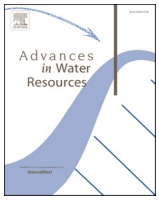
Advances in Water Resources 160 (2022) 104108

Contents lists available at ScienceDirect

Advances in Water Resources

journal homepage: www.elsevier.com/locate/advwatres



Where groundwater seeps: Evaluating modeled groundwater discharge patterns with thermal infrared surveys at the river-network scale

J.R. Barclay [a,1,*], M.A. Briggs [b], E.M. Moore [a], J.J. Starn [c], A.E.H. Hanson [b], A.M. Helton [a,d]

[a] Department of Natural Resources and the Environment, University of Connecticut, 1376 Storrs Road, Unit 4087, Storrs, CT 06269-4087, USA
[b] U.S. Geological Survey, Earth System Processes Division, Hydrogeophysics Branch, 11 Sherman Place, Unit 5015, Storrs, CT 06269, USA
[c] U.S Geological Survey, New England Water Science Center, East Hartford, CT, USA
[d] Center for Environmental Sciences and Engineering, University of Connecticut, 3107 Horsebarn Hill Rd., Unit 4210, Storrs, CT 06269-4210, USA

ARTICLE INFO

Keywords:
Groundwater discharge
Seepage
Thermal infrared
MODFLOW
model evaluation

ABSTRACT

Predicting baseflow dynamics, protecting aquatic habitat, and managing legacy contaminants requires explicit characterization and prediction of groundwater discharge patterns throughout river networks. Using handheld thermal infrared (TIR) cameras, we surveyed 47 km of stream length across the Farmington River watershed (1,570 km^2; CT and MA, USA), mapping locations of bank and waterline groundwater discharges based on their thermal signature. Using the observed groundwater discharge locations and predicted groundwater discharge rates from 6 variations of a numerical groundwater-flow model (MODFLOW-NWT), we compared 1) predicted groundwater-discharge rates in areas with and without observed groundwater discharge, 2) spatial patterns of observed and predicted groundwater discharge locations, and 3) density of observed groundwater discharge locations with predicted discharge rates. Five of six models reasonably predicted the spatial patterns of discharge locations along the 5th order mainstem, but fewer models predicted groundwater discharge patterns in smaller streams. Our results highlight 1) the feasibility of using TIR observations to evaluate groundwater models, 2) model parameters that influence discharge prediction accuracy (riverbed sediment and bedrock hydraulic conductivity and river-aquifer connections), and 3) current strengths and future opportunities for improved modeling of groundwater-discharge patterns.

1. Introduction

Model predictions of spatially explicit groundwater discharge patterns along stream networks are not often evaluated using field observations. However, accurate prediction regarding the spatial distribution and other physical characteristics of groundwater discharge to streams and rivers is critical for protecting aquatic habitat and managing contaminant inputs to streams (Briggs and Hare, 2018; Dent et al., 2001; Torgersen et al., 2012). For example, identifying the locations and source depth of groundwater discharge is essential to predicting the occurrence of thermal refuges and refugia in the face of climate change (Hare et al., 2021; Johnson et al., 2020). Similarly, groundwater flowpath depth and discharge location are important for understanding nitrogen loads to river systems (Kolbe et al., 2019; Wherry et al., 2021). Accurate prediction of groundwater discharge across river networks, however, remains challenging due to gaps in model evaluation and a lack of field data at model-relevant spatial extents and resolutions (Barclay et al., 2020a).

Groundwater models are commonly implemented at the river-network scale (drainage areas $> 10^3$ km^2) and simulate groundwater discharge to streams and rivers at the model cell resolution (model-specific, typically hundreds of meters) (Befus et al., 2017; Feinstein et al., 2010; Masterson et al., 2016; Sanford et al., 2012). This suggests that groundwater models can simulate sub-reach heterogeneity in groundwater discharge at the spatial extent of river networks. Yet, the accuracy of simulated discharge patterns has not been assessed at this scale. Recently, Barclay et al. (2020a) demonstrated that groundwater models (MODFLOW-NWT, $n = 11$) with identical recharge inputs but differing parameterization resulted in similar fit to observed water levels but 1) differences of 50% in the fraction of groundwater discharge directed to 1st order streams, 2) three-fold differences in simulated source groundwater flowpath depth, and 3) seven-fold differences in

* Corresponding author.
E-mail address: janet.barclay@uconn.edu (J.R. Barclay).
[1] Current address: U.S. Geological Survey, New England Water Science Center, East Hartford, CT, USA.

https://doi.org/10.1016/j.advwatres.2021.104108
Received 15 June 2021; Received in revised form 17 November 2021; Accepted 20 December 2021
Available online 23 December 2021
0309-1708/Published by Elsevier Ltd.

subsurface travel times among models. Differences of this magnitude in simulated flow distribution could have consequential implications for maintaining environmental flows (including stream temperature), managing contaminant legacies, or understanding nutrient processing (Barlow and Leake, 2012; Barnes et al., 2018; Chen et al., 2018; Kolbe et al., 2019; Kurylyk et al., 2015). This lack of precision regarding critical physical characteristics of groundwater discharge across a river network highlights the need for new approaches to evaluate and refine simulated patterns of discharge such as the position along the river network.

Calibration or evaluation data for groundwater-flow models commonly include groundwater head, stream elevation, stream discharge in sub-basins, and occasionally age tracer data (Sanford, 2011; Starn and Brown, 2007). Information on spatial patterning of groundwater discharge (seepage) to streams is typically not included. Although discharge rates from individual groundwater-discharge zones (Vaccaro, 2011; Yager et al., 2008, 2007) and discharge patterns in small watersheds (< 50 km^2) (Ala-aho et al., 2015; Danielescu et al., 2009; Glaser et al., 2016; Jeannot et al., 2019) have been used in a few instances, studies that explicitly assess sub-reach (hundreds of meters) spatial patterns of simulated groundwater discharge across larger watersheds ($> 10^3$ km^2) are lacking. This absence of assessment is due, in part, to a lack of empirical data that can be meaningfully compared to modeled discharge patterns and a lack of approaches for incorporating new types of field data into the modeling process.

Existing empirical approaches typically identify and quantify groundwater discharge at the point to reach scale (discrete lengths of river, typically hundreds of meters to a few kilometers), but neither point measurement nor reach-scale aggregations of groundwater discharge are straightforward to compare to model outputs due to mismatches in measurement and model resolution. At discrete points, seepage meters and temperature profilers can be robust for quantifying groundwater discharge below the waterline under favorable conditions (Caissie and Luce, 2017; Irvine et al., 2016; Rosenberry, 2008), but measuring the total groundwater discharge for a single model cell using physical point measurements would require many time-consuming individual measurements as extreme heterogeneity precludes interpolation. This is impractical over spatial extents comparable to most predictive models. Spanning the reach and point scale, fiber-optic distributed temperature sensing (FO-DTS) can identify spatial patterns of groundwater discharge along linear cables at sub-meter resolution, as well as high-frequency ($\leq$ 15 min intervals) temporal variation in groundwater discharge (Hare et al., 2015; Matheswaran et al., 2014). Groundwater-discharge observations made with FO-DTS can be easily aggregated to the model resolution, but deploying and managing the cables requires at least a week per 2 – 3 km length of river, making them impractical for river-network scale applications. At the reach scale (hundreds of meters), chemical tracers and differential gaging allow calculation of aggregate groundwater discharge assuming conservation of mass (Kalbus et al., 2006; Kilpatrick and Cobb, 1985; McCallum et al., 2012; Xie et al., 2016), but net groundwater inflow along the measurement reach must be relatively large compared to streamflow for groundwater inputs to be captured with confidence, and spatially-explicit discharge information is lost. Network-scale groundwater discharge analysis conducted using baseflow separation techniques are useful in testing net groundwater exchange (as baseflow) predictions (Cartwright and Miller, 2021; Miller et al., 2017), but as with tracer injections, specific discharge zone information is not preserved.

