brief overview of the physical setting of the NS-RO study area, (2) a description of the construction of the groundwater model and history matching, (3) a description of methods used to delineate the areas contributing recharge to priority wells with probabilistic uncertainty, and (4) a discussion of limitations and assumptions. An appendix is included ([appendix 1](#)) that provides additional information about the data sources used to construct the groundwater-flow model.

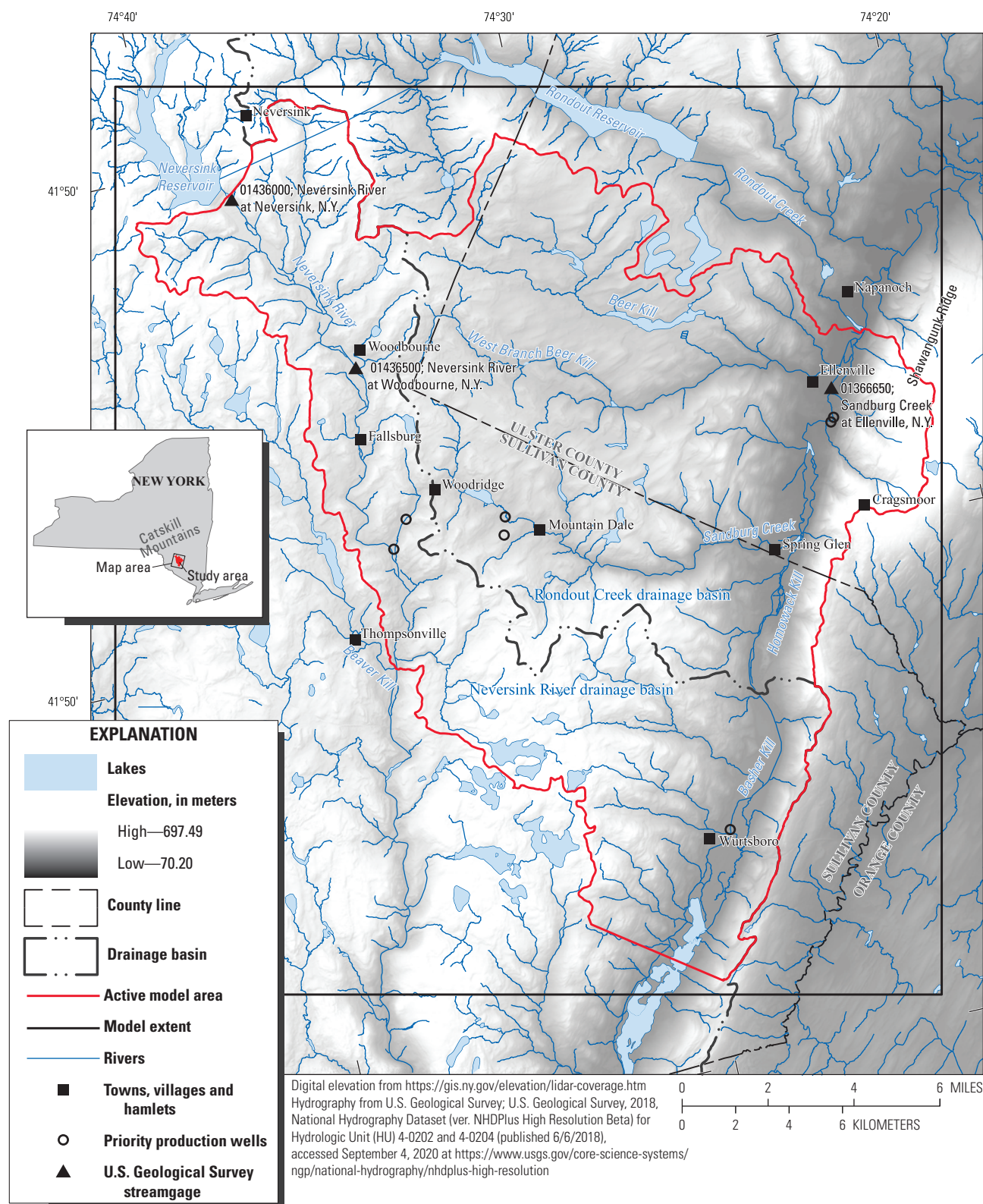


Figure 1. Location of the study area, model domain, and other key features.

Hydrogeologic Framework

The NS–RO study area covers about 183 square miles in eastern Sullivan County and western Ulster County, New York (fig. 1). The study area spans the boundary between the Appalachian Plateaus and the Valley and Ridge physiographic provinces (Fischer and others, 2004). Much of the study area is in the Catskill Mountains, which are part of the Appalachian Plateaus physiographic province. The Catskill Mountains are an uplifted, erosionally dissected plateau with high local relief. Multiple continental glaciations deeply eroded existing stream and river courses in the Catskill Mountains, incising steep valleys with narrow, flat floors that are underlain by stratified sediments (Randall, 2001). The study area is bounded to the east by Shawangunk Ridge, a high, steep bedrock ridge that is part of the physiographically distinct Valley and Ridge physiographic province (Wills and others, 2005). The total amount of relief in the study area is about 1,970 feet (ft; 600 meters).

The study area includes parts of the Neversink River and the Rondout Creek drainage basins. The Neversink River flows into the study area at the northwestern boundary at the outlet of the Neversink Reservoir. The Neversink River flows south across the study area and exits at the southeastern boundary just upstream from its confluence with Beaver Kill near Thompsonville, N.Y. The eastern study area covers a section of the Rondout Creek headwaters, including Sandburg Creek and its tributaries. Sandburg Creek originates near the center of the study area near Woodridge, N.Y., and flows to the northeastern boundary of the study area, above its confluence with Rondout Creek, at Napanoch, N.Y. Tributaries to Sandburg Creek include Beer Kill, West Branch Beer Kill, Homowack Kill, and several streams draining the western flank of Shawangunk Ridge.

Glacially deposited sediments fill the valleys and overlie bedrock in most of the uplands across the study area. The thicknesses of these deposits range from a few meters to more than 100 m (328 feet). Most of these materials were deposited during the advance and retreat of the last (late Wisconsin) continental ice sheet during the late Pleistocene (Randall, 2001). This ice sheet reached its terminus in northern New Jersey and Pennsylvania about 25,000 years ago and retreated northward through southeastern New York by about 21,000 years ago (Randall, 2001). The uplands in the study area are covered by a thin, discontinuous layer of till. The till consists of unsorted fine- and coarse-grained sediments (ranging from clay to boulders) deposited during the advance and retreat of glacial ice (Finkelstein and others, 2022). The valley-fill sequence consists primarily of three depositional facies:

- proximal facies, mostly consisting of sand, gravel, and silt. These facies are commonly referred to as “kame” deposits (Miller and others 1998, Randall, 2001; Finkelstein and others 2022). Where saturated, these deposits form unconfined, semiconfined, and (where overlain by mid-deglacial lacustrine deposits) confined sand and gravel aquifers.

- mid-deglacial facies, primarily lacustrine silts and clays. These layers can act as confining units in the valley-fill aquifer systems.
- late-deglacial facies, primarily well-sorted sands and gravels, deposited as alluvium and outwash. Where saturated, these deposits generally form unconfined, semiconfined, and (where overlain by mid-deglacial lacustrine deposits) confined valley-fill aquifers.

The mid-deglacial facies, consisting of lacustrine silts and clays, are generally the most extensive and thickest of the three facies in the study area. In most valley reaches all three facies are present. In general, the surficial deposits in the valleys are late-deglacial outwash sediments that form unconfined sand and gravel aquifers. These unconfined aquifers are perched above mid-deglacial lacustrine deposits (in the valley center) and overlie bedrock (along valley walls, where the underlying sediments pinch out). The mid-deglacial confining deposits overlie proximal kame deposits that form confined (or semiconfined) sand and gravel aquifers. The base of these confined aquifers is defined by the underlying bedrock surface. In some valley reaches, only one or two facies are present. The upper and lower aquifers are connected where the confining layer is absent. A schematic representation of the valley-fill aquifer system is shown in figure 2. A more detailed description of the regional geology, including the till and bedrock uplands and valley-fill aquifers in the study area, can be found in Finkelstein and others (2022).

The eight priority wells in the study area draw water from the valley-fill aquifer system. These wells are generally screened in transmissive proximal and late-deglacial deposits, such as kame, outwash, alluvium, and lacustrine sand. Information about well screen lithology is documented in table 1. Priority wells in the study area are also generally screened in unconfined or semiconfined valley-fill units, except for well SV 407 near Mountain Dale, N.Y., which is screened in kame deposits overlain by a mid-deglacial lacustrine silt and clay confining unit.

The bedrock in the study area primarily consists of Devonian and Silurian shale, sandstone, and conglomerate, and forms an undulating surface that bounds and underlies the glacial deposits in the valleys. The extents of the till-covered bedrock uplands and the valley-fill aquifers are shown in figure 3. The interior of the bedrock is assumed to be generally unweathered and sparsely fractured, but the surface of the bedrock may include a shallow zone of greater weathering and fracturing. In general, the permeability of the bedrock is low relative to the permeability of valley-fill deposits.

Recharge to the upland till and bedrock is principally from infiltration of precipitation and snowmelt but also, to a lesser extent, from leakage from upland streams. The valley-fill aquifers receive recharge from a variety of sources including infiltration from precipitation, runoff from unchanneled hillsides, and leakage from tributary streams that lose water as they flow onto the permeable valley-fill deposits. Upland

6 Areas Contributing Recharge to Priority Wells in the Neversink River and Rondout Creek Drainage Basins

Table 1. Neversink-Rondout priority and nonpriority production wells with 2010–2015 withdrawal rates

[USGS, U.S. Geological Survey; NYSDC, New York State Department of Environmental Conservation; gal/min, gallons per minute; NYSDOH, New York State Department of Health; --, data not available. Mean and maximum pumping rates for the period 2010–2015. Mean pumping rates were used in the model and were adjusted ± 20 percent during calibration. Maximum pumping rates were used to evaluate uncertainty of the areas contributing recharge to priority wells (see figures 26–31).]

USGS well ID	USGS site number	NYSDEC facility ID	Nearest town/city/ village	County	Surface- water basin	Withdrawal rate ^a , (gal/min)		Well screen	Model layer
						Mean	Maximum ^b	Model lithologic unit	
NYSDEC-NYSDOH priority wells									
SV 193	414210074352501	WWR0000401	Woodridge	Sullivan	Neversink	11	75	Kame	3
SV 227	414138074355701	WWR0000401	Woodridge	Sullivan	Neversink	35	230	Alluvium & outwash	1
U 55	414212074233001	WWR0000479	Ellenville	Ulster	Rondout	43	132	Lacustrine sand	2
U 917	414210074233601	WWR0000479	Ellenville	Ulster	Rondout	182	500	Kame	2
U 7	414212074233201	WWR0000479	Ellenville	Ulster	Rondout	97	500	Kame	2
U 8	414201074230201	WWR0000479	Ellenville	Ulster	Rondout	207	611	Kame	2
SV 407	414147074324701	WWR0001055	Mountain Dale	Sullivan	Rondout	18	120	Kame	3
SV 487	413438074284401	WWR0001748	Wurtsboro	Sullivan	Neversink	75	300	Outwash & alluvial fan combination	1
Nonpriority wells									
--	--	WWR0000209	Woodbourne	Sullivan	Neversink	0.2	--	Bedrock	4
--	--	WWR0000209	Woodbourne	Sullivan	Neversink	0.2	--	Bedrock	4
--	--	WWR0000209	Woodbourne	Sullivan	Neversink	1	--	Bedrock	4
--	--	WWR0000209	Woodbourne	Sullivan	Neversink	1	--	Bedrock	4
SV 61	414301074364401	WWR0000506	Fallsburg	Sullivan	Neversink	133	--	Kame	3
SV 62	414406074361201	WWR0000506	Fallsburg	Sullivan	Neversink	68	--	Bedrock	4
--	--	WWR0000506	Fallsburg	Sullivan	Neversink	71	--	Bedrock	4
SV 63	414419074360701	WWR0000506	Fallsburg	Sullivan	Neversink	71	--	Kame	3
SV 266	414422074360701	WWR0000506	Fallsburg	Sullivan	Neversink	249	--	Alluvium & outwash	1
--	--	WWR0000506	Woodbourne	Sullivan	Neversink	71	--	Bedrock	4
SV 256	414443074355101	WWR0000506	Woodbourne	Sullivan	Neversink	86	--	Bedrock	4
--	--	WWR0000506	Woodbourne	Sullivan	Neversink	116	--	Bedrock	4
SV 406	414128074315601	WWR0001055	Mountain Dale	Sullivan	Rondout	17	--	Bedrock	4
SV 422	413913074263001	WWR0001462	Spring Glen/ Phillipsport	Sullivan	Rondout	6	--	Bedrock	4
SV 510	414457074354701	WWR0001734	Woodbourne	Sullivan	Neversink	15	--	Alluvium & outwash	1
SV 511	414455074354601	WWR0001734	Woodbourne	Sullivan	Neversink	16	--	Bedrock	4
SV 512	414453074351001	WWR0001734	Woodbourne	Sullivan	Neversink	16 ^c	--	Bedrock	4
SV 513	414441074354701	WWR0001734	Woodbourne	Sullivan	Neversink	16	--	Bedrock	4
SV 514	414450074353002	WWR0001734	Woodbourne	Sullivan	Neversink	23	--	Bedrock	4
SV 515	414439074354901	WWR0001734	Woodbourne	Sullivan	Neversink	35	--	Bedrock	4
SV 516	414450074353001	WWR0001734	Woodbourne	Sullivan	Neversink	35 ^c	--	Bedrock	4
SV 463	414236074335601	WWR0001738	Woodridge	Sullivan	Rondout	36	--	Alluvium & lacustrine silt and clay combination	1
SV 465	414236074335501	WWR0001738	Woodridge	Sullivan	Rondout	49	--	Bedrock	4
SV 470	414222074302101	WWR0001738	Mountain Dale	Sullivan	Rondout	57 ^c	--	Bedrock	4
SV 490	413437074284501	WWR0001748	Wurtsboro	Sullivan	Neversink	19	--	Bedrock	4
SV 551	413434074291201	WWR0001748	Wurtsboro	Sullivan	Neversink	44	--	Lacustrine silt & clay	2

^aWell withdrawal rate information compiled from USGS Site-Specific Water-Use Data System (SWUDS); data representative of 2010–2015 period.

^bMaximum withdrawal rates simulated only at priority wells.

^cPumping source of numerical instability in model; withdrawal rate set to zero in model simulations.

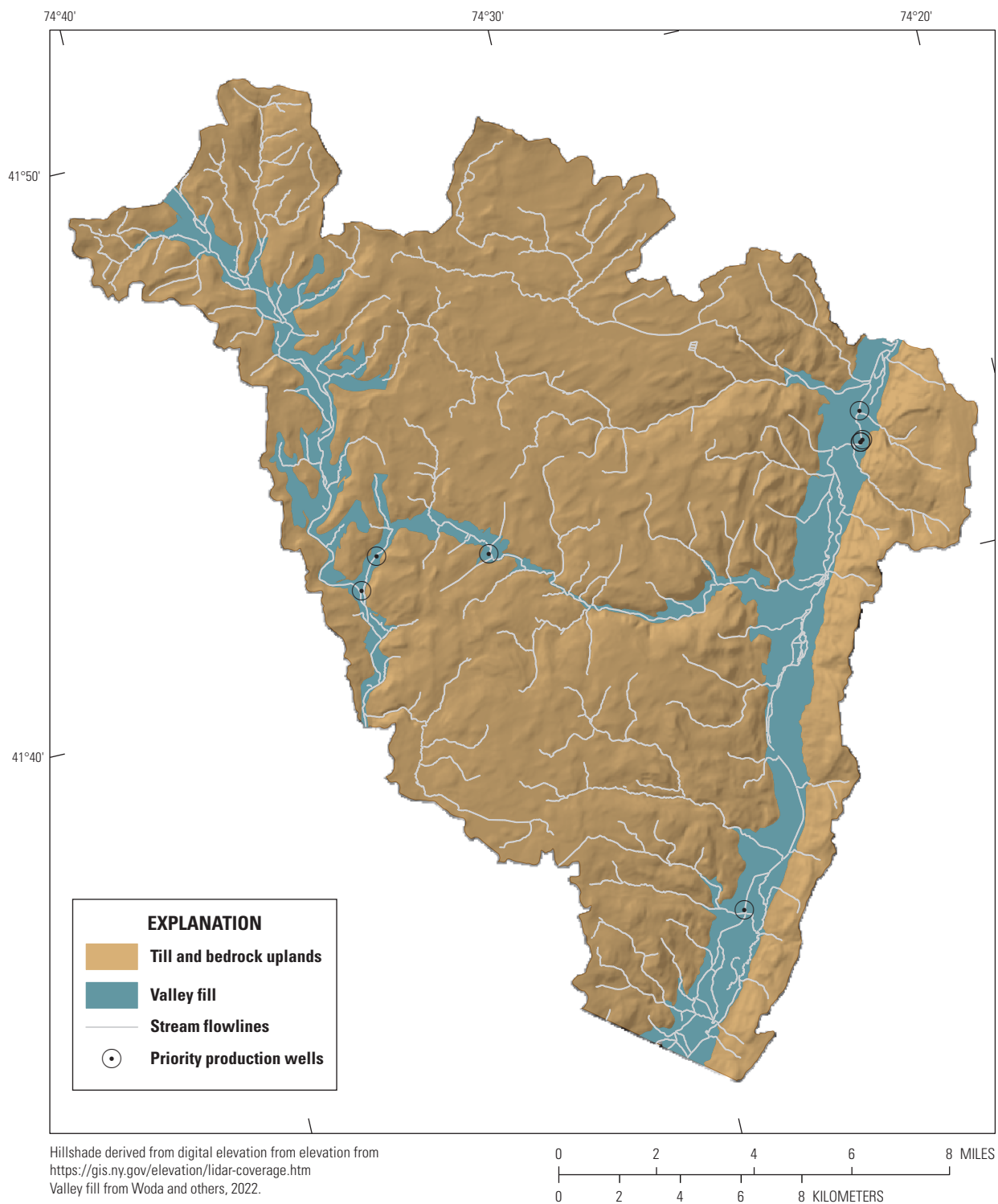


Figure 3. Extent of valley-fill aquifers and till-covered bedrock uplands with priority well locations.

Development of Steady-State Numerical Model

A steady-state, finite-difference MODFLOW 6 (Langevin and others, 2017) groundwater-flow model was developed to simulate the groundwater-flow system in the NS–RO study area. The complete model development process, from construction through history matching and postprocessing, was scripted using the Python programming language (Van Rossum and Drake, 2009) to produce a transparent and repeatable record of the modeling process. We constructed the groundwater-flow model using several USGS-developed Python packages designed to facilitate MODFLOW model development, including *flopy* (Bakker and others, 2016), *sfrmaker* (Leaf and others, 2021), and *modflow-setup*. These tools were used to develop a MODFLOW 6 model and input files from source data, including shapefiles, rasters, and other MODFLOW models. Source data—representing aquifer system geometry, boundary conditions, and hydraulic properties—were read from their native formats and mapped to a regular finite difference grid using *modflow-setup*. The gridded model data for the desired MODFLOW packages generated by *modflow-setup* were then used to construct MODFLOW input files with *flopy*. We estimated parameters for the groundwater-flow model through history matching to be consistent with available observation data including hydraulic heads and flows representing steady-state conditions. We completed history matching using the iES implementation (White, 2018) of the Parameter ESTimation (PEST) software PEST++ (version 5.0.0; White and others, 2021). The resulting parameter ensembles facilitate uncertainty analysis through use of the *pyemu* (White and others, 2016) Python package.

A summary of the groundwater model design, history matching, and parameter optimization are provided in the following sections. MODFLOW input files were generated using georeferenced source data, including shapefiles, rasters, and a parent MODFLOW model. A detailed description of the geospatial source data used to construct the groundwater model is presented in appendix 1. The groundwater-flow model and the Python code used in the modeling workflow are available in the companion data release (Fienen and Corson-Dosch, 2021; https://github.com/usgs/neversink_workflow).

Model Extent and Discretization

The MODFLOW model domain consists of a uniform grid of 680 rows and 619 columns of 50-m-square cells. The model has four layers that represent the glacially deposited sediments and bedrock, with layers 1, 2, and 3 representing surficial glacially-deposited sediments—including outwash, alluvium, kame, lacustrine sediments, and till—and layer 4 representing the underlying bedrock. Layers 1–3 varied in thickness across the model domain and were spatially discontinuous. The bedrock unit, represented by layer 4, was set

to a uniform thickness of 30 m. Model top elevations were derived from lidar data (U.S. Geological Survey, 2015) by sampling the mean elevation within each cell and mapping that value to the model grid. Model layer elevations were developed using a variety of information, including well construction reports, boring logs, geologic cross-section data, Soil Survey Geographic Database (SSURGO) data, and lidar data. The methods used to develop model layer elevations and thicknesses are described in detail by Finkelstein and others (2022) and the data sets are made available in the USGS data release by Woda and others (2022).

The lithology of the valley-fill aquifer systems (represented in model layers 1–3) is complex. The individual units that comprise these layers have irregular geometries that interfinger and pinch out (fig. 2, for example). The lateral continuity of the units also is variable—valley-fill facies present in one valley reach may be different than those present in another. These complex aquifer geometries were simulated using the MODFLOW 6 idomain array to specify active (idomain=1), inactive (idomain=0), and vertical pass-through (idomain=-1) cells. Vertical pass-through cells were used to route water in the model between the uppermost active and the next highest active layer; for example, if the geologic units represented by model layers 2 and 3 pinched out along the edge of a valley, vertical passthrough cells applied to layers 2 and 3 allow water to be exchanged between layer 1 and layer 4 directly, without the need for the intermediate layers to be included. This improves the computational efficiency of the model and the accuracy of the model's physical representation of the flow system.

The geographic extent of the model was established using the topographical divides defined by 5 adjacent 12-digit hydrologic unit code (HUC) drainage basins. These drainage basins include the Upper Neversink River (020401040305), Gumaer Brook-Basher Kill (020401040201), Basher Kill (020401040203), Beer Kill (020200070503), and Sandburg Creek (020200070504). The combined extent of these drainage basins defined the active model domain with two modifications: (1) the model boundary downgradient from the Neversink Reservoir was adjusted about 0.25 mile southeast—from the top of the Neversink dam structure to its base—to improve model stability; and (2) the model extent in Basher Kill River drainage basin was truncated south of Wurtsboro, N.Y., downgradient from any priority wells to improve the computational efficiency of the groundwater-flow model. The spatial extent of the model domain is shown in figure 1.

Boundary Conditions

Hydrologic boundary conditions represent flows of water into and out of the model domain. Boundary conditions in this model include areal recharge from net infiltration percolating below the root zone, groundwater flow across the model perimeter in the valley-fill areas, stream-aquifer interactions, and groundwater extraction at wells.

Recharge

Steady-state recharge rates were estimated for each model cell using a gridded estimate of mean recharge (Yager and others, 2018, 2019). Yager and others (2019) calculated mean net infiltration recharge rates during 2000–2013 for the glaciated conterminous United States at a 250-m grid cell resolution using the Soil Water Balance code (SWB; Westenbroek and others, 2010). SWB uses spatially distributed variables, such as soil type, land use, climate data, and soil water capacity to partition precipitation into runoff, evapotranspiration, and deep infiltration. At this regional scale, all deep infiltration is assumed to enter the groundwater system as recharge.

The gridded mean deep infiltration product from Yager and others (2019) was resampled to the finer 50-m MODFLOW groundwater model grid using the [modflow-setup](#) Python package. The mean annual recharge rate across the study area for the 2000–2013 period was 14.9 inches per year. These estimates of recharge were assumed to be representative of overall steady-state conditions and were adjusted during groundwater model history matching to reflect uncertainty in the recharge estimate.

Groundwater Flow Across Model Perimeter

Groundwater flow across the model perimeter in valleys were simulated using the MODFLOW 6 specified head boundary package (CHD). CHD boundaries for this model were developed from two previously published regional groundwater-flow models (Zell and Sanford, 2020a; Zell and Sanford, 2020b). Zell and Sanford (2020a) produced 75 steady-state MODFLOW 6 groundwater-flow models that span the contiguous United States and simulate the long-term mean surficial groundwater system. The NS–RO model domain intersects parts of two Zell and Sanford (2020a) regional models. The discretization packages (DIS) and binary head output files from the two intersecting Zell and Sanford (2020a) models were combined to create a single synthetic “parent” model. NS–RO CHD values were extracted from the “parent” model head output using the [modflow-setup](#) Python package. The locations of CHD cells in the groundwater-flow model and their corresponding values are shown on [figure 4](#).

Stream-Aquifer Interactions

Stream-aquifer interactions were simulated using the MODFLOW 6 implementation of the Streamflow Routing (SFR) Package (Langevin and others, 2017). The SFR package simulates stream stage and accumulated base flow by accounting for gains and losses of water in each stream cell and routing streamflow from upstream cells to downstream cells. The SFR network was developed using hydrography from the U.S. Geological Survey National Hydrography Dataset Plus High Resolution (NHDPlus HR; U.S. Geological Survey, 2018). MODFLOW SFR input files were constructed using the

[sfrmaker](#) (Leaf and others, 2021) Python package—a tool that facilitates construction of streamflow routing networks from hydrography data—using NHDPlus HR flowlines and the associated streamflow routing and physical property information from 4-digit HUC 0202 and HUC 0204. Additional flowlines were added to the SFR network in select locations to reduce simulated groundwater flooding near priority wells. These flowlines were added manually following topography in drainages where NHDPlus HR flowlines were not already present. The MODFLOW 6 SFR flowlines are shown on [figure 4](#).

The Neversink Reservoir impounds flows from the Neversink River northwest of the model boundary. Outflow from the Neversink Reservoir discharges back to the Neversink River which in turn flows into the northwest part of the groundwater-flow model. Continuous streamflow is recorded at USGS streamgage 01436000 (Neversink River streamgage at Neversink, N.Y.), where the Neversink River flows into the model domain at the base of the Neversink Reservoir ([fig. 1](#)). The median daily flow rate from streamgage 01436000 (20.0 cubic feet per second; U.S. Geological Survey, 2020a) was applied as an inflow to the SFR reach of the Neversink River at the northwest model boundary ([table 2](#)). This inflow was assumed to be representative of overall steady-state conditions.

Water Withdrawals

Estimates of groundwater pumping from production wells—priority and nonpriority wells—were developed from 2010 through 2015 pumping records available in the USGS Site-Specific Water-Use Data System (SWUDS) database ([table 1](#); [fig. 4](#)). It was assumed that the 2010–2015 mean well withdrawals adequately represented pumping rates from these wells during steady-state conditions; however, pumping rates were adjusting during history matching to reflect the uncertainty in these estimates. Priority wells were assigned to model layers based on lithology. Nonpriority wells were assigned to the thickest model layer intersecting with the well open interval.

Relatively small amounts of groundwater are pumped for agricultural uses in Sullivan and Ulster Counties (Fischer and others, 2004; U.S. Geological Survey, 2020a) and no irrigation-related groundwater pumping or return flows were represented in the model. Additionally, withdrawals from smaller wells were not included. It was assumed that these wells are negligible at the scale of the model because they return most water through on-site septic systems, resulting in little-to-no net withdrawal.

Hydraulic Properties

Hydraulic properties in the model were defined by zones based on lithology data and surficial geology mapping. The initial hydraulic conductivity values assigned to these zones

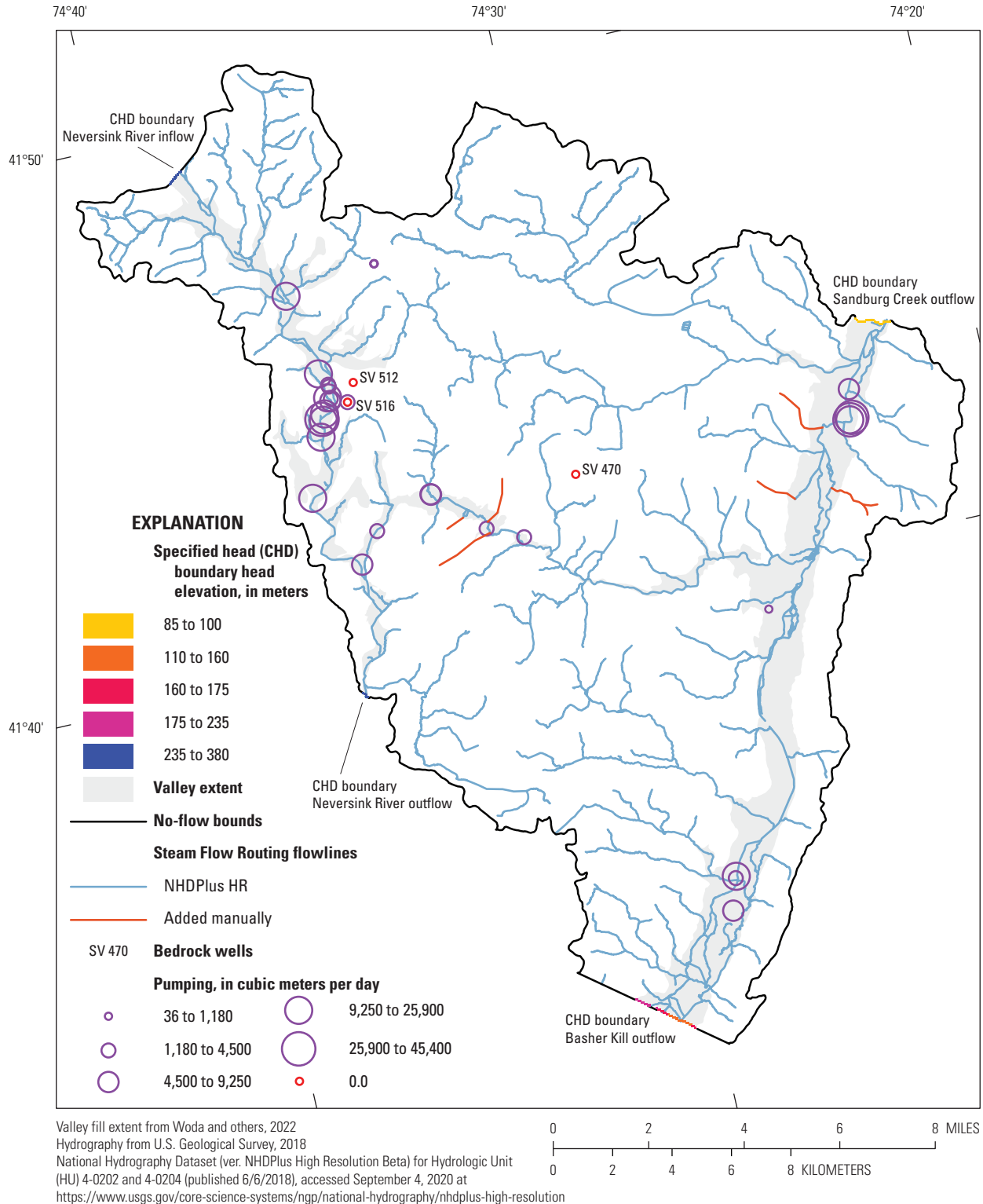


Figure 4. Model perimeter boundary conditions, streamflow routing (SFR) lines, and pumping wells in groundwater-flow model. Perimeter boundary cells intersecting valleys are represented by specified head (CHD) cells. Pumping well symbols (purple circles) are scaled to base realization pumping rate, except for three bedrock wells (red circles) where pumping rates were manually set to zero. CHD cell colors are scaled to CHD head elevations. Parts of the perimeter represented with no-flow boundary conditions are in gray.

Table 2. U.S. Geological Survey streamgages and data used for streamflow routing (SFR) inflow and base flow calibration targets.

[USGS, U.S. Geological Survey; BFI, base-flow index; SFR, MODFLOW 6 Streamflow Routing package; ft³/sec, cubic feet per second; N.Y., New York; Dates given in mm/dd/yyyy. Longitude and latitude are given in North American Datum of 1983]

SFR inflow							
Station name	USGS site number	Latitude	Longitude	Period of record used	Median daily flow rate (ft³/sec)		
Neversink River at Neversink, N.Y.	01436000	41.8200	-74.6356	10/1/1954– 9/30/1993 ^a	20.0		
Base-flow flux calibration targets							
Station name	USGS site number	Latitude	Longitude	Period of record used	Mean annual flow (ft³/sec)	Mean BFI value ^b	Base-flow calibration target (ft³/sec)
Neversink River at Woodbourne, N.Y.	01436500	41.7568	-74.5968	10/1/1957– 9/19/1977 ^c	91	0.522	47.5
Sandburg Creek at Ellenville, N.Y.	01366650	41.7142	-74.3892	10/1/1954– 9/30/1993	101	0.502	50.7

^aPeriod of record that overlaps with operation of downstream streamgage 01436500 (Neversink River at Woodbourne, N.Y.) and excludes period of record before the construction of the Neversink Reservoir.

^bU.S. Geological Survey, 2020b.

^cPeriod of record that overlaps with upstream streamgage 01436000 (Neversink River at Neversink, N.Y.) and after construction of the Neversink Reservoir.

(and initial upper and lower calibration bounds) were estimated using published regional aquifer test data (Como, 2013; Smith and Williams, 2020). The lithologic zones are shown, by layer, in panel “B” of figures 5–12. Initial (prior) hydraulic conductivities, aggregated by geologic zones, are summarized in table 3. Information about the history matching process and the development of posterior hydraulic properties are presented in the “Parameter Estimation by Ensemble History Matching” section.

Parameter Estimation by Ensemble History Matching

We did history matching to systematically adjust parameter values in the model such that they were consistent with historical observations of model outputs including hydraulic head and base-flow values. The general history matching approach follows Bayes’ theorem (Tarantola, 2005) for parameter estimation. In this approach, we start with a prior estimate of model parameters and their credible range (in other words, uncertainty). These prior parameter values and uncertainty are informed by expert knowledge, literature values, and available

direct measurements. Through a systematic conditioning step, both the parameter values and the uncertainty are updated to be consistent with observations that correspond to model outputs. This step is referred to as an “update” and results in a posterior set of parameter values and uncertainty range (“posterior” meaning “after the update”). There are many algorithms that can be used to complete this conditioning, and for this work we used the iES implementation (White, 2018) of PEST++ (version 5.0.0; White, 2018; White and others, 2021).

iES is an ensemble method, meaning at every stage of analysis, an ensemble of parameter sets (realizations) consistent with their inherent uncertainty and the assumed uncertainty in the observations is available. Each parameter set from this ensemble can be run through the model to provide a range of model outputs. iES uses empirical correlations between parameter and observation ensembles to iteratively reduce the uncertainty and discrepancy between model outputs and observations. iES produces a posterior parameter ensemble that is characteristic of the inherent uncertainty in the parameters and model and conditioned on the available data. This ensemble can then be used for predictive analysis—in this case, the delineation of probabilistic areas contributing water to priority production wells with results that span the likely range of outcomes.

12 **Areas Contributing Recharge to Priority Wells in the Neversink River and Rondout Creek Drainage Basins**

Table 3. Summary of prior and posterior base realization hydraulic conductivities aggregated by geologic zones.

[Kh, horizontal hydraulic conductivity; m/day, meters per day; Kv, vertical hydraulic conductivity; Prior, pre-history matching (or starting) value; Posterior, post-history matching value; Base, base realization representing the central tendency of the posterior parameter distribution; Min, minimum; Max, maximum; Anisotropy values are unitless]

Lithologic zone and description	Prior Kh (m/day)			Posterior base Kh (m/day)			Prior Kv (m/day)			Posterior base Kv (m/day)			Prior anisotropy			Posterior base anisotropy		
	Min	Mean	Max	Min	Mean	Max	Min	Mean	Max	Min	Mean	Max	Min	Mean	Max	Min	Mean	Max
11: Outwash	60.96	60.96	60.96	28.25	35.21	47.07	6.10	6.10	6.10	2.93	11.71	15.20	0.10	0.10	0.10	0.07	0.34	0.53
12: Peat and outwash	60.96	60.96	60.96	152.00	152.00	152.00	6.10	6.10	6.10	3.69	3.75	3.81	0.10	0.10	0.10	0.02	0.02	0.03
15: Kame	22.86	22.86	60.96	2.82	14.85	129.01	2.29	2.29	6.10	0.24	0.72	4.18	0.10	0.10	0.10	0.02	0.09	0.18
17: Alluvium and outwash	45.72	45.72	60.96	1.35	3.08	63.34	4.57	4.57	6.10	0.22	3.78	5.15	0.10	0.10	0.10	0.004	2.24	3.27
18: Kame moraine	9.14	9.14	9.14	2.57	3.97	5.40	0.91	0.91	0.91	0.32	5.66	15.20	0.10	0.10	0.10	0.08	1.35	4.26
20: Till and bedrock	0.17	0.23	60.96	1.87	2.74	152.00	0.02	0.02	6.10	0.68	4.23	15.20	0.10	0.10	0.10	0.10	1.80	2.82
21: Till	0.17	0.17	0.17	1.39	18.09	22.77	0.02	0.02	0.02	0.13	0.24	0.92	0.10	0.10	0.10	0.01	0.01	0.61
22: Bedrock	0.17	0.17	0.17	0.50	1.41	5.89	0.02	0.02	0.02	0.06	0.53	0.70	0.10	0.10	0.10	0.04	0.38	0.60
23: Lacustrine silt and clay	0.05	0.06	22.86	0.36	12.33	152.00	0.01	0.01	2.29	0.20	0.28	15.20	0.10	0.10	0.10	0.01	0.07	0.95
24: Outwash and alluvial fan combination	45.72	45.72	45.72	83.58	144.49	152.00	4.57	4.57	4.57	2.79	9.92	13.29	0.10	0.10	0.10	0.03	0.07	0.09
25: Kame and alluvial fan combination	22.86	22.86	22.86	31.11	38.88	45.33	2.29	2.29	2.29	1.57	1.76	2.23	0.10	0.10	0.10	0.04	0.05	0.05
26: Peat outwash and alluvium combination	22.86	22.86	22.86	152.00	152.00	152.00	2.29	2.29	2.29	15.20	15.20	15.20	0.10	0.10	0.10	0.10	0.10	0.10
27: Peat and kame combination	22.86	22.86	22.86	152.00	152.00	152.00	2.29	2.29	2.29	1.45	1.50	1.66	0.10	0.10	0.10	0.01	0.01	0.01
28: Alluvium and kame combination	22.86	22.86	22.86	10.48	15.75	19.40	2.29	2.29	2.29	6.21	10.28	12.14	0.10	0.10	0.10	0.45	0.66	0.87
29: Artificial fill	3.81	3.81	3.81	152.00	152.00	152.00	0.38	0.38	0.38	15.20	15.20	15.20	0.10	0.10	0.10	0.10	0.10	0.10
30: Lacustrine sand	3.81	3.81	3.81	0.33	4.42	11.65	0.38	0.38	0.38	0.10	0.58	1.20	0.10	0.10	0.10	0.01	1.30	2.76
31: Lacustrine sand and alluvium combination	8.38	8.38	8.38	152.00	152.00	152.00	0.84	0.84	0.84	6.79	7.62	8.22	0.10	0.10	0.10	0.04	0.05	0.05
32: Till/alluvium combination	1.68	1.68	1.68	16.32	17.02	17.66	0.17	0.17	0.17	7.32	7.67	7.98	0.10	0.10	0.10	0.45	0.45	0.45
33: Lacustrine sand and alluvium fan combination	8.38	8.38	8.38	152.00	152.00	152.00	0.84	0.84	0.84	14.27	15.18	15.20	0.10	0.10	0.10	0.09	0.10	0.10
34: Lacustrine silt and alluvium fan combination	1.68	1.68	1.68	152.00	152.00	152.00	0.17	0.17	0.17	1.81	2.23	2.67	0.10	0.10	0.10	0.01	0.01	0.02
35: Alluvium and lacustrine silt and clay combination	8.38	8.38	8.38	12.24	15.65	18.07	0.84	0.84	0.84	1.58	1.79	2.18	0.10	0.10	0.10	0.10	0.12	0.14
36: Till and kame combination	0.17	9.12	9.14	1.79	89.78	111.27	0.02	0.91	0.91	0.05	2.44	2.94	0.10	0.10	0.10	0.02	0.03	0.04
37: Till and outwash combination	9.14	9.14	9.14	97.40	118.60	145.05	0.91	0.91	0.91	1.89	2.30	2.77	0.10	0.10	0.10	0.01	0.02	0.03
38: Kame and water	22.86	22.86	22.86	103.74	117.42	132.79	2.29	2.29	2.29	2.34	2.51	2.73	0.10	0.10	0.10	0.02	0.02	0.02
39: Peat and till combination	0.17	0.17	0.17	42.40	53.62	67.10	0.02	0.02	0.02	1.42	1.75	2.10	0.10	0.10	0.10	0.02	0.03	0.05

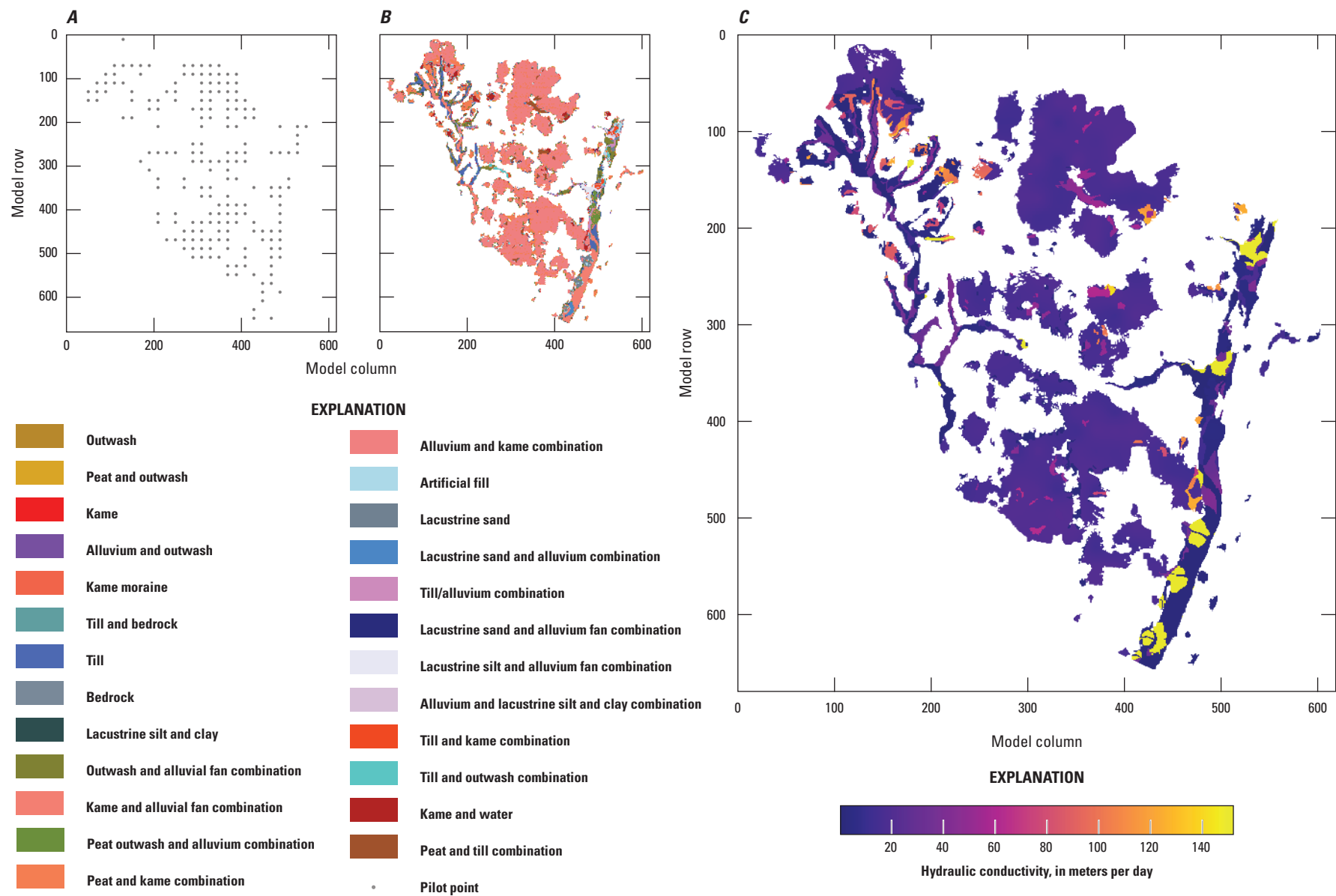


Figure 5. Base realization values and parameterization for horizontal hydraulic conductivity in model layer 1 with, *A*, pilot point network, *B*, lithologic zones, and, *C*, base realization conductivity array.

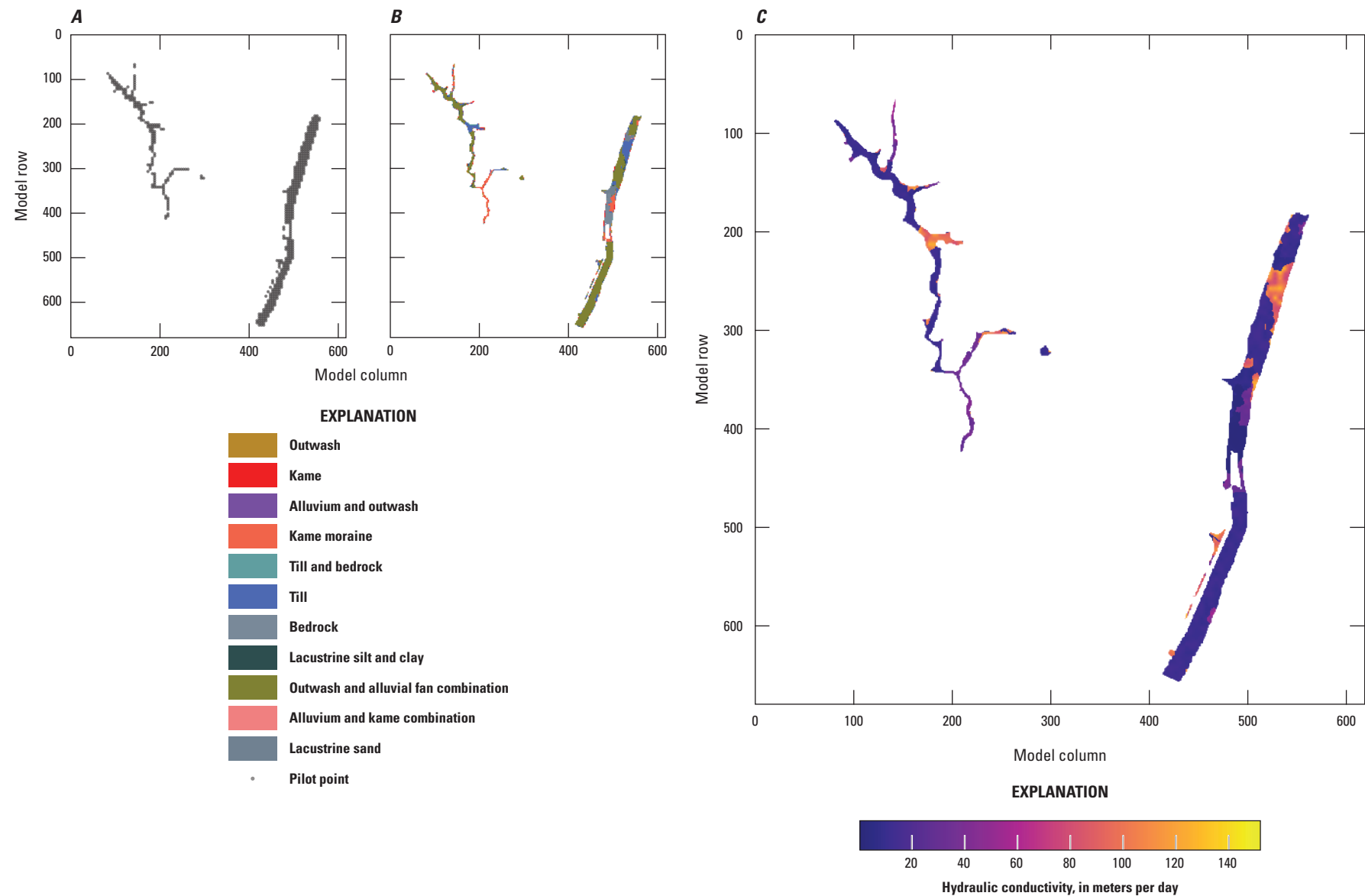


Figure 6. Base realization values and parameterization for horizontal hydraulic conductivity in model layer 2 with, *A*, pilot point network, *B*, lithologic zones, and, *C*, base realization conductivity array.