Newly refined remote-sensing and geophysical-based methodologies offer promise in better aligning the scales of discharge-specific field-data collection and watershed groundwater-flow model resolution (Briggs and Hare, 2018; Toran et al., 2015). In particular, thermal infrared imagery (TIR) (e.g. Ala-aho et al., 2015; Briggs et al., 2016; Schuetz and Weiler, 2011), which greatly enhances potential spatial coverage while preserving high spatial resolution of discharge-zone mapping (Briggs and Hare, 2018), may be an ideal empirical method for integrating with river-network scale groundwater models. TIR provides a time-specific map of water surface and bank temperatures along contiguous surveys. This surface temperature map can be used to identify zones, or areas, of 'preferential' groundwater discharge at times of year with contrasting groundwater and surface-water or stream-bank temperatures (Torgersen et al., 2001). Preferential groundwater discharge differs from spatially diffuse discharge in that groundwater emerges at a relatively high flux rate in discrete patches such that source groundwater temperature signatures are preserved, allowing identification via heat tracing (Briggs and Hare, 2018). Generally, groundwater discharge observed with TIR at or above the water line (on riverbanks) occurs as discrete points at the sub-meter scale, or when multiple discharge points occur in close proximity over more extensive (tens of meters) bank areas, as preferential discharge faces (Mundy et al., 2017). Drone-based TIR has recently been applied to the evaluation of groundwater discharge-based large wetland restoration design (Harvey et al., 2019), validation of discharge patterns from integrated hydrologic models in small (< 5 km^2) watersheds using handheld instruments (Glaser et al., 2016; Jeannot et al., 2019) and mapping potential thermal refugia zones based on the occurrence of discharge faces along higher-order rivers across hundreds of kilometers using airborne platforms (Ala-aho et al., 2015; Dugdale et al., 2015).

Field surveys conducted with TIR have the potential to provide field observations at a spatial extent and resolution that can be meaningfully compared with river-network scale groundwater-flow models. TIR is more efficient to implement than other empirical methods (e.g., seepage meters, differential gaging, chemical tracers), though unlike several such methods TIR cannot be used to determine groundwater discharge flux rates, except potentially under highly specific circumstances such as frozen discharge faces (Pandey et al., 2013). Warm weather riverbank discharge plume size and temperature is governed by a range of localized hydrodynamic factors, and relatively cold discharge water tends to plunge downward into the water column out of the view of TIR, hampering efforts to directly infer discharge rates from TIR images. However, TIR can be used to quickly identify discharge *areas, or zones,* across contiguous space in perhaps a more complete manner than any other current groundwater/surface water exchange processes methodology. The mapped locations of observed preferential groundwater-discharge, particularly when they occur along more extensive discharge faces, could potentially be used to evaluate spatial patterns of groundwater discharge predicted by river-network scale flow models as the development of discharge faces is likely driven by geologic and hydraulic factors more similar to model scales than singular discharge points.

Our primary objective in this study was to test the utility of TIR-based field measurements of preferential groundwater-discharge points and faces for evaluating model-cell scale spatial patterns of groundwater discharge predicted by a river-network scale groundwater model. This is a substantial step towards resolving the disparate scales of detailed stream-reach field studies and coarse river-network scale simulations. In this project, we characterized spatial patterns of groundwater discharge at the river-network scale using extensive TIR field surveys and then evaluated six groundwater models for a 5th order watershed representing different spatial patterns of groundwater discharge but similar model fit metrics to groundwater levels (Barclay et al., 2020a) against the field observations. Ultimately, this study demonstrates an approach to field verifying the predicted spatial patterns of groundwater discharge from river-network scale groundwater models, thereby highlighting areas for future improvement in our ability to simulate groundwater discharge at the river-network scale.

2. Methods

Briefly, we characterized spatial patterns of preferential groundwater discharge across a river network using spatially extensive ground-

based TIR field surveys, then compared these empirical results to river-network scale groundwater-flow models. We surveyed 47 km of total stream length (across stream orders) with TIR imagery to map areas of groundwater discharge across headwater areas, 3rd order tributaries and the mainstem of a 5th order watershed. We then selected six meaningful physical variations of a previously published groundwater-flow model (Barclay et al., 2020a, 2020b) with similar calibration-fit metrics but differing predicted spatial patterns of groundwater discharge. We used the field-mapped spatial distributions of groundwater discharge locations to evaluate predicted patterns of groundwater discharge from the calibrated models. This independent evaluation of the models with observed spatial patterns of discharge is not typical for groundwater models and is a unique strength of our approach. Details of our methods are given in the following sections.

2.1. Site description

The study site was the Farmington River watershed (MA and CT, USA) (Fig. 1). The Farmington River drains an area of 1570 km^2 and is a 5th order tributary of the Connecticut River. Glacial till overlies crystalline bedrock over most of the watershed with areas of stratified glacial sediments over sedimentary rock in the river valleys (Olcott, 1995; Soller et al., 2012). Previous geophysical mapping has indicated that groundwater discharge patterns may be driven in part by near-surface bedrock type, in addition to the unconsolidated surficial geology (Lane et al., 2020). Land cover is predominantly forest, particularly in the central and northern regions of the watershed; developed land is focused along the eastern river valley (Homer et al., 2015). In the uplands the stream channels were predominantly cobbled with some areas of exposed bedrock and the canopies were typically closed. In the eastern river valley, the stream gradients were lower, bed sediment was primarily coarse or fine sand, and the canopy was open, particularly along the 5th order section of the Farmington River. Further description of the site is given in Barclay et al. (2020a).

2.2. Field surveys

To map locations of groundwater discharge along stream banks, we surveyed 47 km of non-contiguous stream reaches within the Farmington River watershed using TIR. Due to practical considerations, we surveyed small streams by wading/walking along banks and larger rivers by canoe from the center of the channel. We waded eight 1st order stream sites (4.7 km total), 6 km of a 2nd order stream, and three 3rd order stream sites (1.6 km total). We paddled 2.7 km of a 3rd order stream and 32 km of the main stem of the Farmington River (5th order) (Table 1). We conducted the surveys over 13 days between July 6 and October 5, 2017 and 7 days between August 2 and September 22, 2019. Streams were surveyed once, though some reaches required multiple days.

During the surveys, we scanned the banks using one of four FLIR cameras (T640, T620, E8, and i7; FLIR Systems, www.flir.com), with instrument emissivities set consistent with fresh water in summer (0.96 – 0.98) (Handcock et al., 2012), and viewfinder temperature ranges set for the T640, T620, and E8 cameras to span stream and groundwater temperatures. In areas with apparent colder thermal anomalies of multiple degrees on the banks and along waterline, we collected TIR images, noted the latitude and longitude using a hand-held GPS unit (Garmin Colorado 400t or Garmin GPSMAP 64 s, www.garmin.com), and recorded the subsurface (11 cm depth) temperature using a precise hand-held digital thermometer (https://www.traceable.com/4000-traceable-digital-thermometer.html). Absolute temperatures calculated from TIR images can be affected by reflected radiation (Baker et al., 2019; Handcock et al., 2012), therefore we identified groundwater discharge based on multiple lines of evidence: thermal anomalies in the surface temperature, subsurface temperatures consistent with discharging groundwater, and qualitative indicators of groundwater discharge, such as wet soil or flowing water above the water line. Because the TIR images were used for mapping the locations of cold thermal anomalies attributable to groundwater discharge and not for