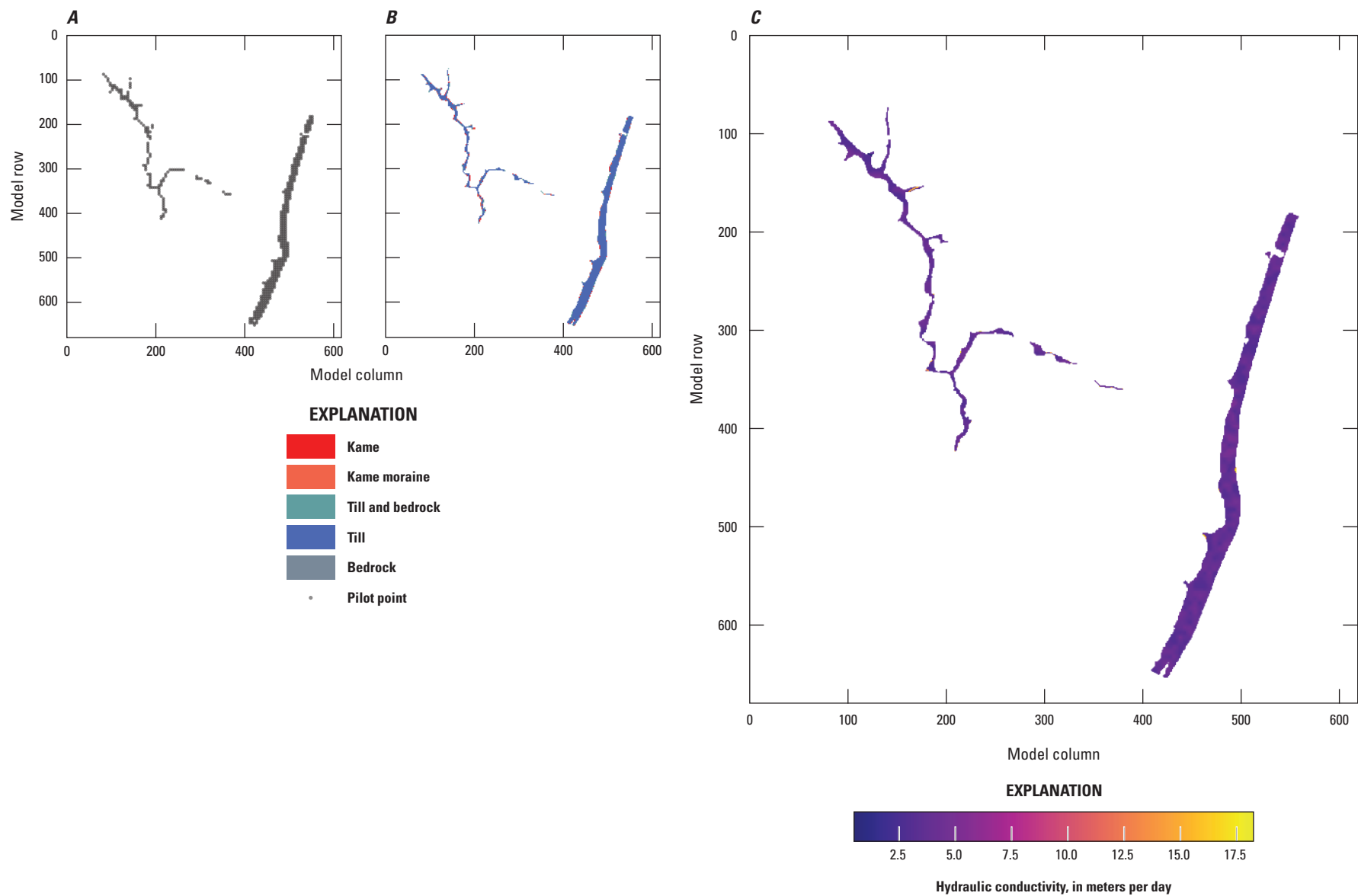


Figure 7. Base realization values and parameterization for horizontal hydraulic conductivity in model layer 3 with, *A*, pilot point network, *B*, lithologic zones, and, *C*, base realization conductivity array.

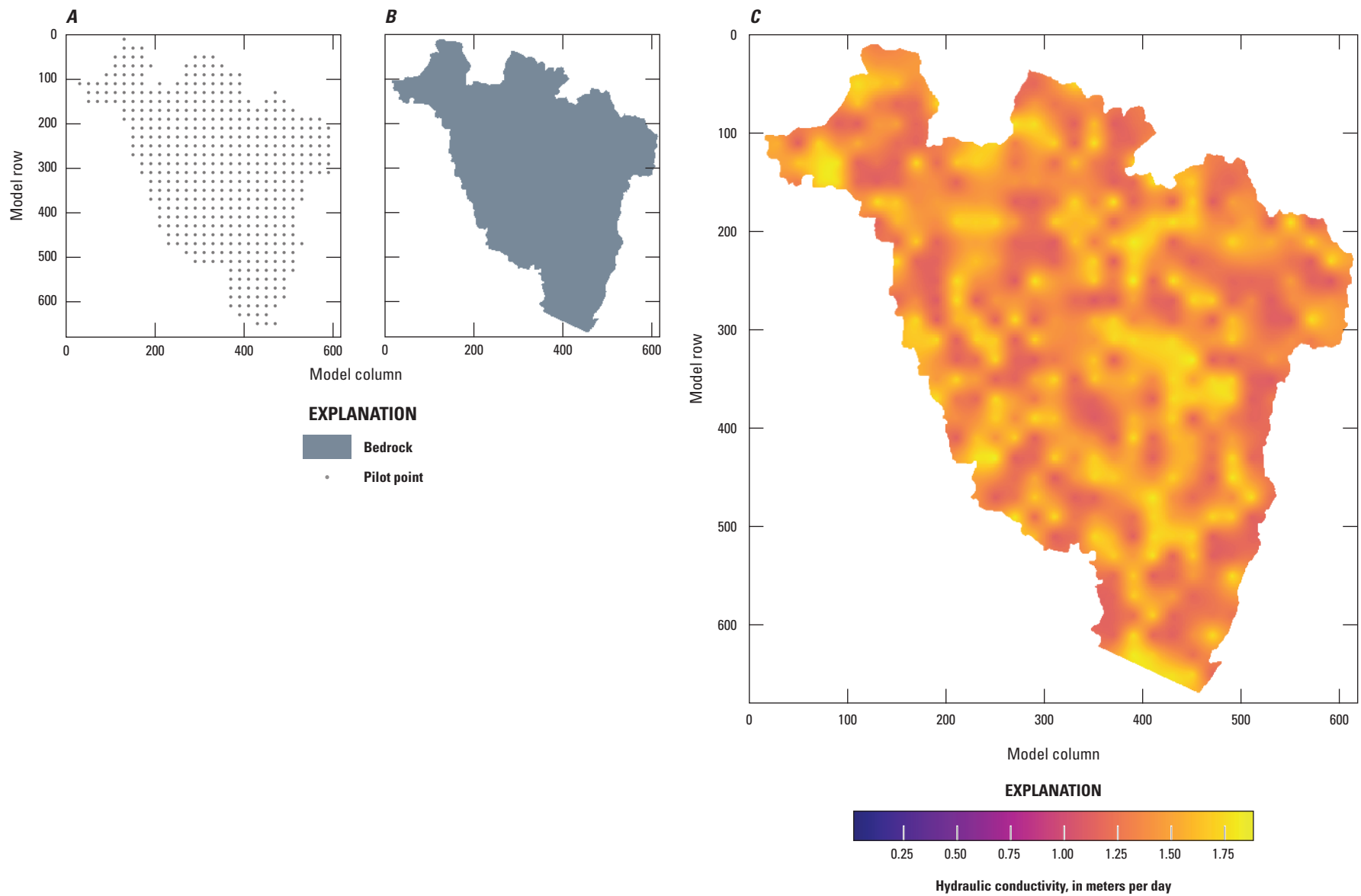


Figure 8. Base realization values and parameterization for horizontal hydraulic conductivity in model layer 4 with, *A*, pilot point network, *B*, lithologic zones, and, *C*, base realization conductivity array.

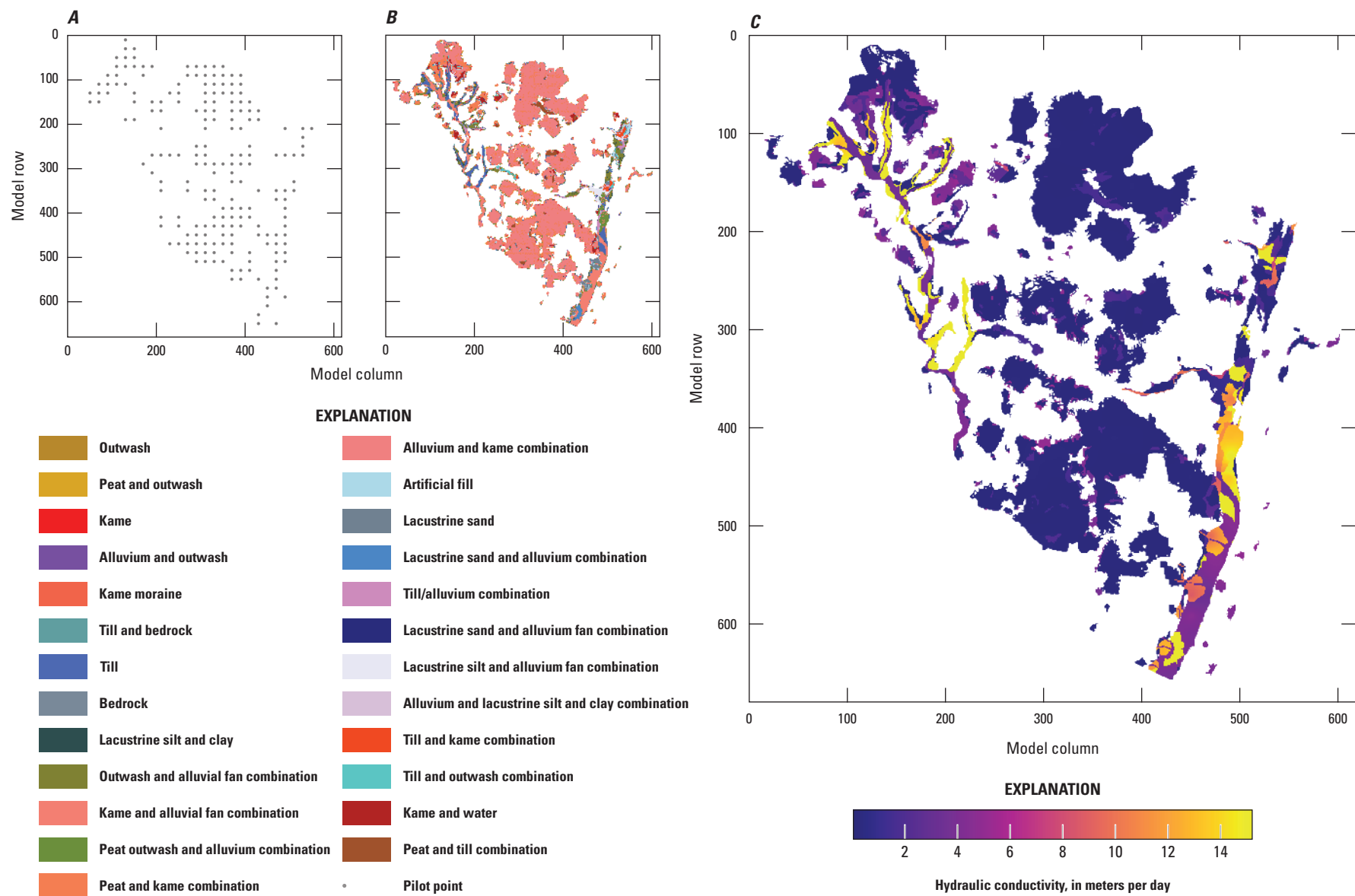


Figure 9. Base realization values and parameterization for vertical hydraulic conductivity in model layer 1 with, *A*, pilot point network, *B*, lithologic zones, and, *C*, base realization conductivity array.

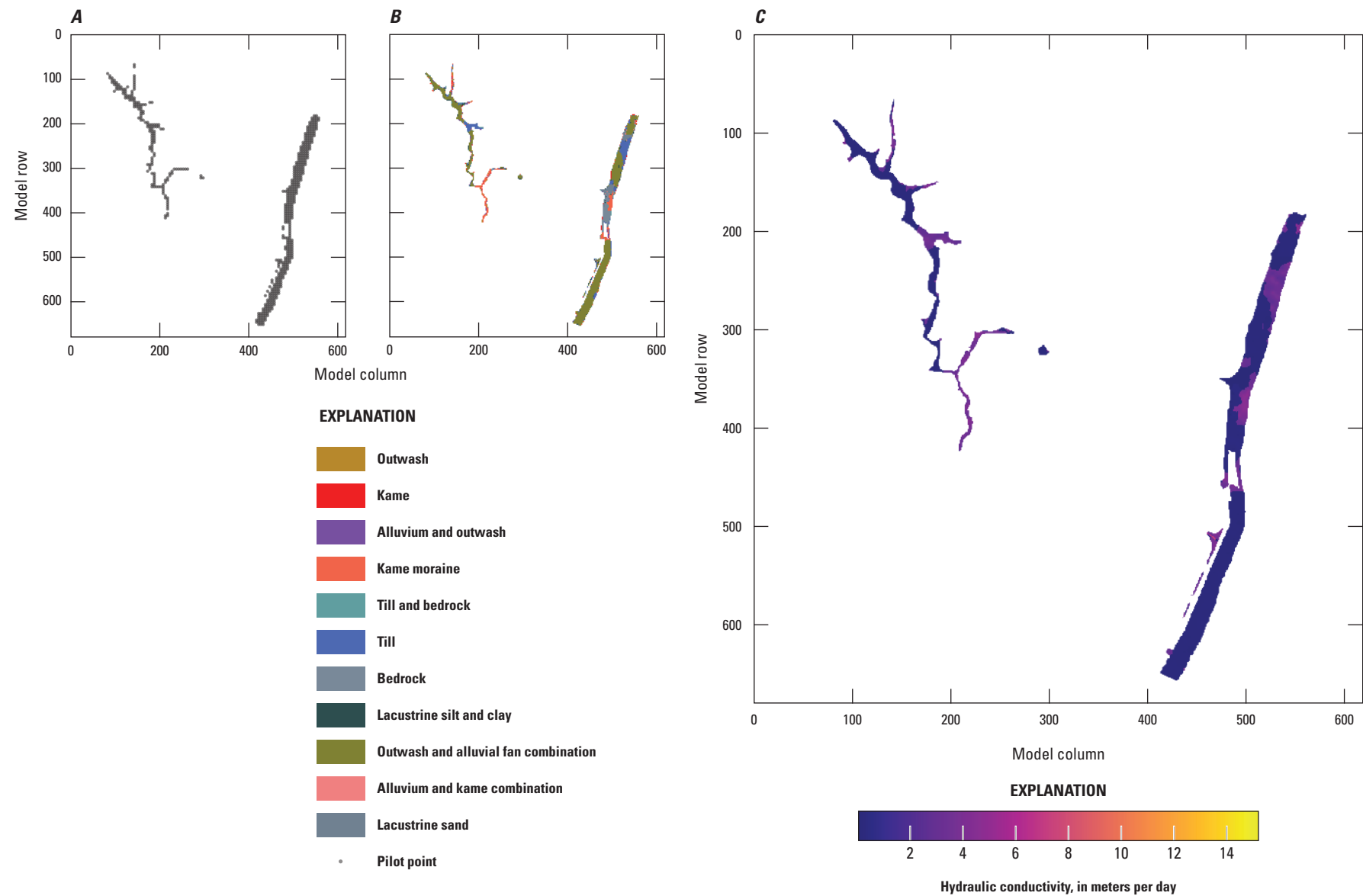


Figure 10. Base realization values and parameterization for vertical hydraulic conductivity in model layer 2 with, *A*, pilot point network, *B*, lithologic zones, and, *C*, base realization conductivity array.

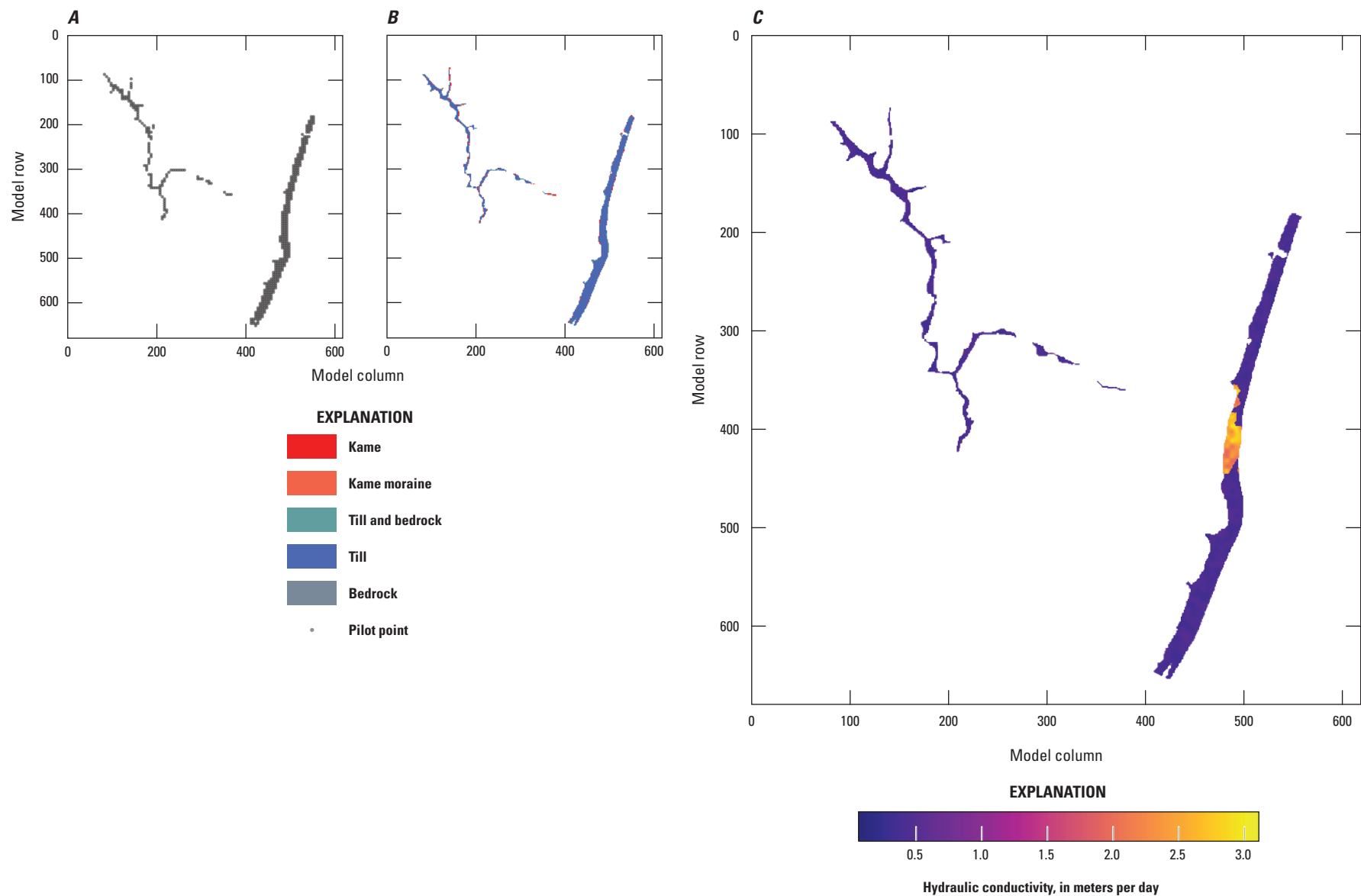


Figure 11. Base realization values and parameterization for vertical hydraulic conductivity in model layer 3 with, *A*, pilot point network, *B*, lithologic zones, and, *C*, base realization conductivity array.

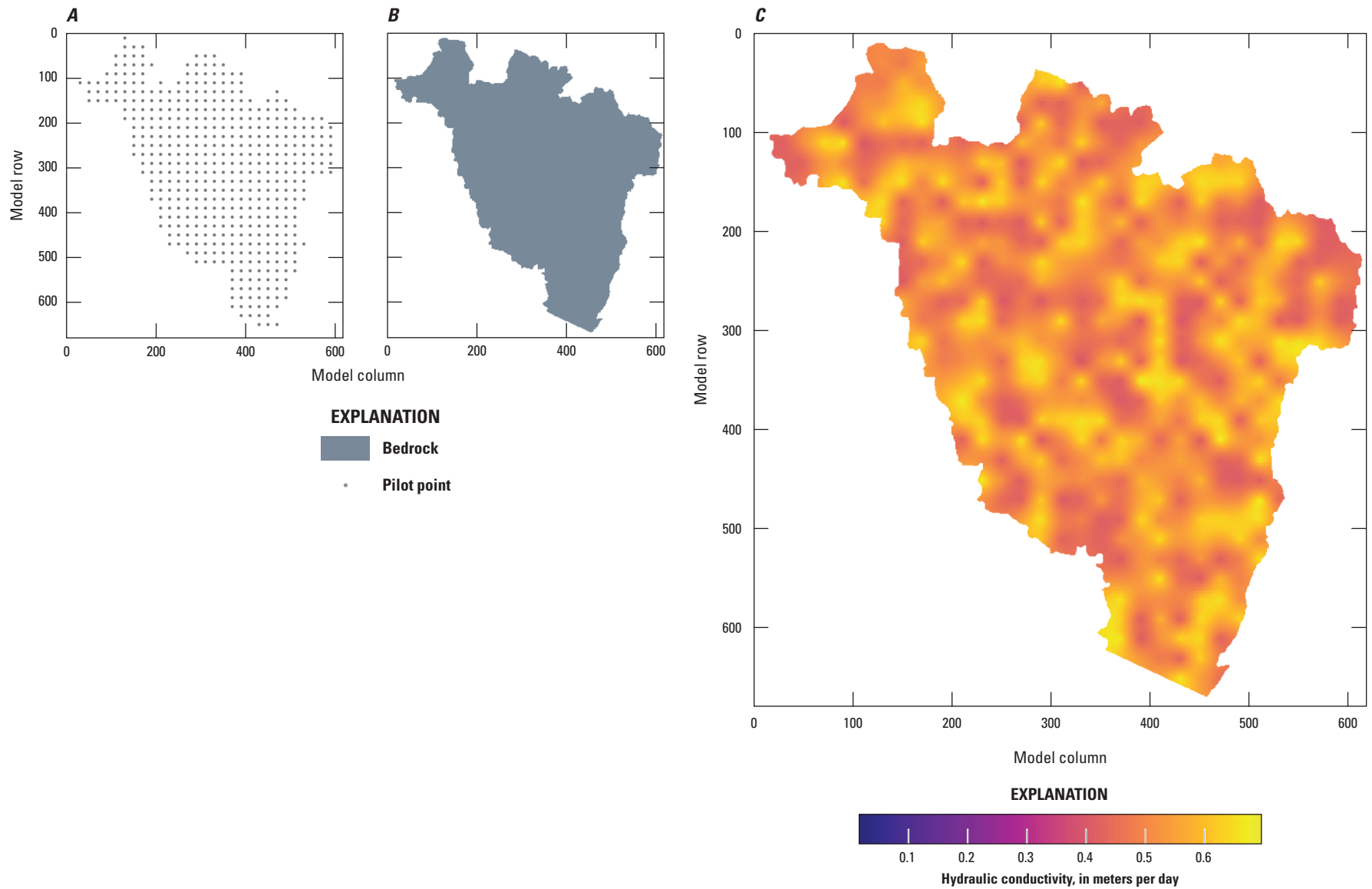


Figure 12. Base realization values and parameterization for vertical hydraulic conductivity in model layer 4 with, *A*, pilot point network, *B*, lithologic zones, and, *C*, base realization conductivity array.

Observation Data

An observation dataset was developed to compare model outputs with field measurements of the system to facilitate ensemble-based history matching. Historical water-level measurements were obtained from the National Water Information System (NWIS) and NYSDEC water well completion reports (New York State Department of Environmental Conservation, 2020). Only a single water-level measurement was available in the NWIS database (at site 414525074360601; U.S. Geological Survey, 2020a). Where present, multiple hydraulic head measurements were averaged across time to derive a single, steady-state value for a given location. In instances where water-level measurements were reported as depth to water below land surface, water-level elevations were computed by subtracting depth to water values from the model top elevation. This subtraction is accompanied by substantial uncertainty given the steep terrain and the resolution of the elevation data informing the model's top. In total, 448 hydraulic head targets were used for history matching. The single NWIS measurement includes more accurate elevation information than the NYSDEC measurements; however, the need to average over time causes enough uncertainty that the entire group of hydraulic head measurements was assumed to have the same uncertainty. This assumed hydraulic head measurement uncertainty was used to assign weights to hydraulic head observation for the history matching. During history matching, hydraulic head observations were compared against transmissivity-weighted hydraulic heads simulated by the model at the location of the observation well.