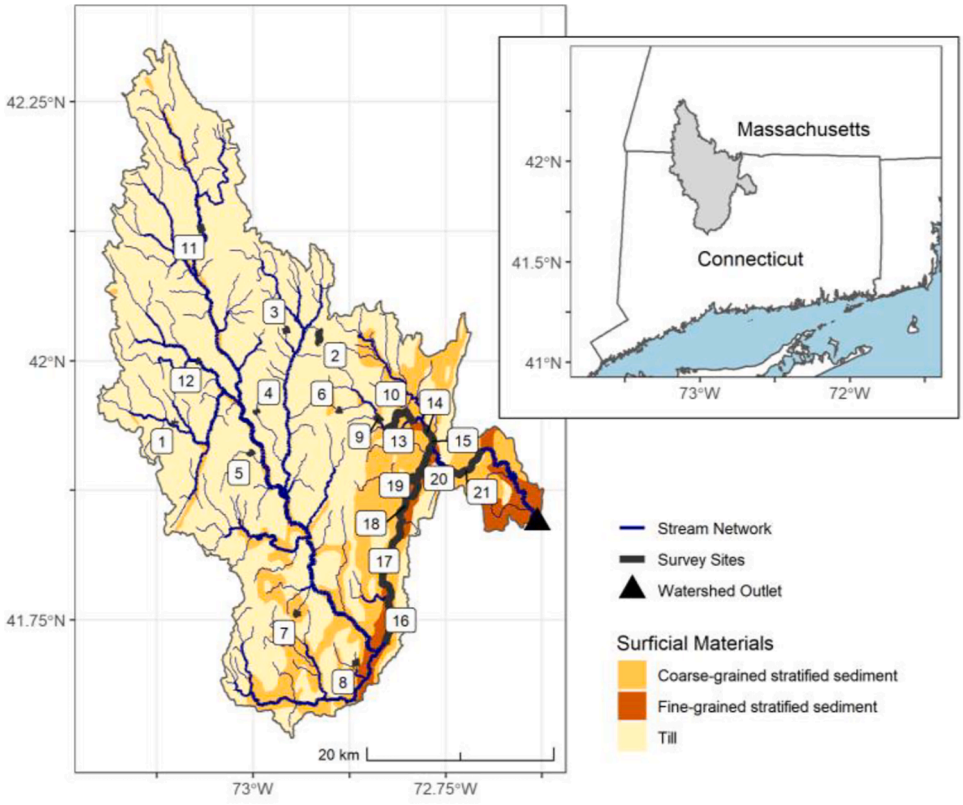


Fig. 1. Farmington River Watershed. Numbers refer to survey sites listed in Table 1.

Table 1
Field Sites (*P* = Paddled, *W* = Waded; DF = Discharge Face, Pt = Discharge Point; GT = Glacial Till, CG = Coarse-grained stratified sediment, FG = Fine-grained stratified sediment).

Site Number	Site Name	Stream Order	Survey Length (m)	Access	Discharge Types	Survey Dates	Drainage Area (km^2)	Survey Elevation (m)	Surficial Materials in Survey Reach
1	Tributary to Mad River	1	250	W	Pt	9-Aug-17	0.9	306	GT
2	Hurricane Brook	1	1300	W	Pt	7-Aug-17	2	298	GT
3	Falls Brook	1	400	W	Pt	7-Aug-17	0.8	347	GT
4	Beaver Brook	1	300	W	Pt	9-Aug-17	6	237	GT
5	Morgan Brook	1	500	W	DF, Pt	8-Sep-17	18	169	GT
6	Tributary to Salmon Brook, West Branch	1	400	W	Pt	8-Sep-17	8	266	GT
7	Tributary to Punch Brook	1	650	W	Pt	2-Aug-17; 8-Aug-17	0.7	201	CG
8	Tributary to Scott Swamp Brook	1	700	W	DF, Pt	8-Aug-17	3	62	CG; GT
9	West Branch Salmon Brook	2	750	W	DF, Pt	12-Aug-19; 19-Sep-19	49	96	CG; GT
10	West Branch Salmon Brook	2	5300	W	DF, Pt	2-Aug-19; 5-Aug-19	55	68	CG; GT
11	Farmington River, West Branch	3	700	W	Pt	11-Sep-17	133	327	GT
12	Sandy Brook	3	100	W	Pt	9-Aug-17	89	245	GT
13	West Branch Salmon Brook	3	350	W	DF, Pt	6-Aug-19	67	60	CG; GT
14	Salmon Brook	3	1400	P	DF, Pt	20-Aug-19	174	50	FG; CG
15	Salmon Brook	3	1250	P	DF, Pt	22-Sep-19	188	50	FG; CG
16	Farmington River, Access Point 21 - 22	5	9750	P	Pt	15-Sept-17; 2-Oct-17	1202	47	FG; CG
17	Farmington River, Access Point 22 - 23	5	5500	P	DF, Pt	26-Jul-17	1235	45	CG
18	Farmington River, Access Point 23 - 24	5	3400	P	DF, Pt	6-Jul-17; 26-Jul-17	1279	45	CG; FG
19	Farmington River, Access Point 24 - 25	5	5100	P	DF, Pt	26-Jul-17; 1-Aug-17	1292	45	FG; CG
20	Farmington River, Access Point 25 - 26	5	2500	P	Pt	6-Jul-17; 27-Sept-17	1303	45	FG; CG
21	Farmington River, Access Point 28 - 29	5	5400	P	DF, Pt	13-Sep-17	1518	29	CG; GT

quantitative analysis of the absolute temperatures, we did not post-process the images or apply corrections to the calculated temperatures. We delineated the approximate linear discharge-zone length along the bank. We categorized two types of preferential groundwater discharge observed along streambanks: 1) preferential groundwater-discharge points (one or more individual sub-meter scale zones of preferential discharge) and 2) discharge faces (extensive lengths, tens to hundreds of meters long, of grouped cold anomalies

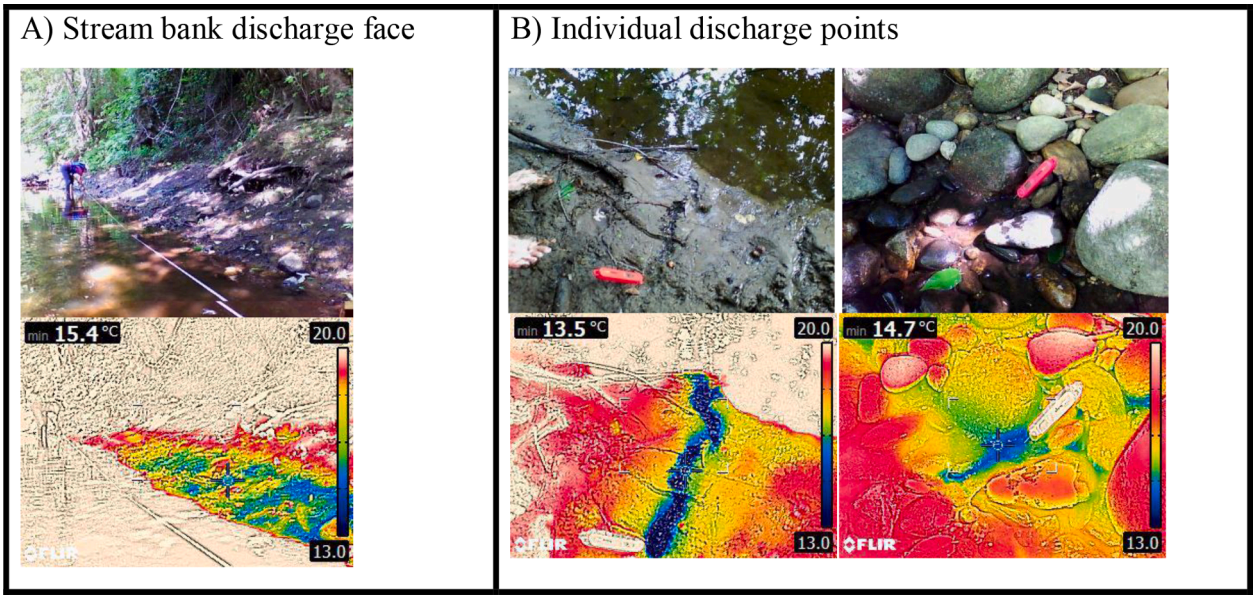


Fig. 2. Examples of preferential, bankside groundwater-discharge observed with handheld thermal infrared (TIR) cameras: A) Stream-bank groundwater discharge face and B) Individual groundwater-discharge points. In each box, photos on the top are visible color and photos on the bottom are the corresponding thermal infrared image. In the thermal infrared images, cooler colors indicated colder temperatures and warmer colors indicate warmer temperatures. Groundwater discharge can be identified by the dark blue, cold anomalies. Images provided by Eric Moore.

along the bank) (Fig. 2). In this study, we use the term "discharge zones" to refer to all areas of groundwater discharge, regardless of spatial extent, and "discharge point" and "discharge face" to distinguish between spatially limited and spatially extensive areas of groundwater discharge as defined above. Images from the TIR surveys are available in Barclay et al. (2019) and Moore et al. (2020).

2.3. Groundwater-flow model

We selected a subset (n = 6) of previously implemented groundwater-flow models for our study site that had similar fit to groundwater level and stream elevation calibration data, but varied in their predictions of spatial patterns, flowpath depth, and travel time of groundwater discharge (Barclay et al., 2020a). Similarities across models are described here; differences are discussed in the next paragraph. The models were developed in MODFLOW-NWT (Niswonger et al., 2011) and they simulate flow under steady-state conditions, with a uniform horizontal grid and 4 vertical layers. The unconsolidated surficial materials are split vertically into three layers, with decreasing horizontal hydraulic conductivity with depth. The lowest layer is a bedrock layer. The horizontal hydraulic conductivity was calibrated using PEST++ (Welter et al., 2015), mean head measurements from 990 wells and water levels from 1st order streams in flat areas (land surface slope <= 2%), and a grid of pilot points. Pilot points are preselected points at which hydraulic conductivity is calibrated; hydraulic conductivity at all other locations is interpolated from the pilot points. Simulated water levels were compared to head measurements from 984 wells and 525 stream water levels with root mean squared errors (RMSE) ranging from 9.8 to 11.6 m (Table 2). The models were also evaluated using an OverallError metric, defined as the percent of flooded terrestrial cells (incorrectly simulated as flooded) plus the percent of dry river cells (incorrectly simulated as dry) plus a penalty for excessive losing river reaches (in excess of 5 percent of recharge). The OverallError metric is based on the work of Starn and Belitz (2018).

The selected models differ in the zonation of pilot points, degree of variation in bedrock and riverbed hydraulic conductivity, allowed directionality of river-aquifer exchange and horizontal grid resolution (Table 2). The *Initial* model was developed based on readily available data and common simplifying assumptions. It has a 300 m horizontal grid, uniform riverbed hydraulic conductivity (0.1 m d^{-1}), literature-based bedrock hydraulic conductivity (Domenico and Schwartz, 1997; U.S. Geological Survey, 2007), and a 4000 m grid of pilot points. Streams were simulated using the RIV package, which allows stream reaches to be net gaining or losing at the model cell resolution (Harbaugh, 2005). In *SurfMatZone*, the pilot points are assigned to one of four zones based on surficial material type (glacial till, coarse stratified sediments, fine stratified sediments, and open water) and a finer grid (1000 m) was used for the pilot points in the coarse sediments, fine sediments and open water zones because these zones occupied a smaller area where the 4000 m spacing was impractical (442 pilot points total vs 102 with the *Initial* model). *RiverbedK* is identical to the *Initial* model, except that the riverbed hydraulic conductivity is spatially varying and assigned as 0.01 times the hydraulic conductivity of the surrounding surficial material. *RivK_BedK* is similar to *RiverbedK*, except that the hydraulic conductivity in the bedrock is calibrated on a 4000 m grid of pilot points, rather than being based on literature values. *HighRes* is identical to *RivK_BedK*, except that the grid cell size is reduced to 100 m. Finally, *RivK_BedK_drn* is similar to *RivK_BedK* except that rivers are represented using the DRN package in MODFLOW, which allows flow only from aquifer to river (streams are all gaining or disconnected), instead of the RIV package (Harbaugh, 2005). Further details of model development are available in Barclay et al. (2020a) and the model input and output files are available in Barclay et al. (2020b).

Despite the similarity in model fit as indicated by the RMSE and the OverallError metrics, the models had substantial differences in groundwater flowpath characteristics (Table 2 and Fig. 3). For example, the fraction of groundwater discharge directed to small (1st order) streams ranged by a factor of 1.4, from 29 – 40%; the median travel time ranged by a factor of 3.8, from 1.1 to 4.2 years, and the median flowpath depth ranged by a factor of 2.6, from 6.8 to 17.9 m (Table 2). Similar differences in simulated discharge characteristics were seen in the subset of reaches included in the field surveys of this study (Fig. 3). For example, mean groundwater discharge rates to 1st order streams ranged by a factor of 4, from 0.4 to 1.2 $m^3m^{-1}d^{-1}$, with a mean of 0.8 $m^3m^{-1}d^{-1}$. In this paper, simulated discharge rates are expressed as a volume of discharge per length of stream per day.

2.4. Comparing locations of modeled and observed groundwater discharge

To evaluate the simulated spatial patterns of groundwater discharge, we compared the simulated discharge predictions with our field observations of preferential discharge points and faces. First, we quantitatively compared simulated groundwater-discharge rates in sections with

Table 2
Groundwater model cases. A notation of "–" indicates the model is identical to the *Initial* Model. Modified from Table 1 in Barclay et al. (2020a).

Name	Grid	Calibration Data	Pilot Points	River Package	Riverbed K	Bedrock K	RMSE (m)	Overall Error[1]	Percent of Discharge to 1st Order Rivers	Subsurface Travel Time[3], IQ[4] Range (median) (yr)	Max Flowpath Depth[5], IQ[4] Range (median) (m)
Initial	300 m	990 wells, 525 stream elevations	4000 m grid	RIV	Uniform (0.1 m d^{-1})	Literature Values	11.2	7%	40%	0.5 – 6.3 (1.8)	1.6 – 25.7 (6.8)
SurfMatZone	–	–	Zones	–	–	–	11.6	5%	36%	0.3 – 5.8 (1.4)	3.5 – 28.4 (7.0)
RiverbedK	–	–	–	–	Spatially Varying[2]	–	10.7	6%	29%	0.3 – 7.2 (1.1)	4.0 – 27.8 (7.1)
RivK_BedK	–	–	–	–	Spatially Varying[2]	Calibrated	10.0	5%	29%	0.7 – 45.0 (4.0)	4.0 – 55.4 (16.0)
RivK_BedK_drn	–	–	–	DRN	Spatially Varying[2]	Calibrated	10.2	6%	35%	0.7 – 46.8 (4.2)	4.5 – 58.9 (17.9)
HighRes	100 m	984 wells, 2962 stream elevations	–	–	Spatially Varying[2]	Calibrated	9.8	8%	32%	0.7 – 47.2 (4.2)	5.8 – 54.4 (13.4)

[1] OverallError is a percentage of model cells in error, as noted by flooded terrestrial cells and dry river cells, plus a penalty for excessive losing reaches (modified from Starn and Belitz, 2018). .
[2] Riverbed hydraulic conductivity in each reach was 0.01 times the hydraulic conductivity of the surficial materials in the surrounding cell.
[3] Based on particle tracking.
[4] IQ is the interquartile range of the values.
[5] This is the interquartile range and median of the maximum depths of each flowpath within the model domain, calculated using particle tracking.

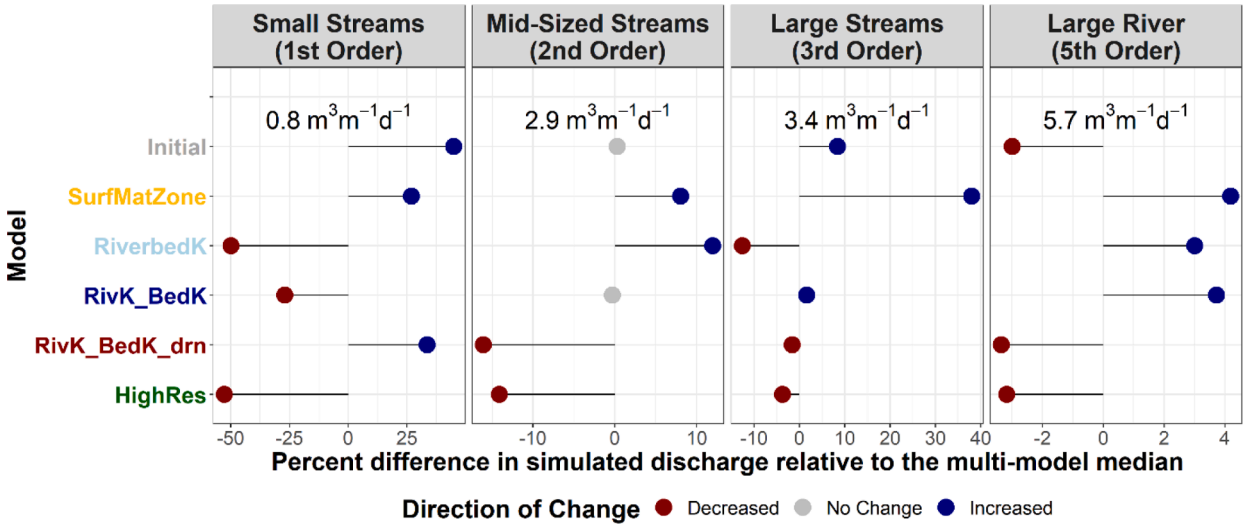


Fig. 3. Variation in simulated groundwater discharge in surveyed reaches by network position, represented as differences from the multi-model median. Numbers at the top of each panel indicate the mean discharge per length of stream for the multi-model median to streams of the respective size. Colors indicate changes (red = decrease, blue = increase, gray = no change) from the multi-model median. Note variation in x-axis scale across panels.