Steady-state base flows were computed from continuous records at two USGS streamgages 01436500 (Neversink River at Woodbourne, N.Y.), and 01366650 (Sandburg Creek at Ellenville, N.Y.). Long-term mean base-flow index (BFI; the average proportion of streamflow that is base flow) was computed using the USGS StreamStats web application (U.S. Geological Survey, 2020b). Mean annual flow rates were multiplied by the BFI values to compute an estimate of steady-state base flow (table 2). At streamgage 01366650, mean annual flow rates were computed using the entire period of record (October 1, 1957–September 29, 1977). At streamgage 01436500, only data collected after the construction of the Neversink Reservoir were used to compute mean annual flow rate (October 1, 1954–September 30, 1993). It was assumed that base flow computed over these periods is representative of long-term steady state conditions.

Land-surface elevations were also used as inequality targets on a grid every 50 model cells in each direction. We initially met a challenge common to this setting of hydraulic head solutions above the land-surface elevation (in other words, flooding) in early versions of the model, particularly before history matching. The locations of calibration targets are shown on figures 13–15.

We grouped the observations in three groups: flux, hydraulic head, and land surface. We assigned initial weights based on a general assumption of the observation quality

(table 4), which corresponded to a coefficient of variation of 10 percent for flux observations, and standard deviation of 5 and 10 m for the hydraulic head and land-surface observations, respectively. These assumptions are qualitative and result in an imbalanced objective function (the squared sum of residuals between model outputs and observations, each multiplied by their weight, and referred to as the 'PHI' value; fig. 16A). Subjectively, we judged that a reasonable balance of the objective function is 20 percent for hydraulic heads, 30 percent for flux, and 50 percent for land surface. This results in the balanced objective function in figure 16B with adjusted weights shown in table 4. The adjusted weights imply a reversal of standard deviation between hydraulic head and land-surface observations (11.2 and 2.4 m, respectively). Furthermore, the implied standard deviation for the two flux observations decreased dramatically because of this rebalancing; however, without adjusting the weights, the history matching algorithm would largely ignore the flux observations, allowing great variability and not constraining the overall water balance of the model. While many hydraulic head observations were available, they are individually subject to substantial uncertainty as discussed previously.

Stepwise Estimation of Model Parameters

Model parameters that are estimated in the history matching process are those for which we assume are informed by the observations. In choosing which parameters to allow to be adjustable through history matching, we acknowledge that any parameters not allowed to be estimated through history matching are assumed perfectly known. In this section, we outline the strategy for selecting the parameters to adjust and estimate in the history matching process. We then describe the approach for evaluating the uncertainty of parameters and observations to set initial parameter values and then the algorithm for history matching.

Parameterizing the Model

We followed the technique proposed by White and others (2020) to use multiple scales of variability in the parameterization. We typically used multipliers to allow adjustment at multiple scales—for example, single multipliers applied at the model layer scale combined with zone-wise multipliers applied to local areas (zones) within the model, in the case of parameters representing hydrogeologic properties (for example, hydraulic conductivity). Within zones, multipliers on pilot points (Doherty, 2003) were used to allow for finer-scale variability in hydraulic conductivity and recharge. Table 5 shows the categories of parameters and the number of adjustable parameters for each type. For horizontal and vertical hydraulic conductivity, we assigned zone-based homogeneous multipliers based on the mapped lithologic zones. Additionally, pilot points were placed every 20 cells in layers 1 and 4 and every 5 cells in layers 2 and 3. The higher frequency in

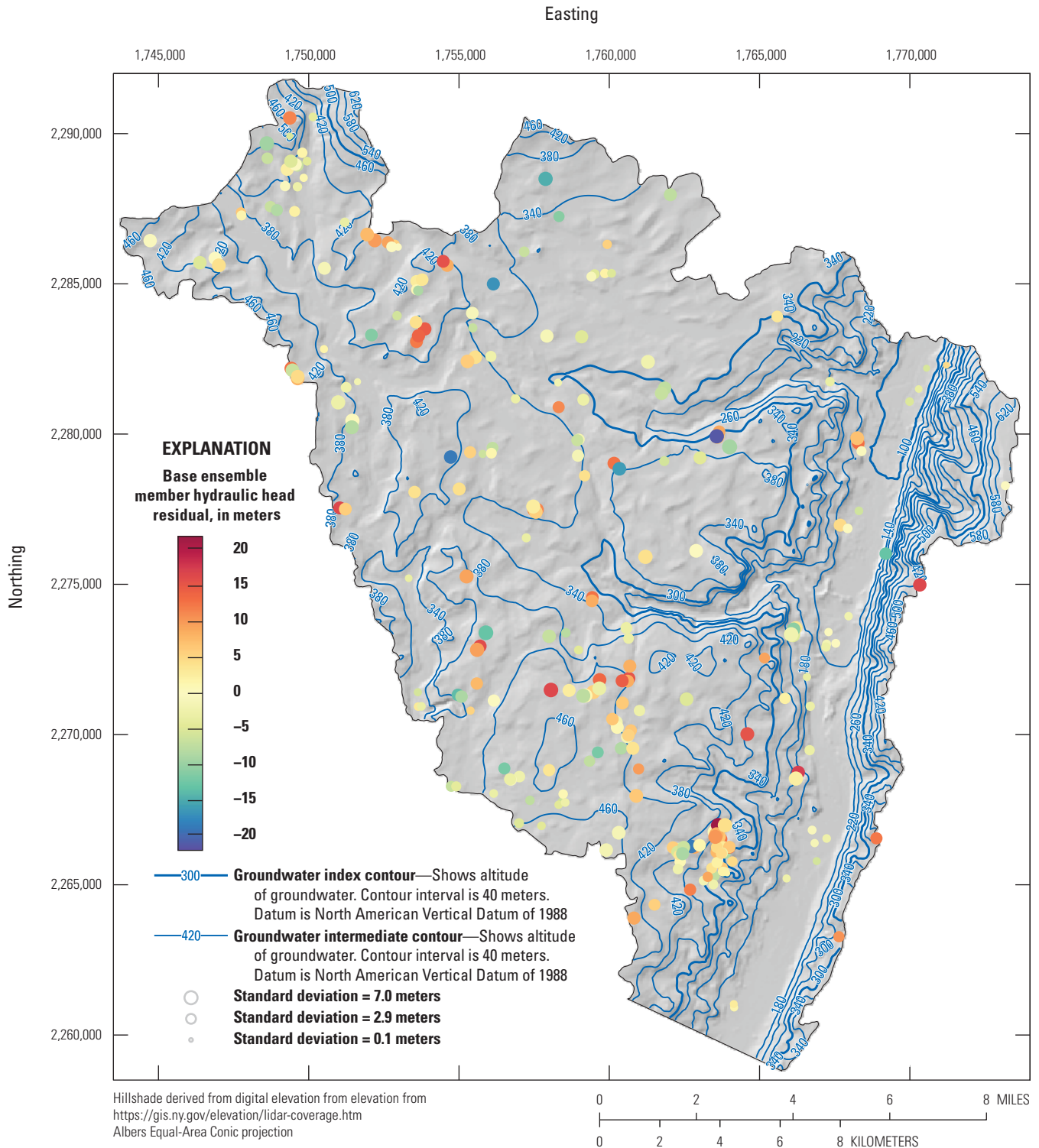


Figure 13. Spatial distribution of hydraulic head residuals (observed minus simulated) for the posterior parameter ensemble. Symbol colors are scaled to the base realization residual. Blue values indicate simulated values that are higher than observed. Symbol sizes are scaled to the standard deviation of the residual across the posterior parameter distribution; residuals are in meters. Contour lines show altitude of groundwater-table. Contour interval is 40 meters.

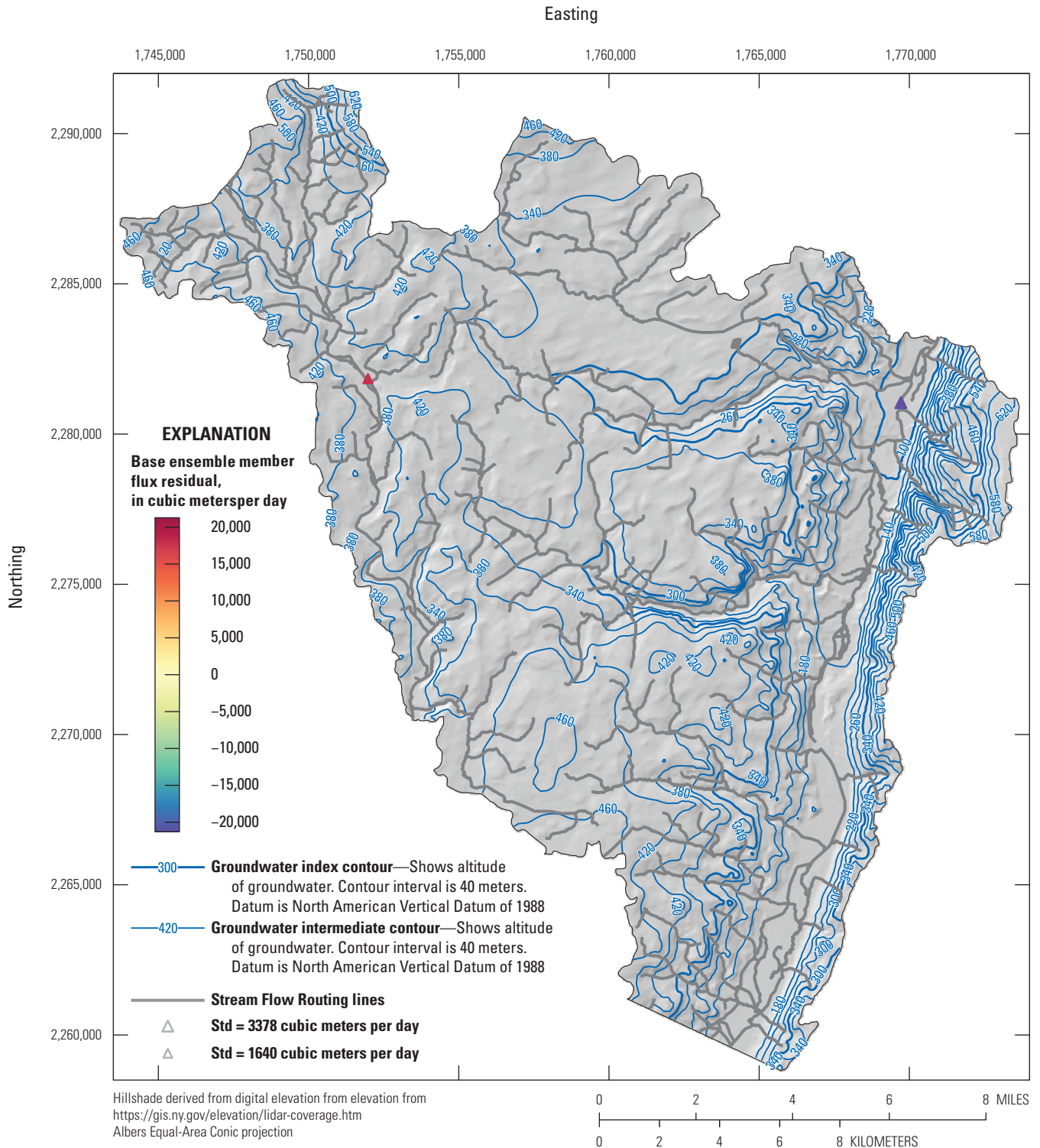


Figure 14. Spatial distribution of flux residuals (observed minus simulated) for the posterior parameter ensemble. Symbol colors are scaled to the base realization residual. Blue values indicate simulated values that are higher than observed. Symbol sizes are scaled to the standard deviation of the residual across the posterior parameter distribution; residuals are in cubic meters per day. Contour lines show altitude of groundwater-table. Contour interval is 40 meters.

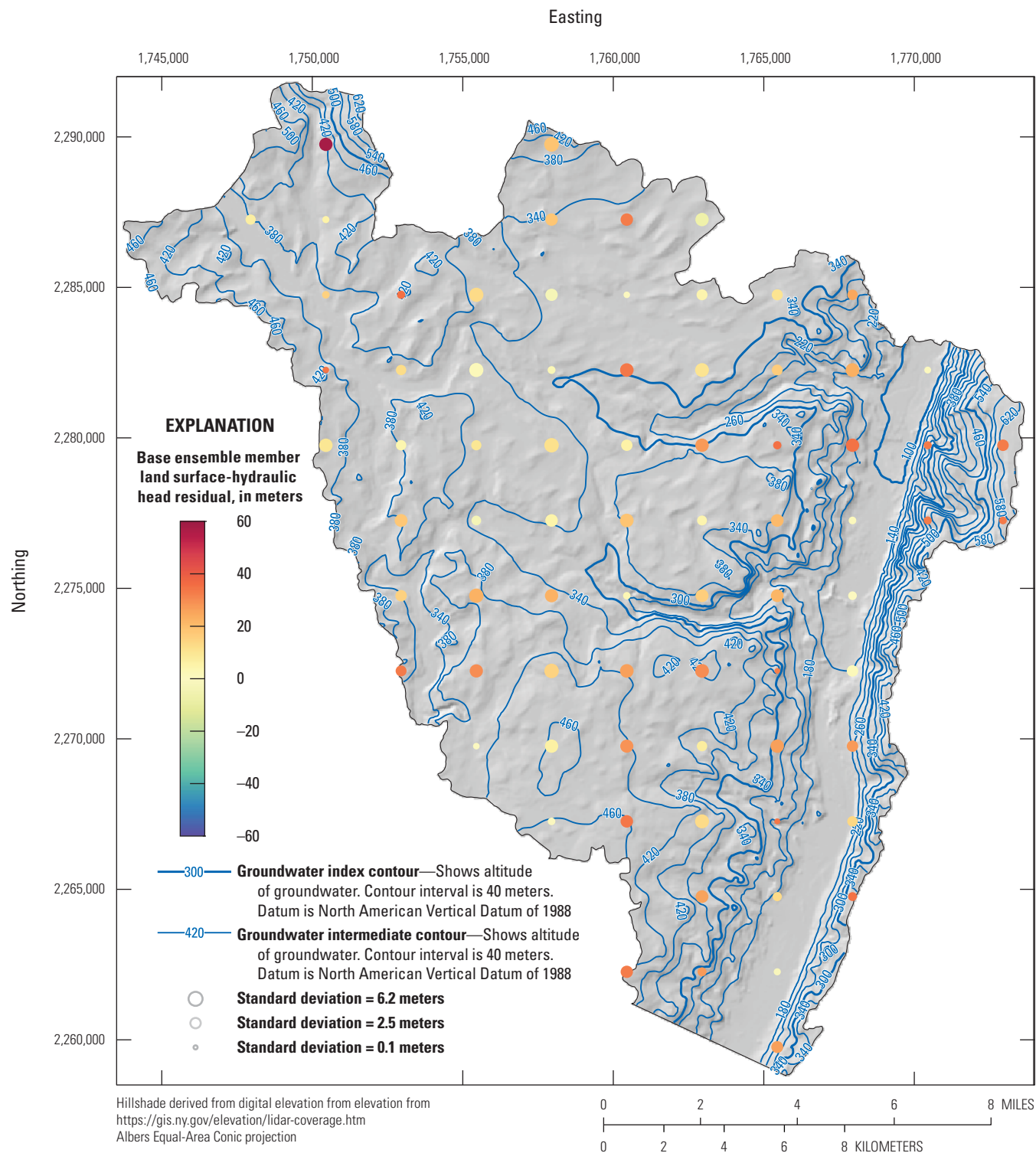


Figure 15. Spatial distribution of land surface inequalities (land-surface elevation minus simulated hydraulic head) for the posterior parameter ensemble. Symbol colors are scaled to the base realization residual. Red values indicate simulated hydraulic head elevations below land surface. Symbol sizes are scaled to the standard deviation of the residual across the posterior parameter distribution; residuals are in meters. Contour lines show altitude of groundwater-table contour interval is 40 meters.

layers 2 and 3 was to accommodate the irregular geometry of the geological zonation in those layers. The hydraulic conductivity zones and pilot point networks are displayed in panel “C” of figures 5–12. In a similar multiscale approach, we implemented a single recharge multiplier for the entire recharge array to allow for relatively large-scale adjustment, augmented by pilot points to allow for finer scale adjustment. The recharge pilot points were also placed every 20 model cells. The recharge pilot points are collocated with the layer 4 hydraulic conductivity pilot points as displayed in figure 17. The final parameters we implemented was a multiplier for the pumping rate in each well and a multiplier for the SFR streambed conductance by reach.

Using Uncertainty Analysis to Set Initial Values

An initial prior Monte Carlo analysis was completed on the parameterized model. A parameter ensemble was generated assuming a normal distribution of values and that the parameter bounds represent a spread of 4 standard deviations (in other words, 95 percent confidence) about the initial parameter values, as shown in table 5 in the section “Prior Monte Carlo.” The bounds were intentionally limited to keep parameters within initial ranges based on literature values. This was particularly problematic for the bedrock hydraulic conductivity values that assume deep, generally unweathered, and sparsely fractured bedrock when, in reality, the bedrock units include a shallow zone of greater weathering and fracturing that must be represented by the same unit.

The initial parameter bounds were evaluated by comparing model outputs from the ensemble of parameter samples in the prior Monte Carlo against observation data. We used an ensemble approach to sample the “noise” of the observations such that each realization uses a set of observation values sampled from a normal distribution with the mean being the reported observation value and the standard deviation being a user-defined variance representative of the uncertainty of the observations (table 4).

The left column in figure 18 shows histograms for several representative observations. The blue histograms represent model outputs from the ensemble of parameter samples in prior Monte Carlo. The orange histograms represent the

ensemble of observation values, with the vertical solid line representing the reported observation value, and the vertical dashed lines representing ± 1 standard deviation of the observation ensemble. While observations A and B show that flux is overestimated in one location and underestimated in another, observations C and D show that with the initial parameter assumptions, hydraulic heads are often too high; in the case of observation D, hydraulic head is 25 to 75 m higher than land surface. As a result, we expanded the bounds for hydraulic conductivity parameters by zone and reran the Monte Carlo analysis resulting in generally lower heads (the middle panels of fig. 18) although with relatively minor changes to the flux residuals. Based on this improvement, we used the parameter ensemble from the expanded Monte Carlo as both starting parameter values and upper and lower bounds for the history matching. The 25th, 50th, and 75th quantiles from the expanded Monte Carlo parameter ensemble were assigned as the lower bound, initial parameter value, and upper bound for history matching, respectively. These values are summarized in table 5 in the section “iES start.”

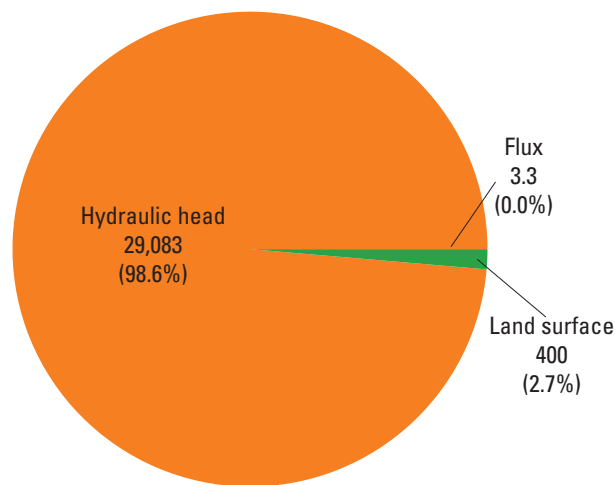
History Matching with the Iterative Ensemble Smoother (iES)

We selected the iES (White, 2018) as implemented in PEST++ (White and others, 2021) for history matching. This technique builds on the Monte Carlo approach by evaluating an ensemble of parameter values, yielding a residual vector and calculating a weighted objective function for each member of the parameter ensemble. From this information, the algorithm calculates an updated ensemble of parameters to evaluate, and iterations are done until a posterior conditioned ensemble is obtained. This results in an ensemble of objective function values from iteration to iteration (fig. 19). In figure 19, the red curve represents a base ensemble member that is taken as representative when a single realization is desired. The iES approach does not depend on local linearization of the inverse problem so inequality observations are possible. In this case, we defined water levels at the land surface on a regular network as inequality observations for the land-surface observation group. If the inequality is not invoked (in other words, if the model output hydraulic head elevation is lower

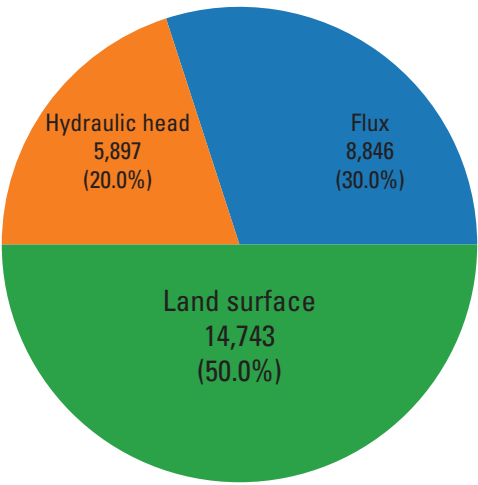
Table 4. Summary of observation data, observation groups, and initial and rebalanced weights.

Calibration target group	Values	Number of observations with nonzero weight	Initial weight			Rebalanced weight		
			Weight	Standard deviation	Percent error	Weight	Standard deviation	Percent error
Hydraulic head	59.3328 to 655.145	448	0.2	5	0.76319 to 8.42705	0.09006	11.10350	1.69481 to 18.7139
Flux	116217 to 124046	2	2.86532E-05 to 8.06151E-05	12404.6 to 34900.1	10 to 30.03	0.0014055 to 0.00395433	252.887 to 711.491	0.203865 to 0.612208
Land surface	88.72 to 673.96	78	0.1	10	1.48377 to 11.2714	0.606785	1.64803	0.244529 to 1.85756

A. Initial objective function components



B. Rebalanced objective function components



EXPLANATION

Land surface	Observation group name
14,743	Objective function amount
(50.0%)	Objective function percent

Figure 16. PHI distribution, A, before and, B, after rebalancing the objective function

than the collocated observation of land-surface elevation) then the weight assigned to this observation is 0.0 and thus the residual does not play a role in the objective function calculation. However, if the simulated hydraulic head exceeded these elevations, a quadratic penalty in the objective function was incurred at their assigned weights. Furthermore, the observation noise distributions allow flexibility in the algorithm to consider the uncertainty of the observations which, as discussed previously, is considered significant in this case where we have many observation values but none of them may be perfectly representative (for example, uncertainty in their location, timing versus the steady state model, and other aspects).

We selected the results of the second iES iteration as optimal based on a subjective judgement that the objective function had lowered to an acceptable level of fit while still retaining variability in the ensemble thus still capturing uncertainty in the system. As an additional step, we applied a rejection sampling process in which objective function values that are unacceptably high are trimmed from the ensemble. As shown in figure 20, a PHI cutoff was selected that removed anomalously high PHI results and results in an approximately normal distribution of PHI results. Figure 21 shows the correspondence between model outputs and observations, broken down by group. These results are shown to have little bias and low variance relative to the uncertainty of the observations. In figure 21, the blue symbols show fit for the representative base model while the bars indicate the 95-percent credible interval as calculated from the trimmed ensemble. Figure 18 shows the posterior ensemble of model outputs compared with the ensemble of observation values, the distribution of outputs and observations are consistent with their uncertainty. Generally, for the examples shown, the distributions of model outputs and observation uncertainty overlap with similar mean values.

Figures 13–15 show the spatial patterns of the residuals for the hydraulic head, flux, and land-surface observation groups, respectively. In each figure (13-15), the color of the symbol indicates the residual value (defined as observed value minus simulated value) for the base ensemble member, and the size of the symbol indicates the standard deviation of the residuals for all the ensemble members. In figure 14, one residual value is positive and one is negative. In figure 13, the residuals are positive and negative as well, highlighting that while there are some clusters of simulated values exceeding observed and vice versa, there is not a consistent spatial pattern that might indicate a structural issue with the model or the estimated parameters (for example, if all valleys were simulated as flooded). This mixed pattern of residuals is the desired outcome from history matching and indicates that the results are generally unbiased. The pattern is different for the land-surface residuals in figure 15. In this case, there are very few negative residuals and nearly all are biased such that the simulated results are lower value than the observed. Because these are inequality constraints and were intended to prevent parameters from driving the solution to simulate hydraulic heads higher than land surface, this bias is the desired outcome.

Table 5. Summaries of initial parameter values and bounds during the prior Monte Carlo, expanded Monte Carlo, and iterative ensemble smoother history matching, aggregated by parameter type.