and without observed groundwater-discharge locations. Based on our field observations, model cells were classified into two groups by the presence or absence of observed areas of groundwater discharge, regardless of the number or spatial extent of the observed discharge, and we analyzed statistical differences in mean modeled discharge rate between the two groups. Models with statistically significant differences at the $p \leq 0.05$ level were considered "strongly distinguishing" and at the $0.05 < p \leq 0.1$ level were considered "weakly distinguishing". All other models were considered poorly performing. We considered the stream sections by stream order due to differences in simulated discharge rate and survey length. We analyzed all data at the model cell resolution and also aggregated the observations and simulated discharge rates to stream lengths of approximately 500 m for the 5th order river. Observations were not aggregated in the small streams due to short survey lengths. We determined statistically significant differences using a comparison of means that does not require normality, equal variance or equal sample size (Herberich et al., 2010). All statistical analyses were completed in R (R Core Team, 2019).

Finally, we qualitatively and semi-quantitatively compared the observed and simulated spatial patterns of groundwater discharge across models and within and across sites for each stream order. To semi-quantitatively analyze the spatial patterns of groundwater discharge we calculated the correlations between the simulated discharge rate and the observed discharge density (number of locations per length of stream, faces were counted once for every 10 m of length) for each stream order. We calculated the correlations at the model cell resolution, and for the 1st – 3rd order surveys across sites / sub-sites, and for the 5th order survey at aggregated lengths of approximately 500 m, 1 km, and 2 km. Because the survey lengths varied substantially among 1st – 3rd order stream sites, sites with long survey lengths were subdivided into subsites with lengths similar to the median for the stream order. We compared the discharge-location density to the simulated discharge rate using a parametric correlation (Pearson's r) for normally distributed model-survey pairings and a non-parametric correlation (Spearman's rho) for the other pairings. Models with statistically significant ($p \leq 0.05$) correlations between observed discharge-location density and simulated discharge were considered better performing models.

3. Results

We observed preferential groundwater discharge at or above the waterline in all mapped stream sections, but discharge points and faces were not evenly distributed along sections and varied in occurrence between sections. Further, some discharge zones were clearly higher flux (based on visible bank-surface flow; Fig. 2B) colder groundwater, with discharge temperature approaching the annual surface mean (6.9 – 10.5 °C, PRISM Climate Group, 2012), while other zones of presumed lower flux (based on the absence of visible flow) discharge were of warmer temperatures. Discharge faces in the 5th order river tended to be longer in length (e.g., Fig. 2A) than discharge faces in the smaller streams, though there was substantial variation. Also, some discharge zones in the 5th order river had flow channels visible in the infrared (note the dark blue flow lines in Fig. 2A) or visible color images (without infrared).

Temperatures in the observed groundwater-discharge zones were consistent with groundwater temperatures in our study area (interquartile range = 10.0 - 13.0 °C, mean = 11.5 °C, U.S. Geological Survey, 2020)(Fig. 4). Bank surface temperatures from the TIR images ranged from 7.1 to 16.6 °C, and subsurface (11 cm depth) temperatures ranged from 7.9 to 16.0 °C, with most groundwater-discharge zones between 10 and 15 °C on both the surface and in the subsurface. Even at similar elevation groundwater can be expected to show a range of temperatures in summer based on the depth of the contributing flowpath (Briggs et al., 2018). Most of the coldest discharge zones (surface temperature < 9.5 °C or subsurface temperature < 11.0 °C) were in the 5th order survey, though they were typically separated by warmer discharge zones and hundreds of meters of inactive banks.

Groundwater-discharge zones were not uniformly distributed along stream reaches. In both the larger river and smaller streams, groundwater-discharge points tended to cluster (Figs. 4 and 5) so that some stream sections contained multiple discharge points and others were relatively inactive. The grouping of discharge faces is even more pronounced (Fig. 5). In the upstream 10 km of the larger-river survey we observed no discharge faces, but in the next 12 km, we observed 12 distinct discharge faces covering approximately 1.5 km of river length. During the 1st to 3rd order stream surveys, we only observed discharge faces at 7 out of 16 sites, Salmon Brook (Sites 9, 10, 13, 14, and 15) Morgan Brook (Site 4) and the Tributary to Scott Swamp Brook (Site 8). Most of the discharge faces found in 1st to 3rd order streams are tens of meters in extent.

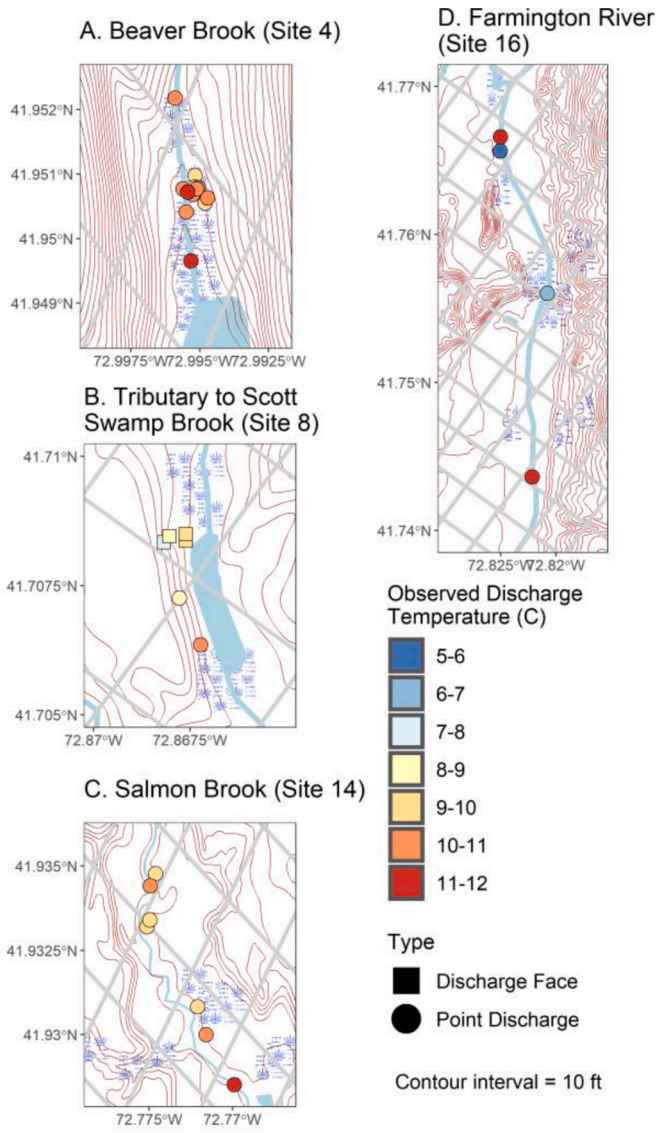


Fig. 4. Observed groundwater discharge at 4 example sites. Sites 4 & 8 (A & B) were waded and sites 14 & 16 (C & D) were paddled. Gray boxes indicate boundaries of the 300 m model cells. Circles and squares indicate observed discharge locations, colored by the observed temperature.

3.1. Model evaluation

3.1.1. Predicted groundwater discharge in areas with and without observed discharge zones

In the 5th order river, most models predicted significantly higher rates of groundwater discharge in areas where we observed groundwater discharge than in areas without observed discharge zones when the observations and simulated rates were aggregated to stream lengths of approximately 500 m ($p \leq 0.05$, Fig. 6, bottom row). The one model that did not predict a statistically significant difference between areas with and without observed discharge zones in the larger river was *SurfMatZone*, where the pilot points were assigned to zones based on the surrounding surficial material. The source of this misfit in the larger river with the *SurfMatZone* model may be anomalously high calibrated hydraulic conductivity values in coarse stratified sediments (median for *SurfMatZone* = 243 m d^{-1} vs 39 – 44 m d^{-1} for all other models). When the observations and simulated discharge rates were compared at the model cell resolution (100 m for *HighRes*; 300 m for all other models), only the *HighRes* model weakly predicted higher rates of groundwater discharge in areas where we observed groundwater discharge than in areas without observed groundwater discharge zones ($p \leq 0.10$).

In contrast, in the smaller streams and rivers (1st – 3rd order, Fig. 6, rows 1 through 3), fewer models aligned with observed discharge patterns, i.e., predicted significantly higher rates of groundwater discharge in areas with TIR-observed discharge zones ($p \leq 0.05$, Fig. 6, rows 1 through 3), with greater agreement in 3rd order rivers than in 1st and 2nd order streams. The best performing model across 1st, 2nd, and 3rd order streams was *SurfMatZone,* where calibration is by surficial deposit zones. This is in contrast to the larger river where *SurfMatZone* performed poorly. The differences in mean simulated discharge rates between areas with and without observed discharge zones was smaller for 1st and 2nd order streams (differences of −0.18 m^3d^{-1} and −0.87 m^3d^{-1}, respectively, across all models, the negatives indicate that some models predicted higher discharge rates where we observed no discharge than where we observed discharge) than for 3rd order rivers (mean difference of 1.77 m^3d^{-1}), suggesting that the models better simulated observed discharge zones in 3rd order streams than in smaller 1st and 2nd order streams.

3.1.2. Location comparison of modeled and observed groundwater discharge

In the 5th order river, all models reflected the general spatial patterns of discharge we observed (Fig. 5D), consistent with the above quantitative analysis (Fig. 6A). All models predicted relatively constant discharge from km 1 to km 10, a section where we saw little evidence of preferential bank groundwater discharge, and all models predicted the apparent increase in discharge that we observed from km 11 - 21. None of the models predicted the increased discharge we observed in km 21 – 22, and only *SurfMatZone* predicted the greater discharge we observed at km 30–32, but *SurfMatZone* predicted discharge we did not observe along the banks at km 35 – 36.

In the 1st – 3rd order streams, most models simulated similar magnitude discharge across the sites (Fig. 5, A-C). Notable exceptions occurred in the 1st order survey at sites 7 and 4 and in the 3rd order survey at sites 11, 13, and 14. At site 7 in the 1st order survey (Fig. 5A), 3 models (*RiverbedK, HighRes*, and *RivK_BedK*) predicted substantial flow from the stream to the aquifer across the entire site, while the remaining 3 models simulated near-zero net discharge. These models (*RiverbedK, HighRes*, and *RivK_BedK*) have spatially varying riverbed hydraulic conductivity *and* simulate streams using the RIV package (which allows net gains or losses to the stream), a combination that can result in unrealistically high flows from river to aquifer. Although the simulation results might suggest predominantly losing conditions at site 7, this is unlikely. Stream temperatures at site 7 suggest substantial groundwater exchange and net gaining conditions, and the lack of a stream network upstream of the site suggests streamflow is sourced locally, likely from groundwater discharge. More likely the groundwater models are unable to simulate the near-surface water table needed to generate gaining conditions, possibly due to the coarse model resolution and headwaters location. At site 4 in the 1st order survey, the three models with calibrated hydraulic conductivity in the bedrock (*HighRes, RivK_BedK*, and *RivK_BedK_drn*) simulated no-flow or slight losses from stream to aquifer, while the other 3 models simulated gaining conditions throughout the site. At site 11 in the 3rd order survey, the higher-resolution model (*HighRes*) simulated slightly different locations for the higher discharge peaks than the remaining models. This could be a consequence of differing model resolutions that led to differing locations of breaks and gradients in the hydraulic conductivity. At sites 13 and 14 in the 3rd order survey (Fig. 5C), *SurfMatZone,* in which hydraulic conductivity was calibrated by zones of surficial-material type, simulated higher rates of discharge over portions of the survey than the other models. Zonation of the hydraulic conductivity results in sharper breaks in the spatial patterns of the calibrated values, which likely caused differing patterns of simulated discharge. In the 2nd order survey, all models simulated similar spatial patterns, and all models simulated higher rates of discharge in the downstream end of site 10 as compared with the