[iES, iterative ensemble smoother; SFR; MODFLOW 6 Streamflow Routing package]

Parameter		Prior Monte Carlo			Expanded Monte Carlo			iES start		
Type	Count	Initial value	Lower bound	Upper bound	Initial value	Lower bound	Upper bound	Initial value	Lower bound	Upper bound
<i>Pilot point multipliers</i>										
Horizontal hydraulic conductivity: Layer 1	202	1	0.442759	1.55724	1	0.442759	1.55724	0.942248 to 1.05538	0.75596 to 0.866352	1.13672 to 1.31561
Horizontal hydraulic conductivity: Layer 2	560	1	0.442759	1.55724	1	0.442759	1.55724	0.934627 to 1.06573	0.741177 to 0.874849	1.1411 to 1.33064
Horizontal hydraulic conductivity: Layer 3	473	1	0.442759	1.55724	1	0.442759	1.55724	0.935518 to 1.07049	0.744879 to 0.870171	1.13199 to 1.33375
Horizontal hydraulic conductivity: Layer 4	470	1	0.442759	1.55724	1	0.442759	1.55724	0.928528 to 1.0651	0.754228 to 0.865706	1.15635 to 1.33452
Vertical hydraulic conductivity: Layer 1	202	1	0.442759	1.55724	1	0.442759	1.55724	0.935312 to 1.06418	0.759673 to 0.879614	1.15972 to 1.3271
Vertical hydraulic conductivity: Layer 2	560	1	0.442759	1.55724	1	0.442759	1.55724	0.93356 to 1.06665	0.746704 to 0.880208	1.13367 to 1.32458
Vertical hydraulic conductivity: Layer 3	473	1	0.442759	1.55724	1	0.442759	1.55724	0.899499 to 1.06398	0.750198 to 0.876427	1.13878 to 1.32987
Vertical hydraulic conductivity: Layer 4	470	1	0.442759	1.55724	1	0.442759	1.55724	0.934159 to 1.08109	0.749258 to 0.884273	1.14991 to 1.36895
Recharge: Highest active layer	470	1	0.8	1.2	1	0.8	1.2	0.978882 to 1.01934	0.912194 to 0.953566	1.04398 to 1.09877
<i>Multipliers by other attributes</i>										
Recharge for entire array	1	1	0.9	1.1	1	0.9	1.1	0.992711	0.953912	1.02459
SFR streambed conductance by reach	915	1	0.1	10	1	0.1	10	0.799918 to 1.29901	0.349099 to 0.637529	1.6714 to 2.77499
Pumping rate by well	31	1	0.8	1.2	1	0.8	1.2	0.987671 to 1.01061	0.91623 to 0.949433	1.03776 to 1.09147
Horizontal hydraulic conductivity by zone	42	1	0.181818 to 0.75	1.25 to 1.81818	0.266667 to 50	0.00266667 to 0.5	26.6667 to 5000	0.195096 to 65.7042	0.0380862 to 12.6817	1.03622 to 314.01
Vertical hydraulic conductivity by zone	42	1	0.181818 to 0.75	1.25 to 1.81818	0.266667 to 50	0.00266667 to 0.5	26.6667 to 5000	0.252877 to 67.3208	0.0611225 to 14.1789	1.48613 to 309.894

Base Realization

The final values of the “base” realization represent the central tendency of the posterior parameter distribution (White and others, 2021). This realization differs from the other realizations that are obtained by making a stochastic sample using the bounds of the initial parameter values. Instead, the base realization is initially made up of the specific starting values provided to PEST++. As a result, this realization starts at values that are considered “most likely” by the users. This realization is then subjected to adjustment during each iteration of the iES algorithm, but the central tendency typically remains, and this single realization typically is less variable than the other realizations. In this section, we present the representative base realization model properties, including hydraulic conductivities and boundary conditions that are representative of the optimal iES ensemble. We also discuss model results that were

simulated using base realization parameters, including the simulated water balance, water table, and streamflow. For the analysis of areas supplying recharge, however, the entire ensemble is used to be sure that rather than focusing on central tendency, we are exploring the range of outcomes consistent with prior knowledge of the parameters and the information from the observation data.

Hydraulic Conductivity

Base realization hydraulic conductivity arrays, zones, and pilot points are shown in [figures 5–12](#). Horizontal hydraulic conductivity arrays from model layers 1–4 are shown in [figures 5–8](#), and vertical hydraulic conductivity arrays from model layers 1–4 are shown in [figures 9–12](#). Base realization hydraulic conductivities, aggregated by geologic zones, are summarized in [table 3](#). Base realization horizontal hydraulic

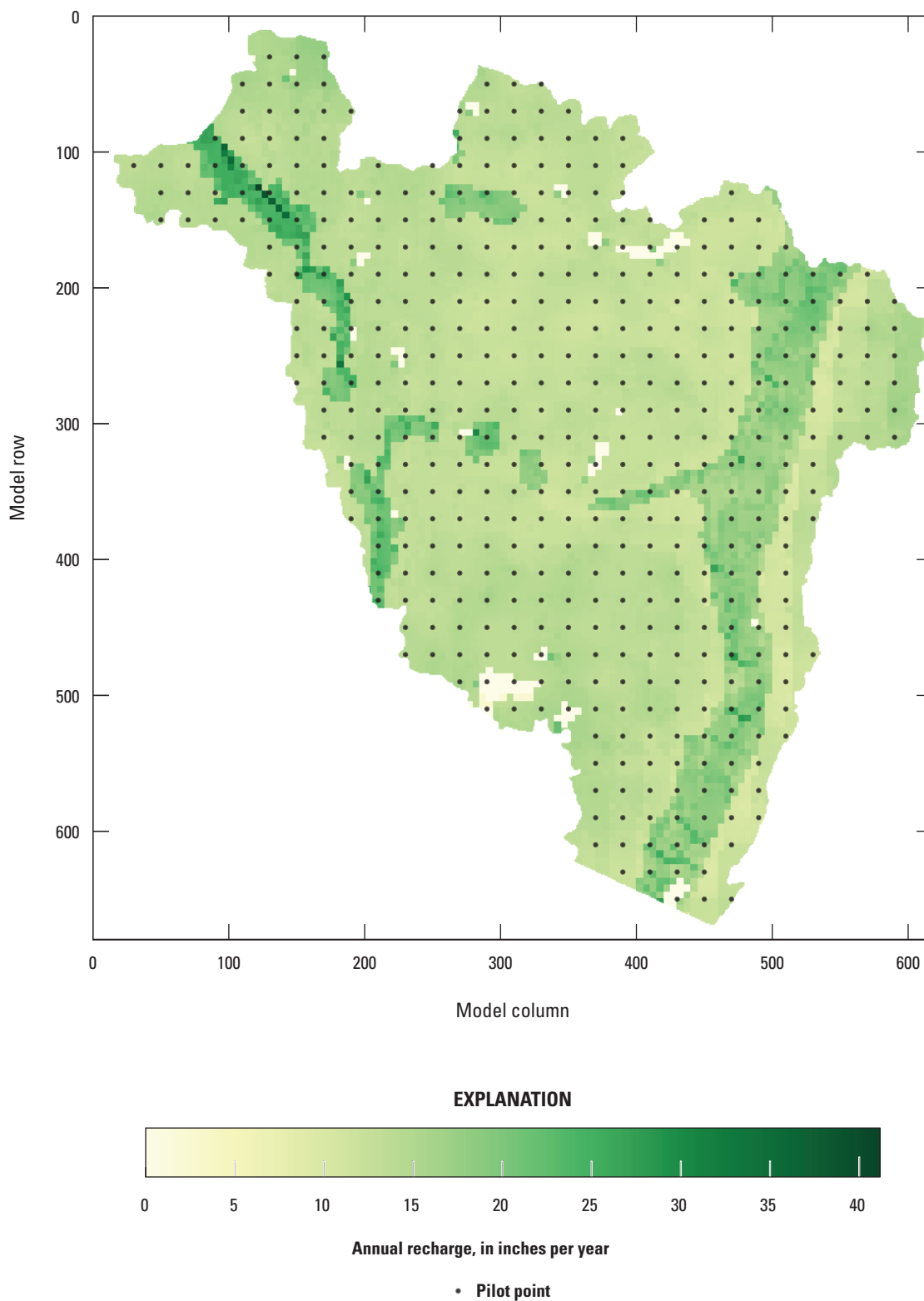


Figure 17. Estimated annual recharge to the groundwater system. Recharge rates shown are from the base realization, representing the central tendency of the calibrated posterior parameter distribution. The mean annual recharge rate across the study area is 14.1 inches per year.

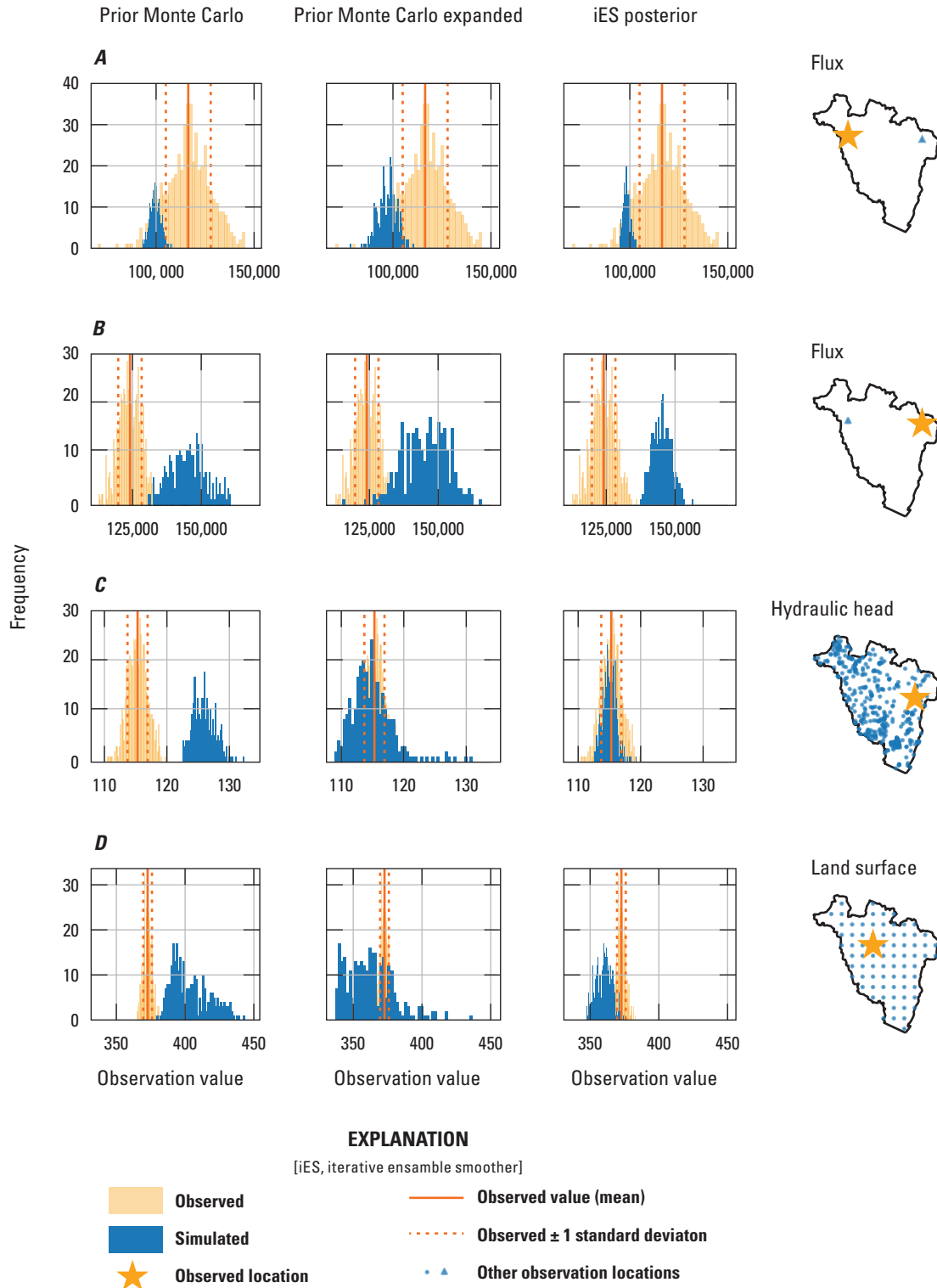


Figure 18. Histograms for selected observations showing model performance with respect to observed values for the “prior Monte Carlo” (left column), “expanded bounds Monte Carlo” (first column from left), and final “iES posterior” (first column from right) parameter ensembles. The location of the observation is shown in the right column. Ensemble results for, *A*, flux observation at U.S. Geological Survey (USGS) streamgage 01436500 (Neversink River at Woodbourne, N.Y.), *B*, flux observation at USGS streamgage 01366650 (Sandburg Creek at Ellenville, N.Y.), *C*, hydraulic head observation at well U 3717, and, *D*, land-surface inequality at model row 250, column 300. Blue histograms show the distribution of model outputs and orange histograms show realizations of the observation value based on the reported observation and a subjectively defined standard deviation.

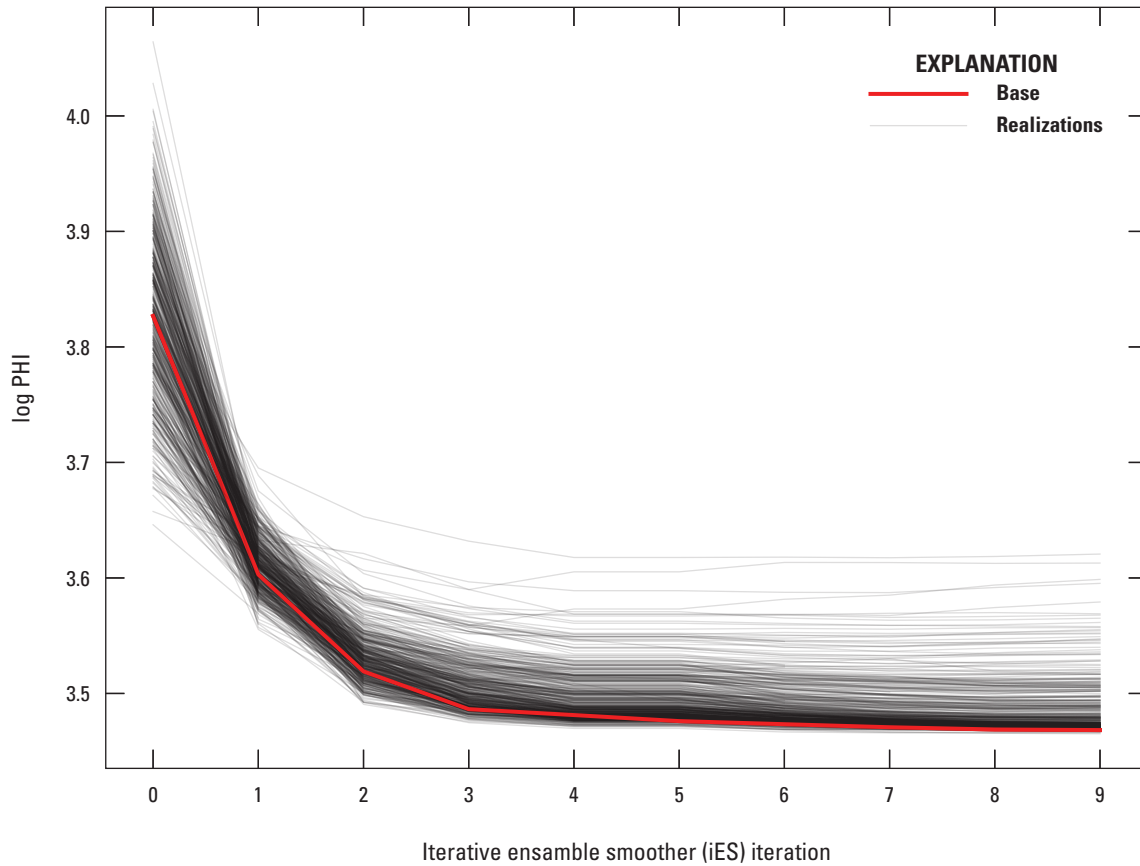


Figure 19. Summary of ensemble (black lines) and base (red line) objective function progress by iterative ensemble smoother (iES) iteration.

conductivity values were generally larger than vertical values, but in some areas of model layers 1 and 3, adjusted vertical hydraulic conductivities (K_v) were larger than horizontal conductivities (K_h). Prior K_v values were set equal to 10 percent of K_h values. This relation was not enforced during history matching, and K_v and K_h were adjusted independently. These areas were limited to relatively small areas in layer 1 (in the highly permeable [sand, outwash] and [or] generally unstructured [kame, till] deposits) where it is not unreasonable for K_v to be similar or slightly larger than K_h . An analysis of model anisotropy—the ratio of K_h to K_v —aggregated by lithologic zone is shown in figure 22. A simple sensitivity analysis was done to evaluate the relative importance of cells with anisotropy greater than 1 to overall model performance. A version of the base realization model was created where K_h was increased to equal K_v in all cells where anisotropy was greater than 1.0. This adjustment did not result in an appreciable change to model performance (less than 0.5 percent change in the objective function PHI), so the few cells with anisotropy greater than 1.0 were allowed to remain for the rest of the analysis.

Boundary Conditions

The base realization recharge array and parameterization is shown in figure 17. The base realization recharge resulted in slightly greater variability from Yager and others (2019) and had a slightly lower mean recharge rate of 14.1 in/yr across the study area. Well pumping rates were adjusted during calibration, but base realization pumping rates compare closely to reported values, with a mean difference of just 1.4 percent from initial pumping rates among all wells. Figure 23 shows the simulated base flow in the streams represented by the SFR package throughout the model. Negative values indicate flow into the streams from the groundwater-flow system (gaining streams), and positive values indicate flow from the streams into the groundwater-flow system from the streams (losing streams). Most of the model domain is characterized by gaining streams, although near the outlet of the Neversink Reservoir in the northwest corner of the model, in some of the steepest valley wall slopes, and at valley margins, losing streams are simulated. Some dry SFR reaches were simulated in upland areas and likely represent intermittent, low-order streams.

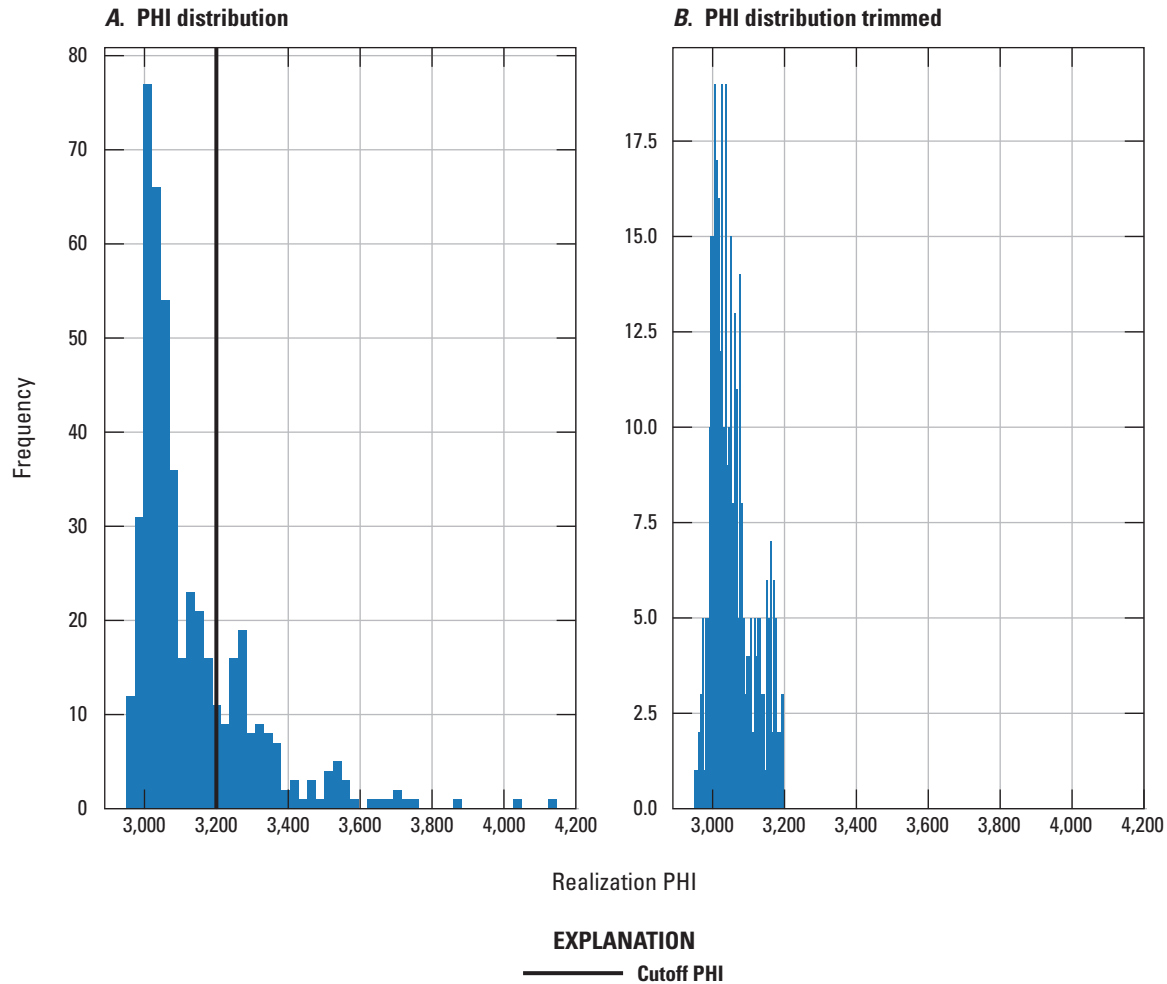


Figure 20. Comparison of posterior parameter ensembles, *A*, before and, *B*, after anomalously high PHI values were trimmed during rejection sampling.

Water Balance, Water Table, and Streamflow

The main components of the water balance, as simulated by the MODFLOW 6 model, are presented in [figure 24](#) for the base realization. As expected from the conceptual model of the system, the main inflow component is recharge and the main outflow component is discharge to streams represented by SFR. Smaller inflow components are recharge from streams represented by SFR and inflow at CHD boundaries. The smallest outflow component is the water removed through well pumping, followed by flow out through CHD boundaries. As required in the formulation of MODFLOW 6, the fluxes in and out of the model balance to zero.

The simulated water table, where hydrostatic pressure is equal to atmospheric pressure, is shown in [figure 25](#) as contours that depict a muted representation of the land-surface elevation, stylized in shaded gray. [Figure 25](#) also shows areas where the model is “flooded”—locations where the simulated water table is higher than the top of the model. Such simulated “flooding” is a common issue in valley-fill regions with

substantial areas of thinly covered bedrock. In reality, flooded areas may correspond to wetlands where groundwater is discharging at the land surface. In early versions of the model, literature-based parameter values, particularly for the bedrock, resulted in extensive simulated flooding. Furthermore, some areas with inadequate SFR representation resulted in the model’s inability to convey water from low-lying areas into streams. With the refined stream network and parameters informed through history matching, the extent of flooding is limited to areas deep in the valleys and some highlands that are generally coincident with wetlands and ponds or lakes. Routing of water simulated above land surface could be implemented through the MODFLOW 6 unsaturated zone flow (UZf) package (Feinstein and others, 2020); however, we judged the extent of flooding to be sufficiently limited that the additional package, accompanied by more computational complexity, was not necessary.

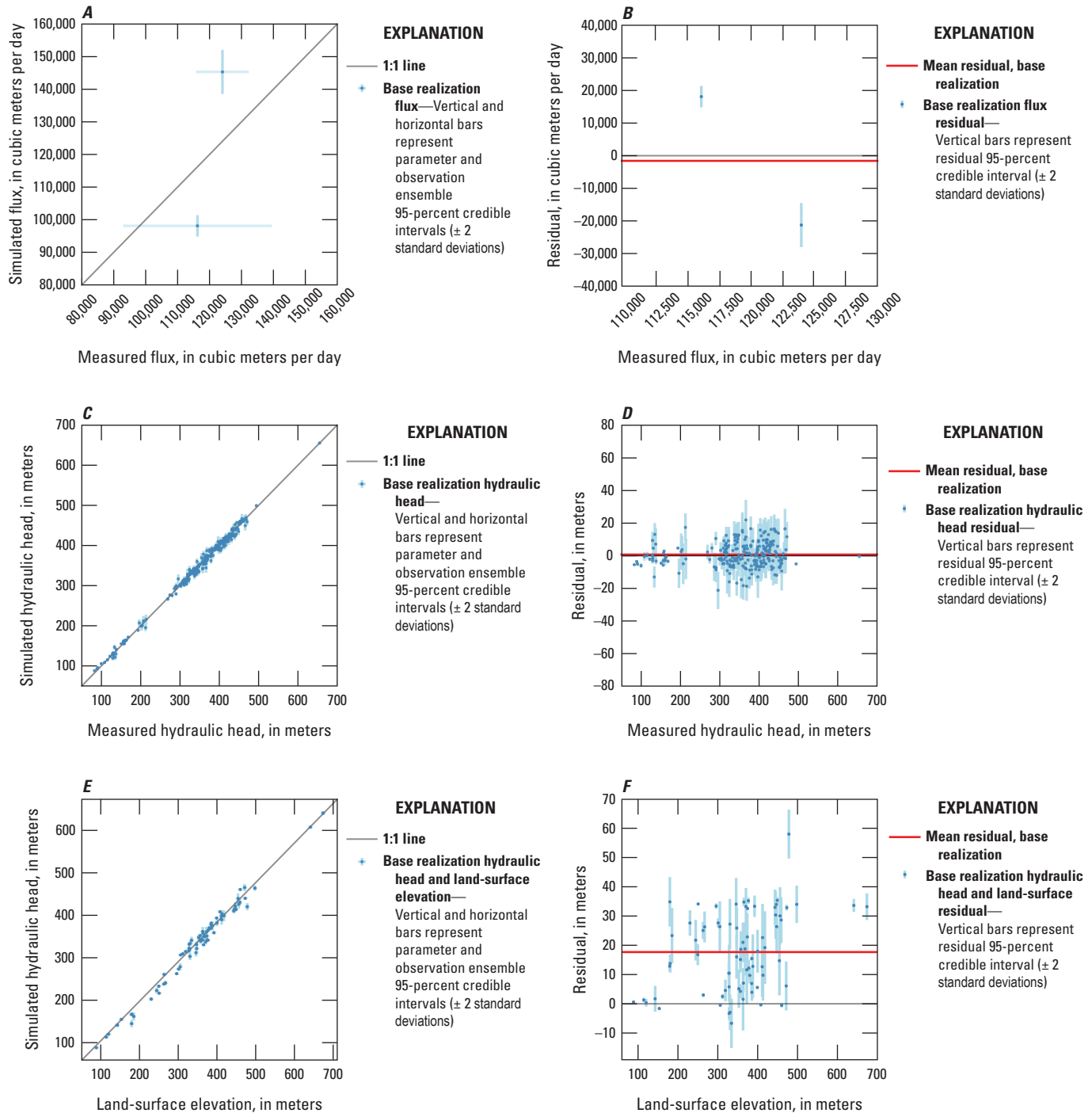


Figure 21. Relations among observed and simulated flux, hydraulic head, and land-surface inequalities, and plots of residuals compared to observed values. Points represent base realization values (the central tendency of the posterior parameter distribution), and the bars represent the 5th–95th percentile ranges of the observation and posterior parameter distributions. Simulated and observed values of, *A*, flux observations, *C*, hydraulic head observations, and, *E*, land surface inequalities. Residuals and observed values of, *B*, flux observations, *D*, hydraulic head observations, and, *F*, land-surface inequalities.

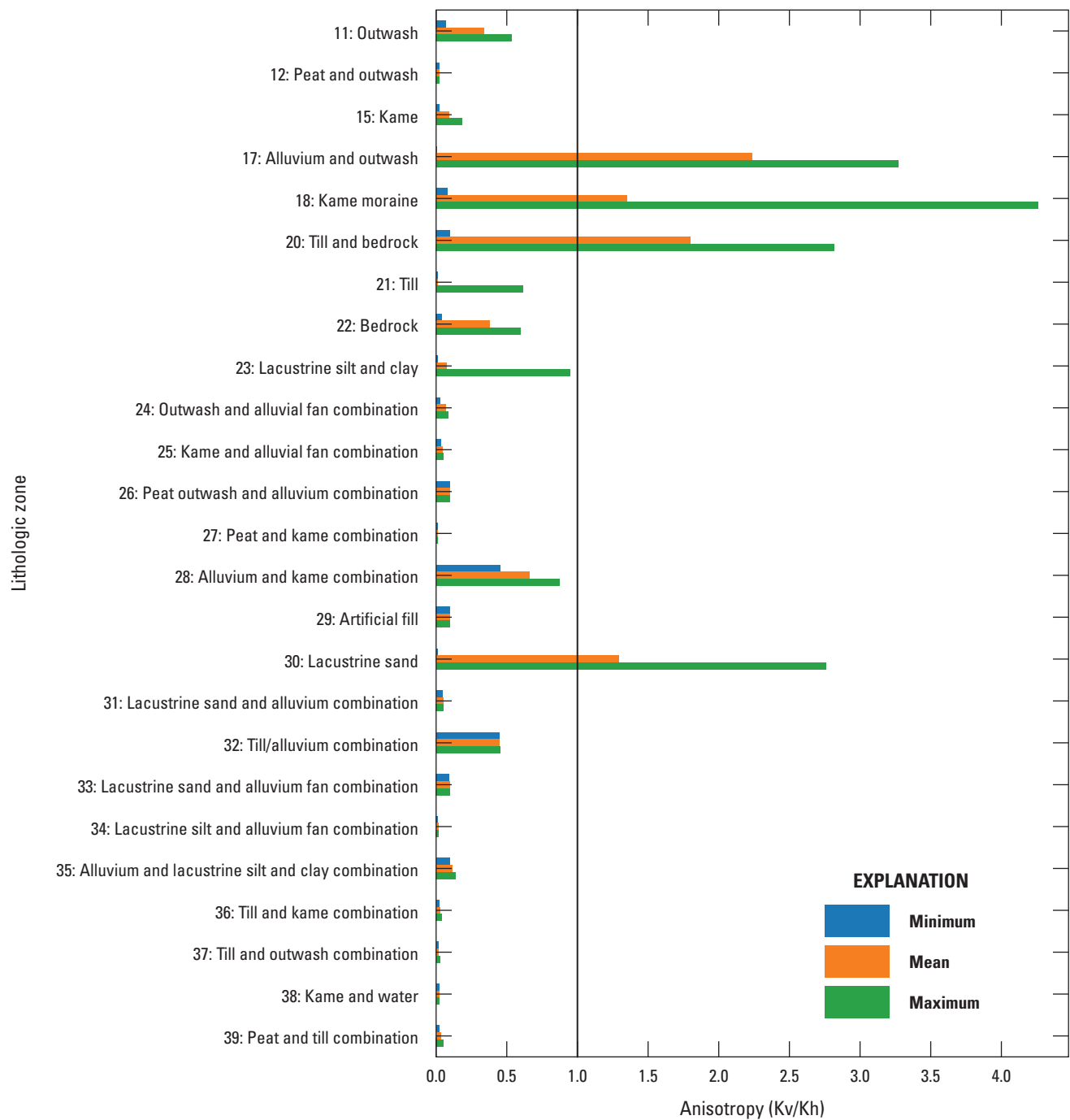


Figure 22. Summary of base realization anisotropy by lithologic zone—the ratio of horizontal (Kh) to vertical hydraulic (Kv) conductivity—in the groundwater-flow model. Anisotropy values less than 1 indicate Kh is greater than Kv. Anisotropy values greater than 1 indicate Kh is less than Kv.

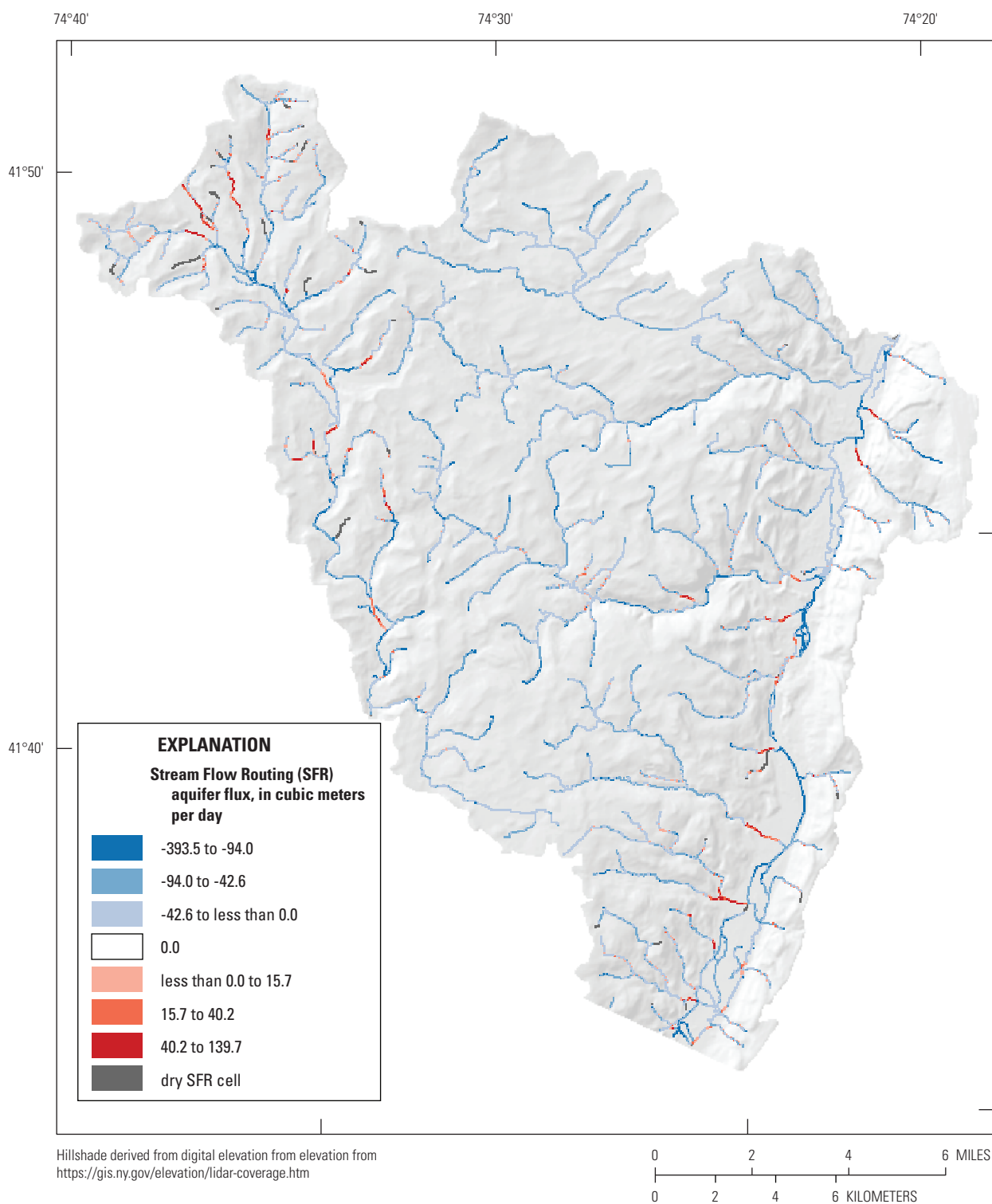


Figure 23. Base realization aquifer flux in Streamflow Routing (SFR) cells. Negative (blue) values are gaining (aquifer discharge to SFR) and positive (red) values are losing (SFR discharge to aquifer). White cells have no aquifer flux and gray cells are dry.

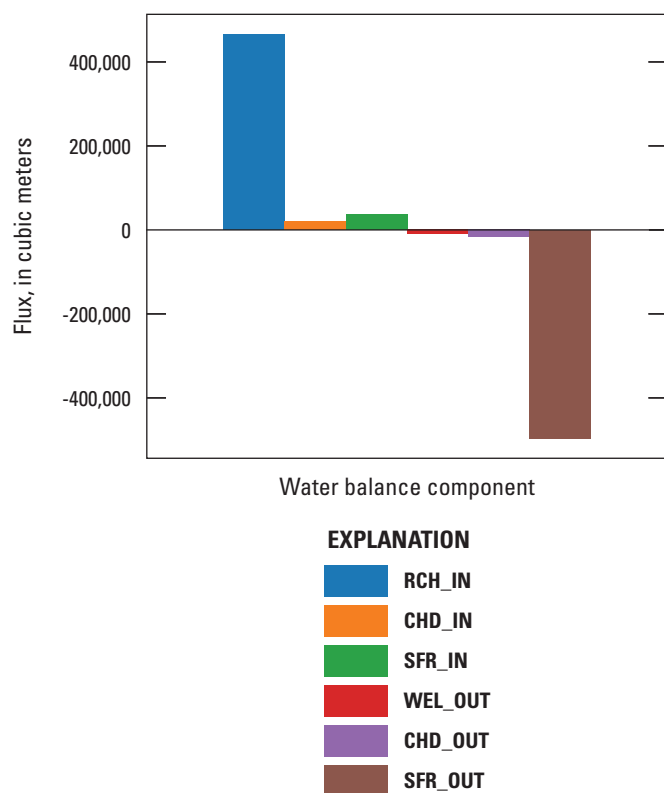


Figure 24. Base realization groundwater model mass balance.

Providing an Ensemble for Monte Carlo Analysis of Areas Contributing Recharge

The end result of the history matching effort was an ensemble of parameter values for the groundwater-flow model that spans the range of values within prior uncertainty bounds and that also is informed by the historical observation data, within a reasonable range of uncertainty on those observations. We carried this ensemble through to the simulations to delineate the areas contributing recharge by running the groundwater-flow and particle tracking models for each ensemble member, as we discuss in subsequent sections.

Simulation of Areas Contributing Recharge and Prediction Uncertainty Analysis

We completed a Monte Carlo analysis to simulate the flow field using the posterior history-matched ensemble of groundwater-flow model parameters to determine the probability of a recharge location being in the contributing area of the priority wells. This section describes the methods used for delineating priority well recharge areas and of the uncertainty analysis of these predictions.

Probabilistic Areas Contributing Recharge

The areas contributing recharge to the priority production wells were delineated using the groundwater-flow model and a particle-tracking simulation. Parameter sets produced by the history matching process were used to run MODFLOW. The unique flow field generated by each parameter realization and MODFLOW were then used to build MODPATH7 (Pollock, 2016) particle tracking input files. For the MODPATH7 simulations, one particle was placed in each cell in the uppermost active layer (in the same way that recharge is applied to the model) at the beginning of the simulation. Particle tracking was then run in the forward direction. The movement of the particles simulate the movement of recharge through the aquifer system, moving downgradient toward sinks (such as wells) and other model boundaries. MODPATH7 tracks the movement of the particles during the simulation and records where they terminate at model sinks. In this study, we instructed particles to stop when they met weak sink cells, which was necessary because some of the priority wells in the model are weak sinks (see Abrams and others, 2013 for additional information about weak sinks). At the conclusion of the particle-tracking simulation, the particles that terminated in priority well sinks were traced back to their starting locations and mapped to the model grid. This produced a single, deterministic map of the model cells that do (and do not) contribute recharge to priority wells.

Prediction Uncertainty Analysis

Results of the history-matching process produced an ensemble of 500 groundwater model parameter realizations that account for both model parameter and calibration data uncertainty. Of these 500 parameter realizations, 142 were trimmed from the ensemble during rejection sampling analysis to remove anomalously high PHI values (fig. 20) and thereby parameter sets that result in unrealistic groundwater levels and streamflows. The remaining 358 parameter realizations were used to complete a Monte Carlo analysis to quantify the uncertainty of the contributing recharge area to priority water-supply wells. The MODPATH particle tracking process, described in the “Probabilistic Areas Contributing Recharge” section, was repeated for each of the 358 realizations in the history-matched parameter ensemble. Deterministic contributing recharge areas produced from these repeated runs were used to calculate the probability of a location being in the area contributing recharge to a priority well.

In the history matching part of the project, best estimates of mean water use were used for the well package in MODFLOW 6; however, these estimates are typically not at the highest range of capacity for the wells, in particular for the priority wells (table 1). As a result, variations in pumping rates related to population growth and other future decisions were added as an additional consideration beyond model parameters

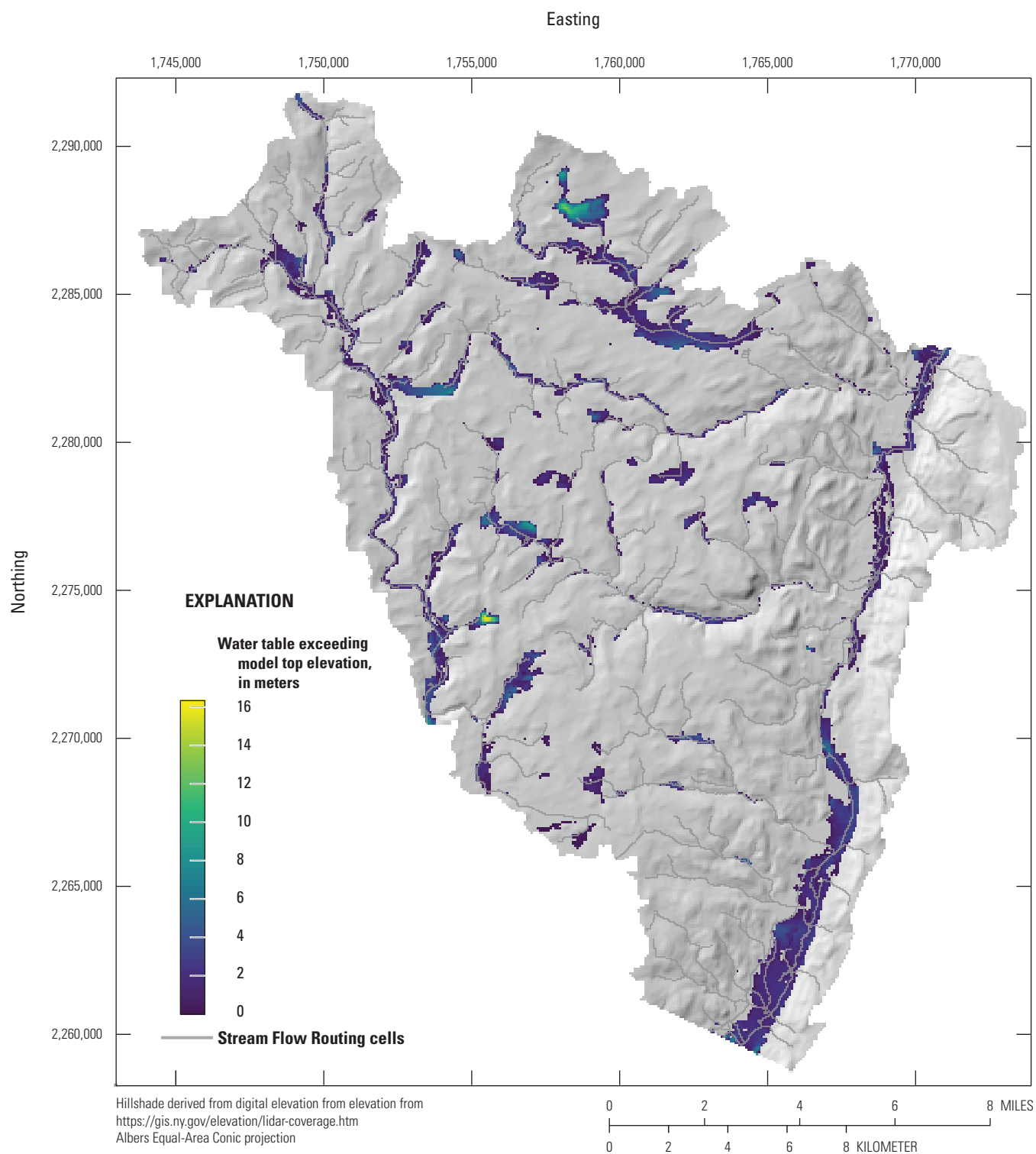


Figure 25. Base realization and simulated flooding.

that contribute to the overall uncertainty of the contributing recharge area. The uncertainty in pumping rates was represented using six pumping scenarios:

- Baseratio: priority well pumping rates only adjusted by history matching;
- Maxratio: priority well pumping rates scaled by the ratio of the maximum capacity to their initial values;
- 0.25ratio: priority well pumping rates scaled by a 25 percent increase (25 percent of the range between base and maximum capacity rates);
- 0.5ratio: priority well pumping rates scaled by a 50 percent increase (50 percent of the range between base and maximum capacity rates);
- 0.75ratio: priority well pumping rates scaled by a 75 percent increase (75 percent of the range between base and maximum capacity rates); and
- Maxrange: priority well pumping rates allowed to vary in a uniform distribution between the base and maximum capacity rates.

The Monte Carlo analysis (358 simulations) was repeated for each of the 6 pumping scenarios (for a total of 2,148 simulations) to incorporate the uncertainty related to the effects of pumping on the areas contributing recharge to priority wells. The results of these analyses are shown in [figures 26–31](#). The general shapes of the contributing recharge areas were consistent among pumping scenarios, with increasing pumping rates leading to larger contributing recharge areas and larger areas of high probability that a location contributes recharge to priority wells; for example, the footprints simulated during the “Maxratio” scenario ([fig. 27](#)) cover larger regions and have larger areas of high probability than those simulated during the “Baseratio” scenario ([fig. 26](#)).

Panel *B* of [figures 26–31](#) shows the probabilistic contributing recharge area to the 4 priority wells near Ellenville, N.Y. The Ellenville contributing recharge area is the largest in size for all scenarios, owing the number (4) and density of wells and their combined maximum withdrawal rates ([table 1](#)). The Ellenville footprint extends to the south, to near Spring Glen, N.Y., and the upper Sandburg Creek drainage ([fig. 1](#)). The footprint also extends laterally to the east and west up the valley walls into the uplands, particularly to the east of Ellenville between the tributary streams draining Shawangunk Ridge. The highest probability areas of the Ellenville contributing recharge area were typically simulated along the eastern part of the valley floor and the eastern uplands, with some high probability areas also simulated in the western uplands.

Panel *C* of [figures 26–31](#) shows the probabilistic contributing recharge areas to the 3 priority wells near Woodridge (2) and Mountain Dale (1), N.Y. Contributing recharge areas near the two western wells near Woodridge cover parts of the valley floor near the wells and, to a larger extent, areas in the uplands to the north, east, and west of the wells. The contributing

recharge area for the eastern well, pumping from a confined aquifer unit near Mountain Dale, includes an area south of the well in the uplands and a small part of the valley floor near the well. During intermediate and higher pumping scenarios, additional contributing areas are simulated in the uplands to the north and south of the well ([fig. 27](#), for example); however, these areas have relatively low probability of contributing recharge.

Panel *D* of [figures 26–31](#) shows the probabilistic contributing recharge area to the priority well near Wurtsboro, N.Y. The contributing recharge area for this well includes a part of the valley floor near the well and the uplands to the west of Wurtsboro. The extent of this well’s contributing recharge footprint does not change greatly among pumping scenarios, but results indicate that increased pumping at this well leads to increased probability of recharge contribution.

Each pumping scenario represents an outcome that, when used for planning purposes, corresponds to a different level of risk tolerance; for example, pumping rates in the “Maxratio” scenario are scaled to maximum capacities, creating the largest contributing recharge footprint and large areas of high probability near wells. Management decisions made using results from this scenario ([fig. 27](#)) would be relatively risk averse. Alternatively, pumping rates in the “Baseratio” scenario are adjusted only by history matching and remain near 2010–2015 reported mean levels, creating contributing recharge footprints that are relatively small and smaller high probability areas. Management decisions made using “Baseratio” results ([fig. 26](#)) would be relatively risk tolerant. Management decisions made using the intermediate scaled and ranged pumping scenarios (“0.25ratio,” “0.5ratio,” “0.75ratio,” and “Maxrange”) therefore represent intermediate levels of risk tolerance. Taken together, these results offer a view of the overall uncertainty of the areas contributing recharge to priority wells in the NS–RO study area and provide a tool for risk-based decision making related to the source water protection of these wells.