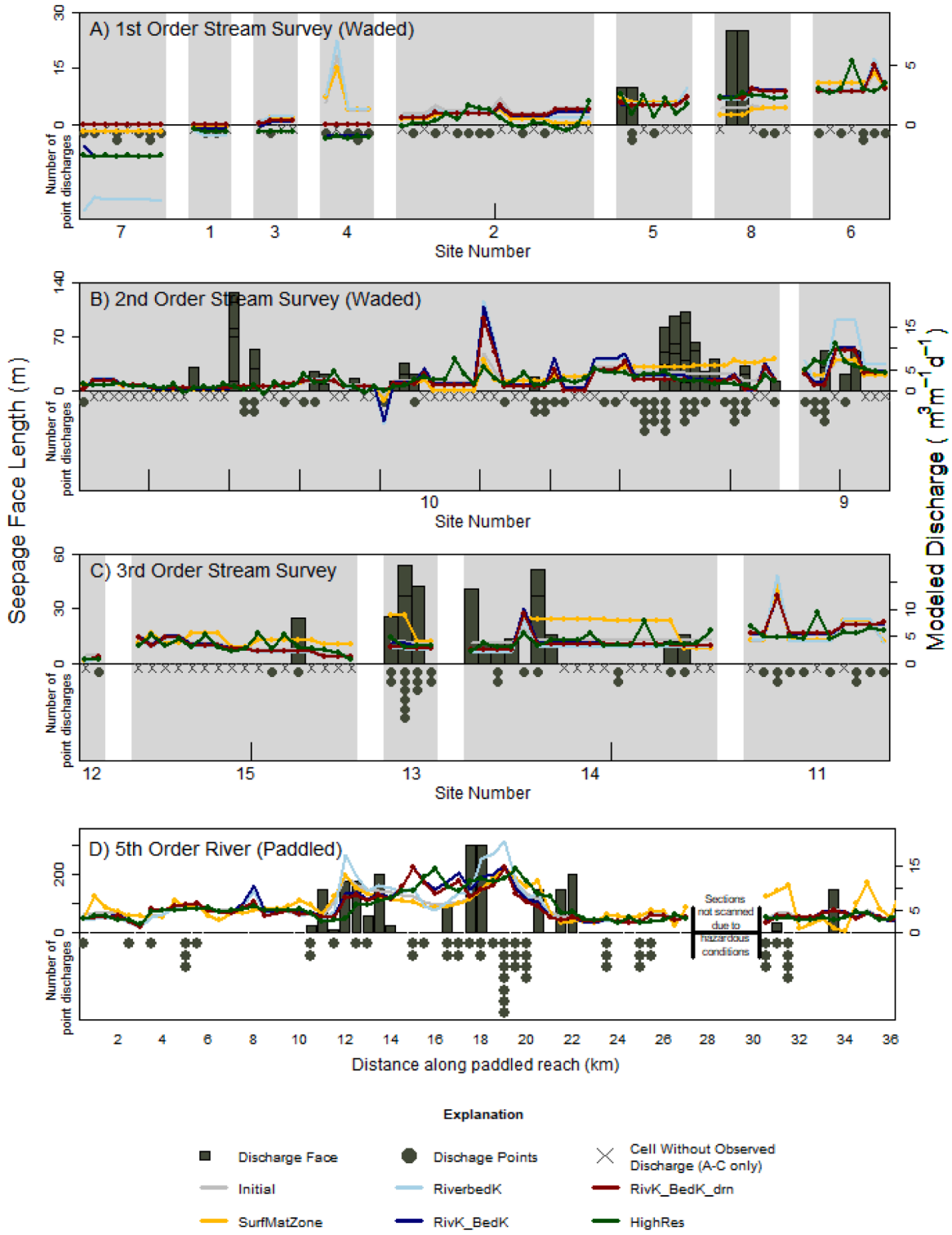


Fig. 5. Observed and simulated groundwater discharge from the 1st order (A); 2nd order (B); 3rd order stream (C); and 5th order river surveys (D). Observations are grouped by site (numbers refer to sites from Table 1) and 100 m model cells, and ordered from left to right by the multi-model mean discharge by site. Models with a 300 m resolution (all models except *HighRes*) were linearly downscaled to 100 m prior to plotting. Gray bars (left axis) indicate the length of discharge face (extended areas of discharge), circles indicate individual discharge points within each model cell and the x's in panels A – C indicate cells without observed groundwater-discharge zones. Short vertical lines on the x axis indicate the subsite breaks used for observed discharge density and simulated discharge rate correlation analysis. The lines are modeled discharge for the respective reach or cell (right axis). In *SurfMatZone*, plotted in yellow, zones of surficial materials were used in calibrating the hydraulic conductivity. *RiverbedK*, plotted in light blue, has spatially varying riverbed hydraulic conductivity, uses the RIV package, which allows net losing reaches, and uses literature values for the hydraulic conductivity in the bedrock. *RivK_BedK*, plotted in dark blue, is similar to *RiverbedK*, but the hydraulic conductivity in the bedrock is calibrated. *HighRes*, plotted in dark green, is similar to *RivK_BedK*, but with a finer resolution grid. *RivK_BedK_drn*, plotted in dark red, has calibrated hydraulic conductivity in the bedrock layer and spatially varying riverbed hydraulic conductivity combined with the DRN package, which allows net gaining or no-exchange reaches. Simulated discharge is expressed as a volume of discharge per length of stream per day.

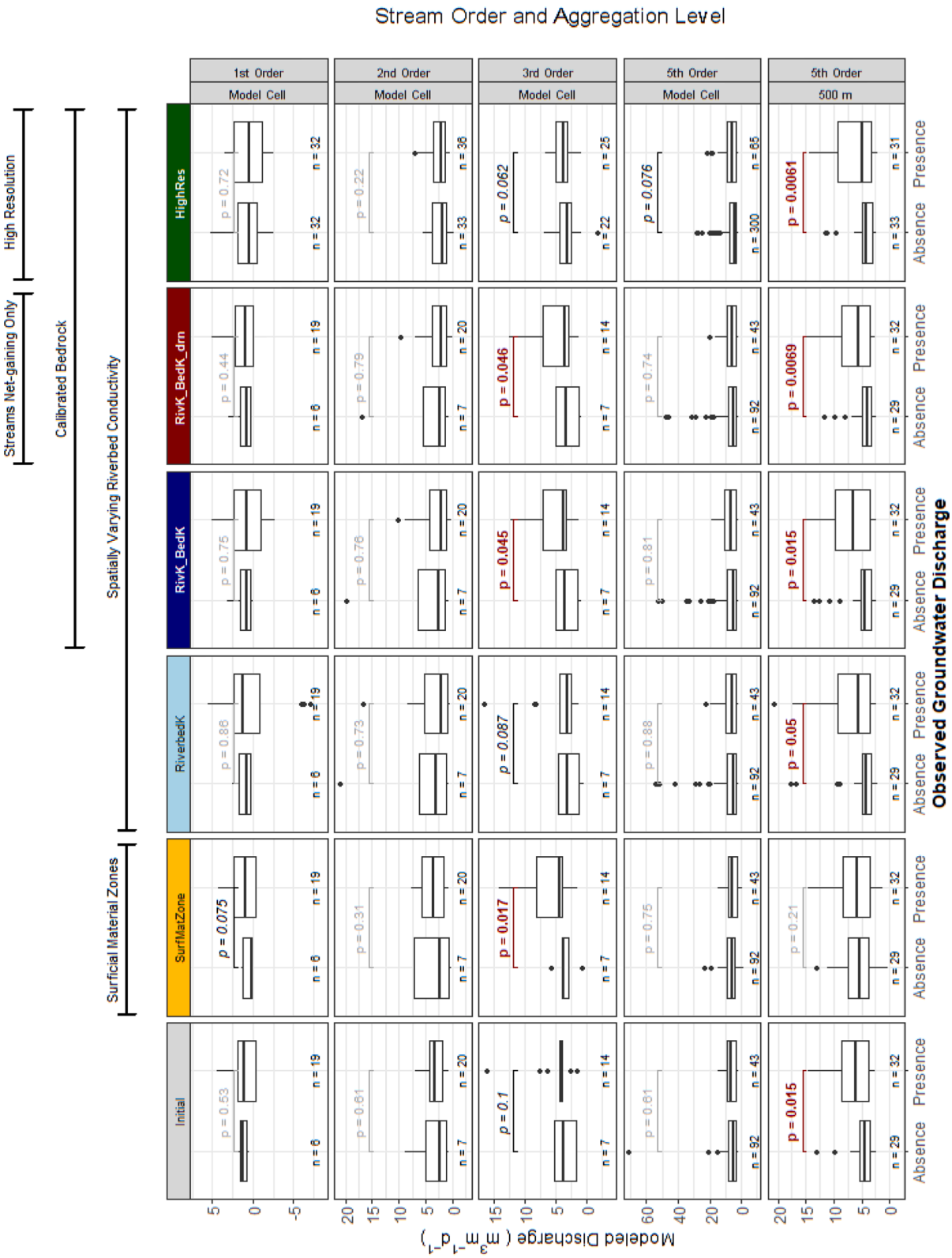


Fig. 6. Modeled discharge grouped by Model (columns), Stream Order (rows) and Aggregation (model cell or 500 m lengths) (rows), and the presence or absence of observed groundwater discharge (x axis) during thermal infrared field surveys. Numbers above the boxplots are the p-values for differences between the groups, with values less than or equal to 0.05 in bold red, values greater than 0.05 and less than or equal to 0.1 in italic black, and all other values in light gray. The non-gray colors in the column headers correspond to similar model characteristics and to the colors in Fig. 5. Simulated discharge is expressed as a volume of discharge per length of stream per day. In the boxplots, boxes depict the first quartile, median, and third quartile, whiskers extend no more than 1.5 times the interquartile range beyond the first and third quartile, and outliers beyond the whiskers are depicted as single points.

upstream end, consistent with the increased aerial extent of observed discharge at the downstream end.

In summary, across all stream orders, the spatial alignment between simulated and observed (river bank) discharge patterns varied. In some locations, such as km 12 in the 5th order river (Fig. 5D), peaks in simulated discharge align spatially with peaks in the spatial extent of observed discharge. In other locations, peaks in simulated discharge occur upstream or downstream of peaks in the spatial extent of observed discharge. For example, in site 14 of the 3rd order river (Fig. 5C), a peak in simulated discharge occurs slightly upstream of an extensive area of observed groundwater discharge. Misalignment in simulated and observed discharge locations is more likely due to lack of precision in the simulated location than to errors in the mapped location of the observed discharge. Observed discharge was aggregated to the model cell resolution (300 m or 150 m for the *HighRes* model), making it unlikely that observed discharge would be mapped to the incorrect model cell, except when the discharge occurred near the cell boundary. The groundwater models were based on relatively coarse datasets, with hydraulic conductivity for individual cells determined by interpolation. Together, this makes it more likely that simulated locations of high discharge will be less accurate than field mapped locations, though our field mapping technique is not expected to capture direct riverbed discharge, which may be substantial in some areas.

3.1.3. Discharge-location density and simulated groundwater discharge

In the 5th order survey, the observed discharge density and simulated discharge were correlated for one model when the results were aggregated to 500 m lengths (*HighRes,* rho = 0.26, $n = 61$; $p \leq 0.05$) and weakly correlated for five model-aggregation pairs (*RivK_BedK*, rho = 0.21, 500 m lengths, $n = 61$; *RivK_BedK_drn*, rho = 0.24, 500 m lengths, $n = 61$; *HighRes,*rho = 0.34, 1 km lengths, $n = 33$; *RivK_BedK,* $r = 0.46$, 1 km lengths, $n = 17$; *RivK_BedK_drn,*$r = 0.48$, 2 km lengths, $n = 17$; $p \leq 0.10$). In the 3rd order survey, the density of observed discharge locations (observations per length of stream) was strongly correlated with the simulated groundwater discharge rate for three models (*RivK_BedK,* $r = 0.73$; *RivK_BedK_drn,* $r = 0.71$; *SurfMatZone,* $r = 0.71$; $p \leq 0.05$) when the results were compared across sites / sub-sites (subdivisions of longer length surveys, shown with vertical bars on the x axis in Fig. 5C, $n = 8$) and weakly correlated for two models (*SurfMatZone*, rho=0.90, model cell resolution, $n = 22$; *RiverbedK,* $r = 0.67$, site / sub-site aggregation, $n = 8$; $p \leq 0.10$). In the 1st and 2nd order surveys, the observed discharge-location density was not significantly correlated with the simulated groundwater discharge.