Assumptions and Limitations of Analysis

The NS–RO groundwater-flow model is a simplification of an infinitely complex natural system. The model integrates large amounts of information about the groundwater system (including aquifer and stream-network geometries, subsurface hydrologic properties, and measurements of water levels and flow), but there are inevitably areas of the model where information is sparse. Considerable effort was put into quantifying model prediction uncertainty stemming from the uncertain model parameters and observation data, but it is more difficult to quantify uncertainty related to “structural error” (Doherty and Welter, 2010). Structural error can be caused by oversimplification or insufficient information and can limit the model’s ability to accurately represent the natural system; for

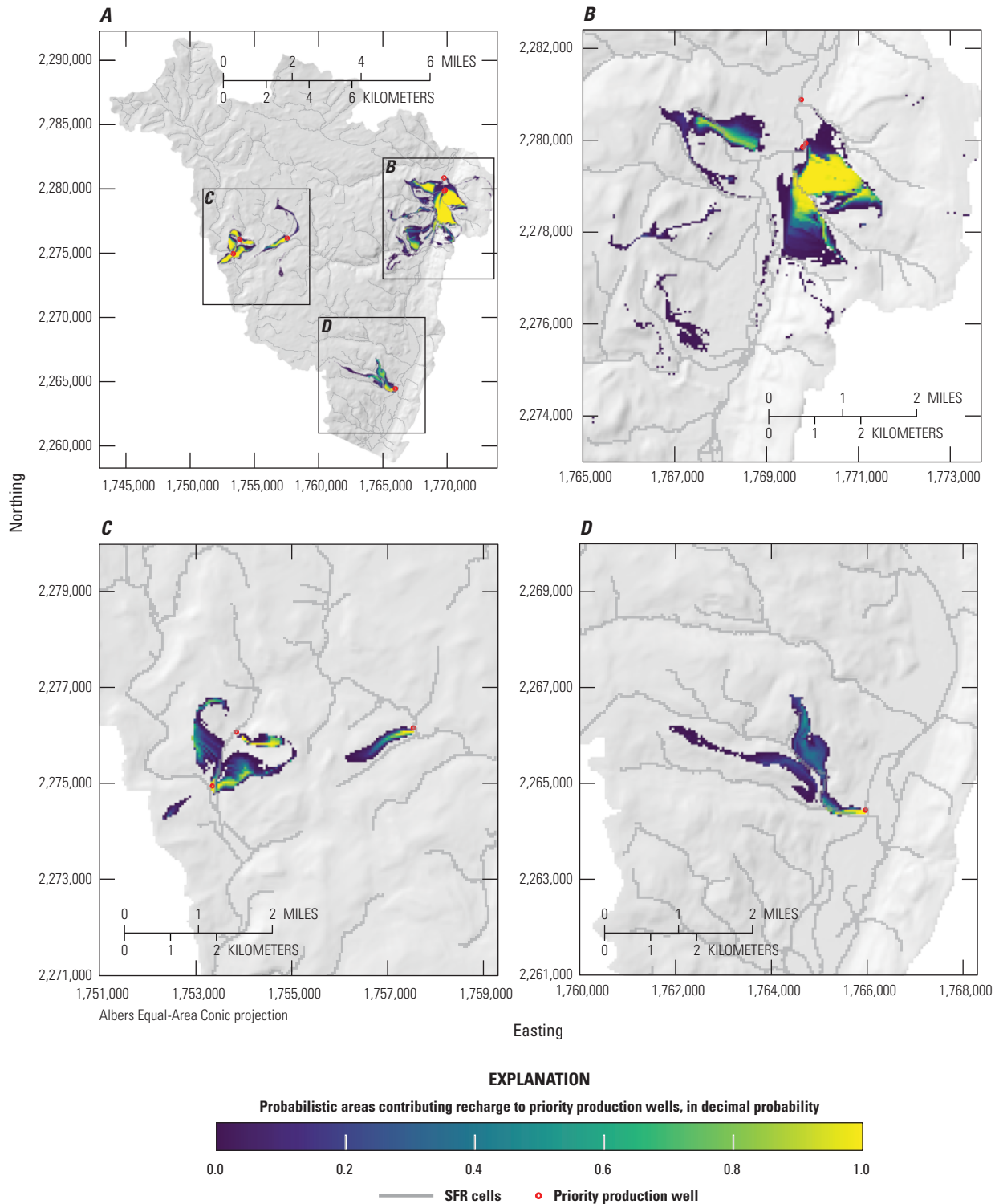


Figure 26. Probabilistic areas contributing recharge to priority wells under the “Baseratio” pumping scenario. Panels show, *A*, full study area results, *B*, detail view of priority wells near Ellenville, *C*, detail view of priority wells near Woodridge and Mountain Dale, and, *D*, detail view of priority well near Wurtsboro. Colormap is scaled to the probability that a location contributes recharge to priority wells (red circles). A location with a value of 1.0 is within the area contributing recharge to a priority well in 100 percent of simulations. A location with a value of 0.5 is within the area contributing recharge to a priority well in 50 percent of simulations. A location outside of the color flood is within the area contributing recharge to a priority well in 0 percent of simulations.

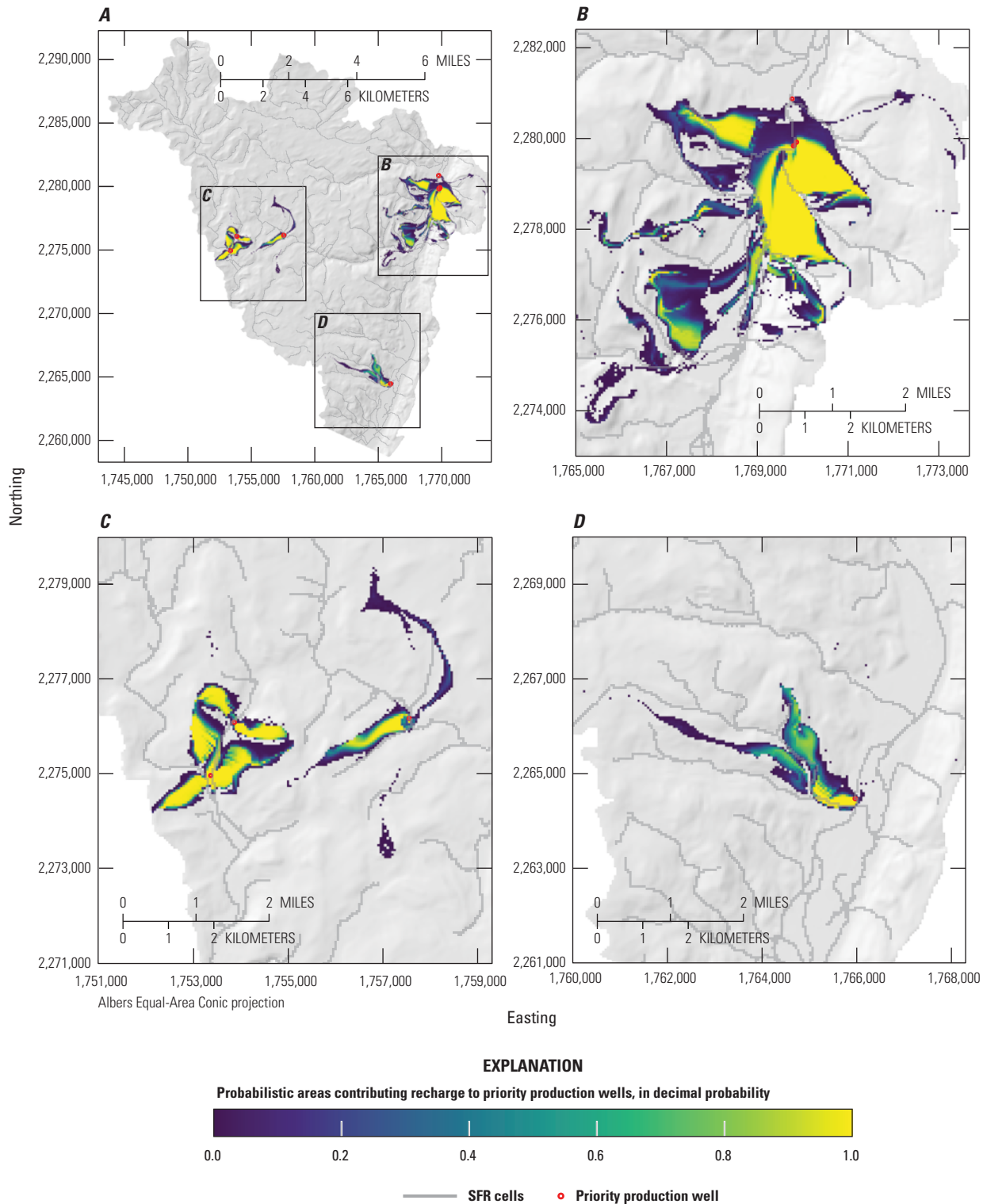


Figure 27. Probabilistic areas contributing recharge to priority wells under the “Maxratio” pumping scenario. Panels show, *A*, full study area results, *B*, detail view of priority wells near Ellenville, *C*, detail view of priority wells near Woodridge and Mountain Dale, and, *D*, detail view of priority well near Wurtsboro. Colormap is scaled to the probability that a location contributes recharge to priority wells (red circles). A location with a value of 1.0 is within the area contributing recharge to a priority well in 100 percent of simulations. A location with a value of 0.5 is within the area contributing recharge to a priority well in 50 percent of simulations. A location outside of the color flood is within the area contributing recharge to a priority well in 0 percent of simulations.

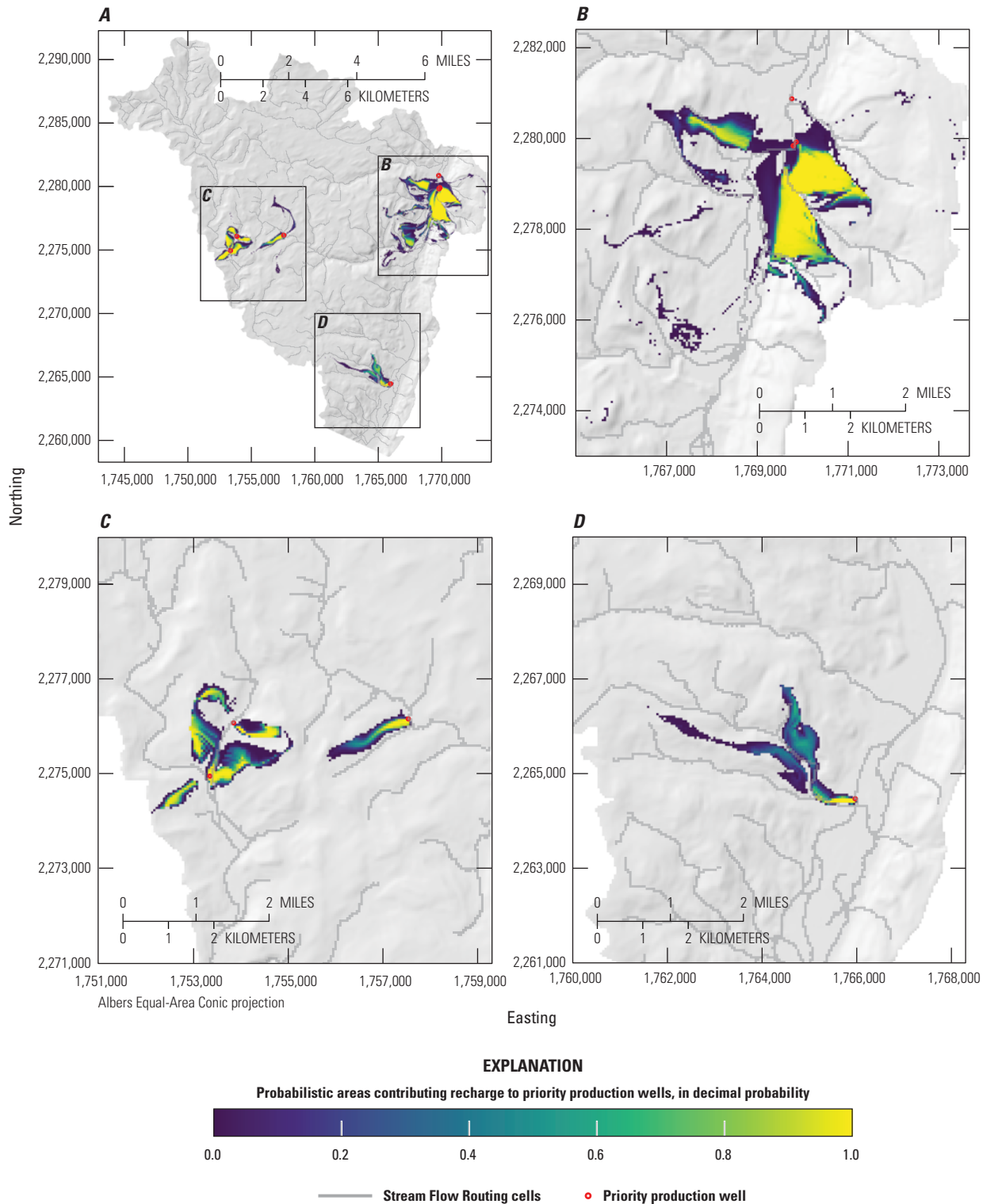


Figure 28. Probabilistic areas contributing recharge to priority wells under the “0.25 ratio” pumping scenario. Panels show, *A*, full study area results, *B*, detail view of priority wells near Ellenville, *C*, detail view of priority wells near Woodridge and Mountain Dale, and, *D*, detail view of priority well near Wurtsboro. Colormap is scaled to the probability that a location contributes recharge to priority wells (red circles). A location with a value of 1.0 is within the area contributing recharge to a priority well in 100 percent of simulations. A location with a value of 0.5 is within the area contributing recharge to a priority well in 50 percent of simulations. A location outside of the color flood is within the area contributing recharge to a priority well in 0 percent of simulations.

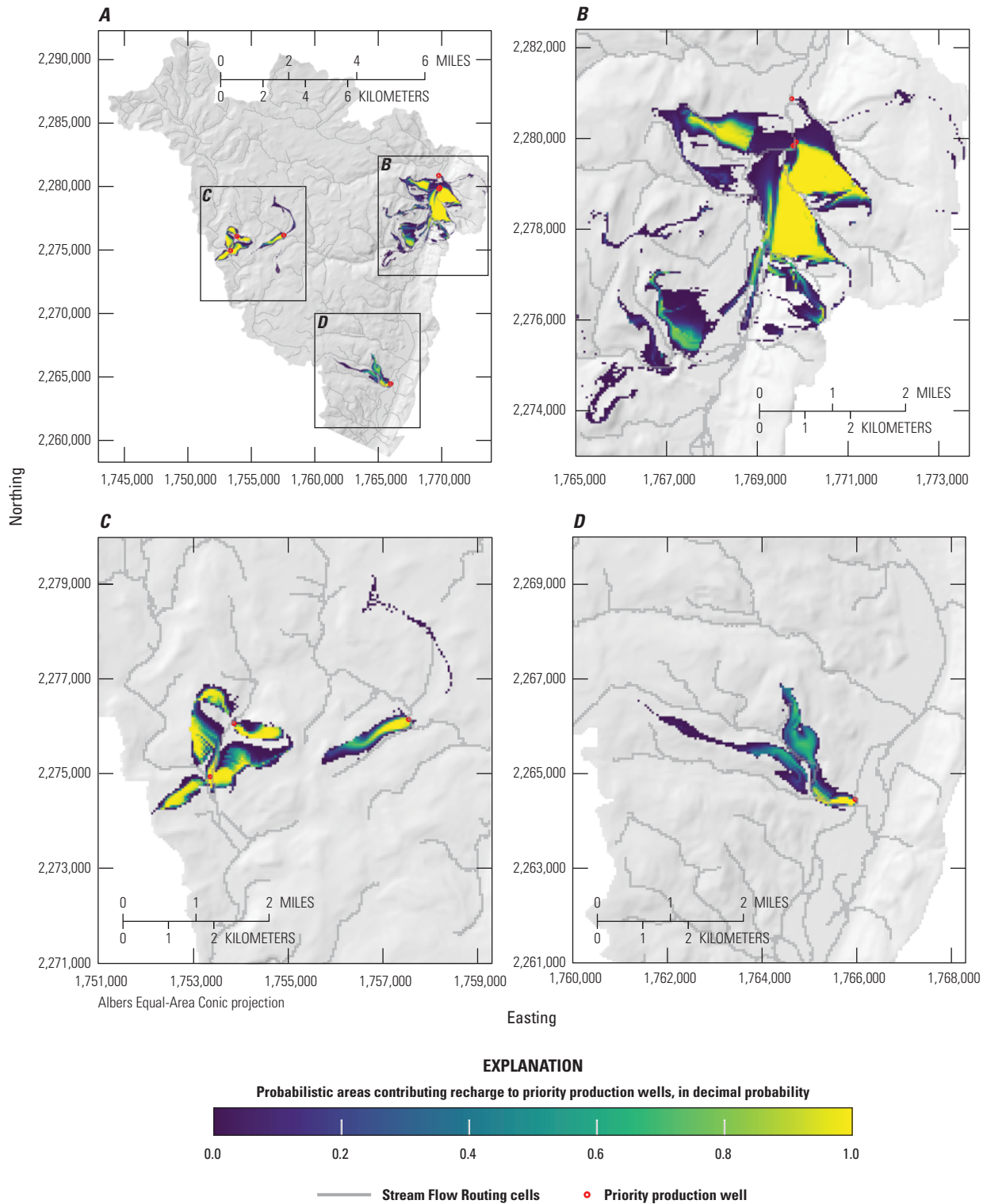


Figure 29. Probabilistic areas contributing recharge to priority wells under the “0.5 ratio” pumping scenario. Panels show, *A*, full study area results, *B*, detail view of priority wells near Ellenville, *C*, detail view of priority wells near Woodridge and Mountain Dale, and, *D*, detail view of priority well near Wurtsboro. Colormap is scaled to the probability that a location contributes recharge to priority wells (red circles). A location with a value of 1.0 is within the area contributing recharge to a priority well in 100 percent of simulations. A location with a value of 0.5 is within the area contributing recharge to a priority well in 50 percent of simulations. A location outside of the color flood is within the area contributing recharge to a priority well in 0 percent of simulations.

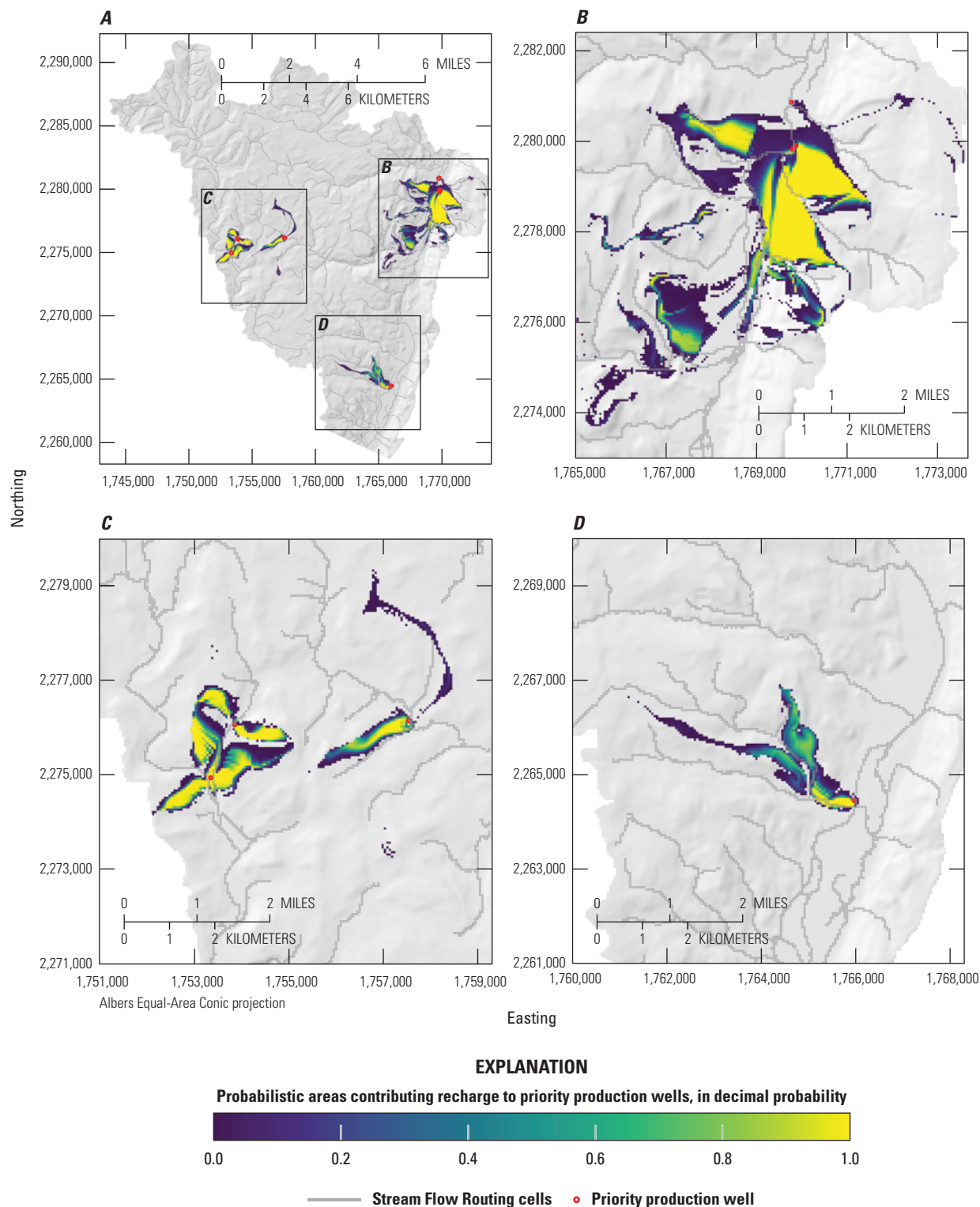


Figure 30. Probabilistic areas contributing recharge to priority wells under the “0.75rate” pumping scenario. Panels show, *A*, full study area results, *B*, detail view of priority wells near Ellenville, *C*, detail view of priority wells near Woodridge and Mountain Dale, and, *D*, detail view of priority well near Wurtsboro. Colormap is scaled to the probability that a location contributes recharge to priority wells (red circles). A location with a value of 1.0 is within the area contributing recharge to a priority well in 100 percent of simulations. A location with a value of 0.5 is within the area contributing recharge to a priority well in 50 percent of simulations. A location outside of the color flood is within the area contributing recharge to a priority well in 0 percent of simulations.

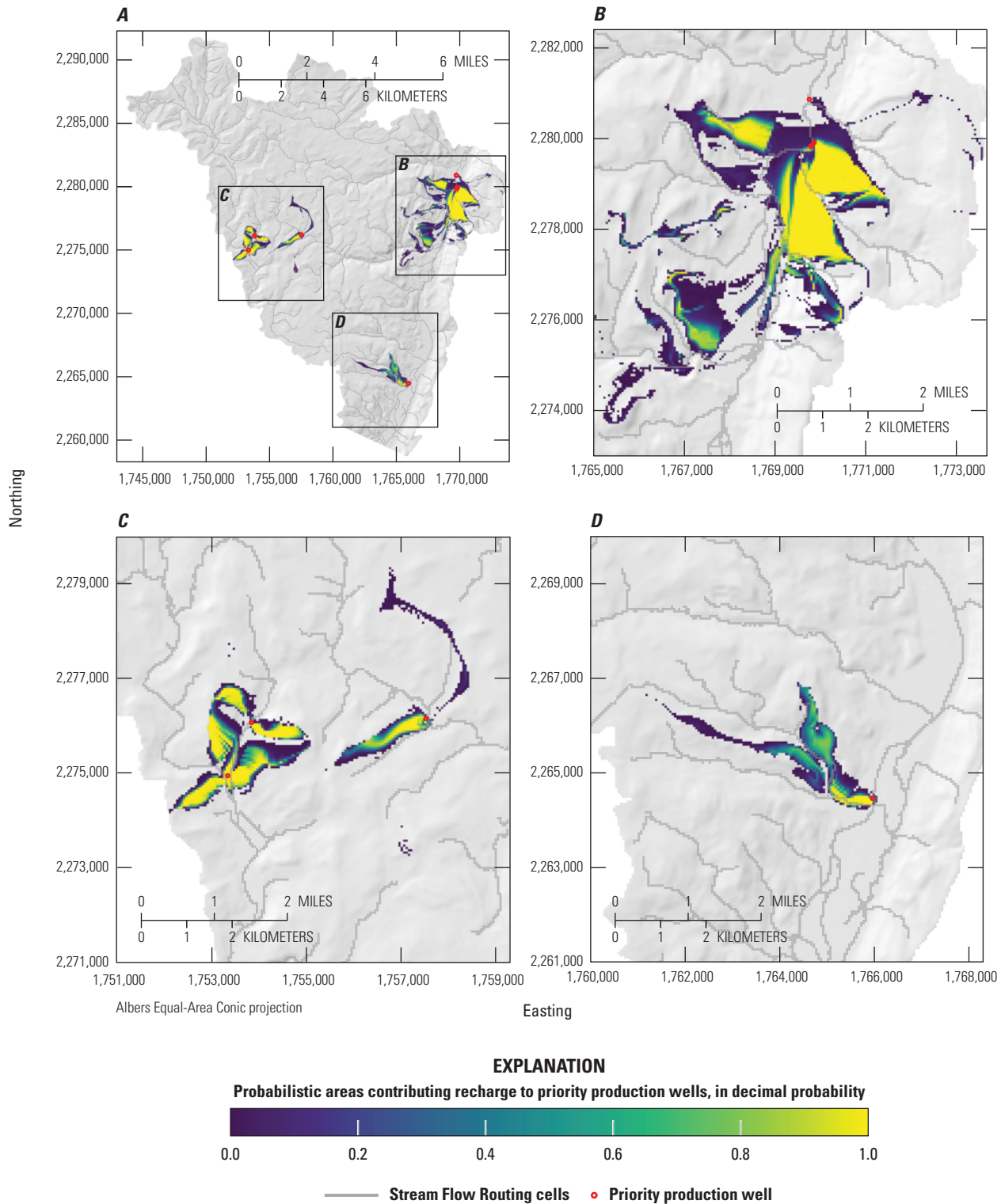


Figure 31. Probabilistic areas contributing recharge to priority wells under the “Maxrange” pumping scenario. Panels show *A*, full study area results, *B*, detail view of priority wells near Ellenville, *C*, detail view of priority wells near Woodridge and Mountain Dale, and, *D*, detail view of priority well near Wurtsboro. Colormap is scaled to the probability that a location contributes recharge to priority wells (red circles). A location with a value of 1.0 is within the area contributing recharge to a priority well in 100 percent of simulations. A location with a value of 0.5 is within the area contributing recharge to a priority well in 50 percent of simulations. A location outside of the color flood is within the area contributing recharge to a priority well in 0 percent of simulations

example, consolidating hydrostratigraphic units into model layers. We attempted to mitigate the effects of structural error by applying a highly parameterized approach to model construction and history matching, using pilot point, zone, and cell multipliers to adjust model properties with as much flexibility as we judged appropriate. Despite these limitations, the model provides a useful framework for investigating broader questions related to the regional flow system and the delineation of areas contributing recharge to priority wells.

Model Discretization

The model discretizes the groundwater-flow system into a simplified model grid, consisting of cells that are 50 m on a side and to more than 100 m in thickness. Values in these model cells therefore represent averages of the respective cell volumes. Similarly, simulated outputs from each cell represent average conditions within that cell.

Steady State Assumption

For the purposes of simulating contributing recharge areas to priority wells it was assumed that the steady state simulations were appropriate. In this model, recharge estimates from 2000 to 2013 overlap with water use estimates developed from the 2010–2015 pumping data. Perfect alignment of stresses and observations with the steady-state period is difficult to obtain which can affect the results. Furthermore, steady state conditions represent a time-averaged delineation of source water areas and can smooth over temporal variability (Rayne and others, 2014). However, the relatively high hydraulic conductivity in the glacial sediments—where the priority wells are screened—likely negates much of the transience that can cause elasticity of the source water areas.

Data Gaps

Observation data used for history matching were few and often of poor or unknown quality. This was especially true for hydraulic head observations, where all but one observation was from NYSDEC well completion reports. Flux observations were developed from BFI estimates, which are also uncertain. We adapted to this by including realizations of observation values consistent with their uncertainty for the history matching effort.

Only two flux calibration targets were available. Furthermore, the flux and hydraulic head datasets did not always overlap in time with the steady-state period. The groundwater-flow model simulated base flow only, so the targets are expected to be representative of the groundwater contribution to flow only. However, the presence of the Neversink Reservoir just upstream from streamgage 01436500 made it difficult to assess the groundwater contribution only. The uncertainty of the reservoir outlet importance at

streamgage 01436500 was, in part, incorporated into the weight for that observation. Nonetheless, more streamflow data on this stream and others in the model would better constrain the water balance.

Particle Tracking

At base pumping levels, some of the priority wells represented in the groundwater-flow model are strong sinks and others are weak sinks. Strong sinks extract water from the entire cell depth, whereas weak sink well cells extract water from some of the cell's depth but also let some water pass through the cell without being extracted (Abrams and others, 2013). Allowing particles to pass through weak sinks in this model could therefore lead to an underprediction of the areas contributing recharge to wells. To avoid underpredicting the size of the contributing recharge areas, the particle tracking analysis in this study stopped all particles when they reached weak sinks. This approach offers a conservative estimate of the areas that contribute recharge to the priority wells but may slightly overpredict the size of and the probability that these areas contribute recharge to the wells.

Each particle tracking Monte Carlo analysis consisted of 384 particle tracking simulations—one for each realization in the posterior parameter ensemble. In this study, it was assumed that the 384 realizations were sufficient to reach statistical stationarity in the probabilistic contributing recharge areas. An evaluation of the effect of ensemble size on contributing recharge area stationarity was completed by Fienen and others (2021). This evaluation indicated that higher probability areas tend to stabilize with fewer realizations than the lower probability areas and that the threshold for contributing recharge area stationarity depends on the user's adopted level of risk tolerance. The probabilistic contributing recharge areas maps produced by this study (figs. 26–31) are considered generally representative but may slightly underpredict the extent of low probability contributing recharge areas. Use of a larger ensemble for the particle tracking Monte Carlo analysis would likely have resulted in similar probability maps—particularly in high probability areas—but with some additional low-probability areas. Further discussion on the effect of ensemble size to map stationarity is presented in Fienen and others (2021).

Summary

Glacial aquifers are the principal source of water for many communities in upstate New York. The high permeability and the shallow depth to the water table in these aquifers make them highly susceptible to contamination. Protecting these aquifers from contamination is critical in areas where groundwater use is high and alternative sources of drinking water are not readily available. Water managers need accurate information about the size and location of the

land surface area that contributes recharge to production wells to make land-use planning decisions to reduce the risk of aquifer contamination.

The New York State Department of Environmental Conservation (NYSDEC) and the New York State Department of Health (NYSDOH) are working to reduce the risk of glacial aquifer contamination by delineating areas that contribute recharge to production wells and to encourage land-use planning within these zones. In southeastern New York, the villages of Ellenville, Wurtsboro, Woodridge, the hamlet of Mountain Dale, and surrounding communities rely on groundwater pumped from shallow valley-fill glacial aquifers for public water supply. In this region, which encompasses adjacent parts of the Neversink River and Rondout Creek drainage basins, the NYSDEC and NYSDOH have designated eight “priority” production wells, for which water supply protection is of primary concern. In 2019, in cooperation with the NYSDEC and the NYSDOH, the U.S. Geological Survey began a study to characterize the regional groundwater-flow system and to delineate the contributing areas for these eight priority wells for the purpose of source water protection.

To complete these objectives, a groundwater-flow model of the valley-fill aquifers in the Neversink-Rondout study area in the Neversink River and Rondout Creek drainage basins was created to simulate the groundwater system and evaluate flow paths to a series of eight priority wells. The MODFLOW 6 groundwater model was built using Python tools (such as [flop](#), [modflow-setup](#), and [sfrmaker](#)) that facilitate transparent and repeatable model development using existing datasets. The model was calibrated to hydraulic head, base flow, and land surface data with a stepwise estimation of model parameters using an iterative ensemble smoother implementation of the Parameter ESTimation software PEST++ (version 5.0.0). We evaluated prior “best guess” parameter bounds with a prior Monte Carlo analysis. Results of the first prior Monte Carlo analysis were used to make informed adjustments to model parameter bounds (typically resulting in expanded bounds, notably in bedrock hydraulic conductivity), and a second prior Monte Carlo analysis was run to identify improved ranges for model parameters during calibration. We used results of the second Monte Carlo analysis to set parameter bounds and starting values for calibration via history matching using the iterative ensemble smoother.

The history matching effort produced an ensemble of parameter values for the groundwater-flow model that spans the range of values within prior uncertainty bounds and is informed by the historical observation data, within a reasonable range of uncertainty on those observations. This history-matched ensemble was used to delineate the areas contributing recharge to priority wells by running the groundwater-flow and particle tracking (MODPATH7) models both for each ensemble member. Deterministic contributing areas computed for each ensemble member were aggregated to produce maps showing the probability that a location contributes recharge to priority wells. Finally, the particle tracking Monte Carlo analysis was repeated for six pumping

scenarios, representing a wide range of possible pumping levels, to incorporate uncertainty in future pumping rates related to population growth or other management decisions. Increasing pumping rates generally led to larger contributing recharge areas and larger areas of high probability that a location contributes recharge to priority wells. These maps show the overall uncertainty of the areas contributing recharge to priority wells in the study area and provide a tool for risk-based decision making for protection of well source water.

References Cited

- Abrams, D., Haitjema, H., and Kauffman, L., 2013, On modeling weak sinks in MODPATH: *Ground Water*, v. 51, no. 4, p. 597–602. [Also available at <https://doi.org/10.1111/j.1745-6584.2012.00995.x>].
- Bakker, M., Post, V., Langevin, C.D., Hughes, J.D., White, J.T., Starn, J.J., and Fienen, M.N., 2016, Scripting MODFLOW model development using Python and flopy: *Groundwater*, v. 54, i. 5, p. 733–739. [Also available at <https://doi.org/10.1111/gwat.12413>].
- Como, M., 2013, Hydrogeologic appraisal of the unconsolidated aquifer in Wawarsing, New York: Stony Brook, Stony Brook University, M.S. thesis, 36 p., 50 pls., [Also available at https://www.geo.sunysb.edu/reports/MComo_FinalReport.pdf].
- Doherty, J., 2003, Ground water model calibration using pilot points and regularization: *Ground Water*, v. 41, no. 2, p. 170–177. [Also available at <https://doi.org/10.1111/j.1745-6584.2003.tb02580.x>].
- Doherty, J., and Welter, D., 2010, A short exploration of structural noise: *Water Resources Research*, v. 46, no. 5, W05525. [Also available at <https://dx.doi.org/10.1029/2009wr008377>].
- Feinstein, D.T., Hart, D.J., Gatzke, S., Hunt, R.J., Niswonger, R.G., and Fienen, M.N., 2020, A simple method for simulating groundwater interactions with fens to forecast development effects: *Ground Water*, v. 58, no. 4, p. 524–534. [Also available at <https://doi.org/10.1111/gwat.12931>].
- Fienen, M.N., and Corson-Dosch, N.T., 2021, Groundwater model archive and workflow for Neversink/Rondout Basin, New York, source water delineation: U.S. Geological Survey data release, at <https://doi.org/10.5066/P9HWSOHP>.
- Fienen, M.N., Corson-Dosch, N.T., White, J.T., Leaf, A.T., and Hunt, R.J., 2021, Risk-based wellhead protection decision support: a repeatable workflow approach. *Groundwater*. [Also available at <https://doi.org/10.1111/gwat.13129>].

- Finkelstein, J.S., Woda, J.C., and Williams, J.H., 2022, Data sources and methods for digital mapping of eight valley-fill aquifer systems in upstate New York: U.S. Geological Survey Scientific Investigations Report 2022–5024, 21 p. [Also available at <https://doi.org/10.3133/sir20225024>].
- Fischer, J.M., Riva-Murray, K., Hickman, R.E., Chichester, D.C., Brightbill, R.A., Romanok, K.M., and Bilger, M.D., 2004, Water quality in the Delaware River Basin, Pennsylvania, New Jersey, New York, and Delaware, 1998–2001: U.S. Geological Survey Circular 1227, 39 p. [Also available at <https://doi.org/10.3133/cir1227>].
- Langevin, C.D., Hughes, J.D., Banta, E.R., and Niswonger, R.G., Panday, Sorab, and Provost, A.M., 2017, Documentation for the MODFLOW 6 Groundwater Flow Model: U.S. Geological Survey Techniques and Methods, book 6, chap. A55, 197 p. [Also available at <https://doi.org/10.3133/tm6A55>].
- Leaf, A.T., Fienen, M.N., and Reeves, H.W., 2021, SFRmaker and Linesink-maker—Rapid construction of streamflow routing networks from hydrography data: Groundwater. [Also available at <https://doi.org/10.1111/gwat.13095>].
- Miller, T.S., Sherwood, D.A., Jeffers, P.M., and Mueller, N., 1998, Hydrogeology, water quality, and simulation of ground-water flow in a glacial-aquifer system, Cortland County, New York: U.S. Geological Survey Water-Resources Investigations Report 96–4255, 84 p., 5 plates. [Also available at <https://doi.org/10.3133/wri964255>].
- New York State Department of Environmental Conservation, 2020, Water Well Contractor and Completion Report Record Search, accessed July 16, 2020 at <https://www.dec.ny.gov/lands/33317.html>.
- Pollock, D.W., 2016, User guide for MODPATH Version 7—A particle-tracking model for MODFLOW: U.S. Geological Survey Open-File Report 2016–1086, 35 p. [Also available at <https://doi.org/10.3133/ofr20161086>].
- Randall, A.D., 2001, Hydrogeologic framework of stratified-drift aquifers in the glaciated Northeastern United States: U.S. Geological Survey Professional Paper 1415-B, 190 p. [Also available at <https://pubs.usgs.gov/pp/1415b/report.pdf>].
- Rayne, T.W., Bradbury, K.R., and Zheng, C., 2014, Correct delineation of capture zones using particle tracking under transient conditions: Ground Water, v. 52, no. 3, p. 332–334, <https://ngwa.onlinelibrary.wiley.com/doi/10.1111/gwat.12141>.
- Smith, M.J., and Williams, J.H., 2020, Aquifer Test Locator: U.S. Geological Survey webpage, accessed July 29, 2020, at <https://doi.org/10.5066/P9XQL1MZ>.
- Tarantola, A., 2005, Inverse problem theory and methods for model parameter estimation: Philadelphia, Pennsylvania, Society for Industrial and Applied Mathematics.
- U.S. Geological Survey, 2015, 3-County LiDAR, point data, online at ftp://ftp.gis.ny.gov/elevation/LIDAR/USGS_3County2014/.
- U.S. Geological Survey, 2020a, USGS water data for the Nation: U.S. Geological Survey National Water Information System database, accessed July 22, 2020, at <https://doi.org/10.5066/F7P55KJN>.
- U.S. Geological Survey, 2020b, The StreamStats program: U.S. Geological Survey web page, accessed August 17, 2020, at <https://streamstats.usgs.gov/ss/>.
- U.S. Geological Survey, 2018, National Hydrography Dataset (ver. NHDPlus High Resolution Beta for Hydrologic Unit (HU) 4-0202 and 4-0204 (published 20180606)): U.S. Geological Survey web page, accessed September 4, 2020, at <https://usgs.maps.arcgis.com/apps/MapTools/index.html?appid=41a5c2ca49bd4a83b239450e61022d53>.
- Van Rossum, G., and Drake, F.L., 2009, Python 3 Reference Manual: Scotts Valley, Calif., CreateSpace.
- Westenbroek, S.M., Kelson, V.A., Dripps, W.R., Hunt, R.J., and Bradbury, K.R., 2010, SWB—A modified Thornthwaite-Mather Soil-Water-Balance code for estimating groundwater recharge: U.S. Geological Survey Techniques and Methods book 6, chap. A31, 60 p. [Also available at <https://doi.org/10.3133/tm6A31>].
- White, J.T., Fienen, M.N., and Doherty, J.E., 2016, A python framework for environmental model uncertainty analysis: Environmental Modelling & Software, v. 85, p. 217–228. [Also available at <https://doi.org/10.1016/j.envsoft.2016.08.017>].
- White, J.T., 2018, A model-independent iterative ensemble smoother for efficient history- matching and uncertainty quantification in very high dimensions: Environmental Modelling & Software, v. 109, p. 191–201. [Also available at <https://doi.org/10.1016/j.envsoft.2018.06.009>].
- White, J.T., Hunt, R.J., Fienen, M.N., and Doherty, J.E., 2021, Approaches to highly parameterized inversion: PEST++ version 5, a software suite for parameter estimation, uncertainty analysis, management optimization and sensitivity analysis: U.S. Geological Survey Techniques and Methods 7-C26, Book 7, Chapter 26. [Also available at <https://doi.org/10.3133/tm7C26>].
- White, J.T., Foster, L.K., Fienen, M.N., Knowling, M.J., Hemmings, B., and Winterle, J.R., 2020, Toward reproducible environmental modeling for decision support: A worked example: Frontiers in Earth Science, v. 8, article 50, 11 p. [Also available at <https://doi.org/10.3389/feart.2020.00050>].

- Wills, J., Merguerian, C., and Bigolski, J.N., 2005. Post-Alleghenian deformation of the Shawangunk ridge in NY and NJ: Conference proceeding, accessed December 15, 2020, at <https://pbisotopes.ess.sunysb.edu/lig/Conferences/abstracts05/abstracts/wills.pdf>.
- Woda, J.C., Finkelstein, J.S., and Williams, J.H., 2022, Interpolated hydrogeologic framework and digitized datasets for upstate New York study areas: U.S. Geological Survey data release, <https://doi.org/10.5066/P96R5K5R>.
- Yager, R.M., Kauffman, L.J., Buchwald, C.A., Westenbroek, S.M., and Reddy, J.E., 2018, Digital products from a hydrogeologic framework for Quaternary sediments within the glaciated conterminous United States: U.S. Geological Survey data release, accessed June 17, 2020, at <https://doi.org/10.5066/F7HH6J8X>.
- Yager, R.M., Kauffman, L.J., Soller, D.R., Haj, A.E., Heisig, P.M., Buchwald, C.A., Westenbroek, S.M., and Reddy, J.E., 2019, Characterization and occurrence of confined and unconfined aquifers in Quaternary sediments in the glaciated conterminous United States (ver. 1.1, February 2019): U.S. Geological Survey Scientific Investigations Report 2018–5091, 90 p. [Also available at <https://doi.org/10.3133/sir20185091>].
- Zell, W.O., and Sanford, W.E., 2020a, Calibrated simulation of the long-term average surficial groundwater system and derived spatial distributions of its characteristics for the contiguous United States: Water Resources Research, v. 56, no. 8. [Also available at <https://doi.org/10.1029/2019WR026724>].
- Zell, W.O., and Sanford, W.E., 2020b, MODFLOW 6 models used to simulate the long-term average surficial groundwater system for the contiguous United States: U.S. Geological Survey data release, accessed June 2, 2020 at <https://doi.org/10.5066/P91LFFN1>.

Appendix 1. Data Sources Used to Construct the Neversink-Rondout MODFLOW 6 Groundwater-flow Model

The Neversink-Rondout groundwater-flow model was constructed using a series of U.S. Geological Survey-developed Python tools that facilitate transparent and repeatable groundwater model development. These tools include the [flopy](#) (Bakker and others, 2016), [modflow-setup](#), and [sfrmaker](#) (Leaf and others, 2021) packages written in the Python programming language (Van Rossum and Drake, 2009). These

tools were used to develop a MODFLOW 6 model and input files from source data, including shapefiles, rasters and other MODFLOW models. Many of the datasets were previously published and are publicly available. Details about the sources of these datasets and the component(s) of the model they were used to build are summarized in [table 1.1](#).

Table 1.1. Summary of publicly available data sources used to build the Neversink-Rondout groundwater-flow model.

[MF6; MODFLOW 6; NWIS, U.S. Geological Survey National Water Information System; SWUDS, U.S. Geological Survey Site-Specific Water-Use Data System; NYSDEC, New York State Department of Environmental Conservation; NHDPlus, National Hydrography Dataset Plus; USGS, U.S. Geological Survey; --, Not applicable]

Model Component	MF6 package ^a	File name(s)	Reference	Notes	Link
Initial hydraulic conductivity values	NPF	--	Smith and Williams, 2020; Como, 2013	Aquifer Test Locator web tool, M.S. Thesis	https://doi.org/10.5066/P9XQL1MZ ; https://www.geo.sunysb.edu/reports/MComo_FinalReport.pdf
Well pumping rates	WEL	--	NWIS/SWUDS	--	https://nwis.waterdata.usgs.gov/nwis/wu
Model top elevation	DIS, SFR	top_50m_from_lidar.tif	Finkelstien and others, 2022; Woda and others, 2022	Derived from Lidar; https://gis.ny.gov/	https://doi.org/10.5066/P9HWSOHP
Layer elevations and thicknesses	DIS	L1_Bottom.tif, L2_Bottom.tif, L3_Bottom_TopOfBedrock.tif	Finkelstein and others, 2022	Top of lacustrine, bottom of lacustrine, top of bedrock	https://doi.org/10.5066/P9HWSOHP
Initial heads	IC	top_50m_from_lidar.tif	Finkelstien and others, 2022; Woda and others, 2022	--	https://doi.org/10.5066/P9HWSOHP
Active model domain	DIS	Model_Extent_HUC12.shp	Finkelstein and others, 2022; Woda and others, 2022	--	https://doi.org/10.5066/P9HWSOHP
Boundary condition constant heads	CHD	0108_0110_0202_0203_MF6_SS_Unconfined_250, 0204_0206_0209_MF6_SS_Unconfined_250	Zell and Sanford, 2020	Parent models; heads file used to build constant head boundaries in valleys	https://water.usgs.gov/GIS/metadata/usgswrd/XML/zell2020_wrr.xml
Streamflow routing lines	SFR	NHDPLUS_H_0202_HU4_GDB.gdb, NHDPLUS_H_0204_HU4_GDB.gdb	NHDPlus HR; U.S. Geological Survey, 2018	NHDLine	https://www.usgs.gov/core-science-systems/ngp/national-hydrography/nhdplus-high-resolution
Streamflow routing information	SFR	NHDPLUS_H_0202_HU4_GDB.gdb, NHDPLUS_H_0204_HU4_GDB.gdb	NHDPlus HR; U.S. Geological Survey, 2018	NHDPlusFlowlineVAA, NHDPlusFlow	https://www.usgs.gov/core-science-systems/ngp/national-hydrography/nhdplus-high-resolution
Stream inflow rates	SFR	--	NWIS	USGS stream gage 01436000	https://doi.org/10.5066/F7P55KJN
Recharge	RCH	SWB_Nawqa_mm.tif	Yager and others, 2018	--	https://doi.org/10.5066/F7HH6J8X
Head and flux observations	OBS	--	NWIS	--	https://doi.org/10.5066/F7P55KJN
Additional head observations	OBS	--	NYSDEC	Water Well Contractor and Completion Report Record	https://www.dec.ny.gov/lands/33317.html

^a <https://water.usgs.gov/ogw/modflow/mf6io.pdf>

References Cited

- Bakker, M., Post, V., Langevin, C.D., Hughes, J.D., White, J.T., Starn, J.J., and Fienen, M.N., 2016, Scripting MODFLOW model development using Python and flopy: *Groundwater*, v. 54, i. 5, p. 733-739. [Also available at <https://doi.org/10.1111/gwat.12413>].
- Como, M., 2013, Hydrogeologic Appraisal of the Unconsolidated Aquifer in Wawarsing, New York: Stony Brook University, M.S. thesis, 36 p., 50 pls., [Also available at https://www.geo.sunysb.edu/reports/MComo_FinalReport.pdf].
- Finkelstein, J.S., Woda, J.C., and Williams, J.H., 2022, Data Sources and methods for digital mapping of eight valley-fill aquifer systems in upstate New York: U.S. Geological Survey Scientific Investigations Report 2022–5024, 21 p. [Also available at <https://doi.org/10.3133/sir20225024>].
- Leaf, A.T., Fienen, M.N., and Reeves, H.W., 2021, SFRmaker and Linesink-maker—Rapid construction of streamflow routing networks from hydrography data: *Groundwater*. [Also available at <https://doi.org/10.1111/gwat.13095>].
- Smith, M.J., and Williams, J.H., 2020, Aquifer Test Locator: U.S. Geological Survey webpage, accessed July 29, 2020, at <https://doi.org/10.5066/P9XQL1MZ>.
- U.S. Geological Survey, 2018, National Hydrography Dataset (ver. NHDPlus High Resolution Beta for Hydrologic Unit (HU) 4-0202 and 4-0204 (published 20180606)): accessed September 4, 2020, at <https://usgs.maps.arcgis.com/apps/MapTools/index.html?appid=41a5c2ca49bd4a83b239450e61022d53>.
- Van Rossum, G., and Drake, F.L., 2009, Python 3 reference manual: Scotts Valley, Calif., CreateSpace.
- Woda, J.C., Finkelstein, J.S., and Williams, J.H., 2022, Interpolated hydrogeologic framework and digitized datasets for upstate New York study areas: U.S. Geological Survey data release, <https://doi.org/10.5066/P96R5K5R>.
- Yager, R.M., Kauffman, L.J., Buchwald, C.A., Westenbroek, S.M., and Reddy, J.E., 2018, Digital products from a hydrogeologic framework for Quaternary sediments within the glaciated conterminous United States: U.S. Geological Survey data release, accessed June 17, 2020, at <https://doi.org/10.5066/F7HH6J8X>.
- Zell, W.O., and Sanford, W.E., 2020, MODFLOW 6 models used to simulate the long-term average surficial groundwater system for the contiguous United States: U.S. Geological Survey data release, accessed June 2, 2020 at <https://doi.org/10.5066/P91LFFN1>.

For additional information contact:

Director, [Upper Midwest Water Science Center](#)

U.S. Geological Survey

1 Gifford Pinchot Drive

Madison, WI 53726

Publishing support provided by the Indianapolis Publishing Service
Center