4. Discussion

Accurate simulation of spatially explicit groundwater-discharge characteristics at the river-network scale requires an approach that links together models and empirical methods at comparable spatial resolutions. We demonstrate an approach to integrating empirical TIR observations of preferential groundwater-discharge occurrence and river-network scale groundwater-flow modeling. To our knowledge, this is one of the most spatially extensive high-resolution sets of observed groundwater-discharge locations that includes small streams to date, and represents a first attempt to use the spatial patterns of discharge locations to evaluate a river-network scale groundwater-flow model at the grid-cell scale.

4.1. Cross-scale integration of groundwater discharge: physical drivers, observations, and simulations

Simulation and assessment of groundwater discharge patterns at the river-network scale involves a complex cross-scale integration of physical drivers of groundwater discharge, empirical observation methods, and groundwater models. Although TIR provides spatially extensive observations at relevant spatial resolutions, our findings suggest that scale-integration challenges remain. In particular, we found that the river-network scale models worked well in the 3rd and 5th order river reaches where the stream channels cut through a thicker sand and gravel aquifer. Along the 2nd order reach, which also overlaid sand and gravel deposits, the models accurately predicted more discharge at the end of site 10 (Fig. 5B) where extensive seep faces were also found. Within 1st order stream reaches, where bedrock was often near-surface and there was prevalence of low-permeability glacial till, the models did not reflect observed variation in discharge.

4.1.1. The scale of physical drivers of groundwater discharge

Groundwater discharge is driven by three-dimensional patterns of hydraulic conductivity and hydraulic head (Freeze and Witherspoon, 1967), which are driven by a complex array of patterns in groundwater recharge, surficial material properties (type, layering, thickness, etc.), bedrock properties (type, layering, fracture patterns, etc.), and geomorphology (slope, sinuosity, confinement, etc.) (Haitjema and Mitchell-Bruker, 2005; Winter et al., 1998). These drivers operate at nested scales ranging from centimeters to tens of kilometers (Freeze and Witherspoon, 1967; Rosenberry and Pitlick, 2009). The importance of each scale varies by network position, with local drivers dominating in headwaters areas and regional drivers dominating in larger rivers, but even in larger rivers, fine scale patterns are important (Rosenberry and Pitlick, 2009).

4.1.2. The scale of TIR observations

TIR observation spatial resolution is a function of the temperature differential between groundwater and surface materials, as well as the survey speed and distance from the discharge points. By paddling down the center of the 5th order river, we were able to survey multiple kilometers a day, but faster travel at a longer distance from the banks resulted in a coarse-resolution survey; we noted the larger areas of groundwater discharge, but likely not every small point. In contrast, slower travel while wading along small narrow streams resulted in a finer level of detail in the smaller stream surveys compared to the large river surveys. Even where we paddled the 3rd order river, the narrower channel and lower water levels resulted in slower travel, and consequently, in a finer-resolution survey than in the 5th order river.

The observed temperature differentials also differed by stream size, likely due to different thermal regimes in the smaller versus larger rivers. In the 5th order river reaches during late summer and early fall, the temperature of the stream channel was consistently $> 20\ ^{\circ}C$ and the observed groundwater-discharge zones on the banks were $< 15\ ^{\circ}C$ (some below $11\ ^{\circ}C$); this large difference made the thermal contrast between the stream banks and groundwater discharge relatively obvious. In contrast, the channel temperatures in the 1st order streams tended to be lower, some as cold as $15\ ^{\circ}C$, with the result that the thermal contrast of groundwater-discharge along the stream banks were less apparent with TIR and the viable survey season was shorter (channel temperatures approached groundwater temperatures earlier in the fall and remained lower later in the spring). Cooler stream channels in the small streams are likely a combination of land cover (the small streams were predominantly in forested areas with closed canopies, whereas the paddled section had an open canopy and were bordered by agricultural and developed land), elevation, and overall groundwater dominance of streamflow in summer. The smaller temperature differentials, finer survey resolution, and shorter survey lengths in the small streams made it more difficult to distinguish river sections with a varying density of groundwater discharge areas as compared to the larger river.

4.1.3. The scale of simulated groundwater discharge

The resolution of simulated groundwater discharge is a function of the model resolution, the resolution of the input data, and the available calibration data. Most input data for river-network scale groundwater models is of a coarser resolution than the model cell. For example, important drivers of groundwater discharge including the thickness and

permeability of surficial materials and bedrock, as well as groundwater recharge rates, are generally known only at a resolution much coarser than the model grid. In addition, local scale heterogeneity in calibrated hydraulic conductivity patterns is constrained by both the spacing of the pilot points and the available calibration data. Increasing the density of pilot points without a corresponding density of observation data, however, will not increase heterogeneity in calibrated values because, in the absence of data, hydraulic conductivity values will reflect default values.

4.1.4. Integration insights

In the 3rd and 5th order rivers, the resolution of the model, groundwater-discharge drivers, and survey method matched well. The model resolution is relatively coarse (300 m or 100 m) and is expected to reflect the general, coarse scale patterns. The 3rd and 5th order rivers traverse a large area of relatively high hydraulic conductivity sediments (coarse and fine-grained stratified sediments), creating large areas of groundwater discharge that were easily observed and modeled. In the 5th order river, we were able to paddle extensive distances (3 −5 km per day), and the combination of longer survey lengths and larger discharges allowed us to identify spatial patterns of groundwater discharge at a similar (coarse) resolution as the model resolution. Likewise, the 3rd order survey lengths were long enough to distinguish between areas of higher and lower discharge-location density. In both the 3rd and 5th order rivers, the better performing models distinguished gradients of observed discharge-location density, not only the difference between areas with and without observed discharge locations.

In contrast, within the 1st and 2nd order stream sites the resolution of the model, groundwater-discharge drivers, and survey method did not match. The model resolution was the same (300 m or 100 m, relatively coarse) as in the larger river. In 1st and 2nd order streams discharge is more strongly controlled by local-scale topography, preferential flow paths, and surficial-material patterns than by regional features. It is likely that the resulting spatial patterns of groundwater discharge were below the resolution of the coarse-scale model, even when we reduced the resolution to 100 m. For example, the complex near-surface geology that is common in headwaters areas, such as near-surface (or scoured to) bedrock dynamics and the prevalence of low-permeability till that inhibits discharge even where predicted based on water table elevations, is not well represented in the groundwater models. In addition, the survey resolution was fine, which may have matched well with the resolution of the drivers of groundwater discharge, but not with the groundwater-flow model.

4.2. Using TIR at the river-network scale for groundwater model evaluation

Use of TIR at the river-network scale requires multiple survey methods. River size and morphology determine how each stream or river can be accessed and surveyed. In this study, wading would have been impractical or unsafe in the larger river, and paddling impossible in the smaller streams. We were unable to survey high-gradient or high-velocity streams, due to practical and safety considerations. The use of airborne TIR (on unmanned aerial systems) could allow surveys in these otherwise inaccessible reaches (Harvey et al., 2019), but cannot be used in small forested settings with a closed canopy and likely would result in coarser resolution than our paddled surveys due to faster travel compared to paddling.

Different TIR survey methods, as demonstrated in this study, have differing spatial resolutions that must be acknowledged and addressed in investigations across stream and river sizes. One aspect of addressing the differing spatial resolutions could be analyzing data from different TIR survey types separately, as well as developing consistent survey metrics to classify observed discharge locations. It is also important to ensure sufficient survey lengths to adequately distinguish between river lengths with higher and lower densities of groundwater-discharge zones. We found this challenging in the small streams, but easy in the larger river. Finally, it will be important to consider the optimum comparison metric for TIR and simulated discharge. We used two metrics: 1) a binary presence / absence metric, which is simple, but does not capture variation in observed discharge, and 2) a correlation of observed discharge-location density and simulated groundwater discharge. Particularly in the small streams, where wading results in high resolution surveys, metrics that capture the extent of discharge, and not simply the presence / absence, are needed. Other possibilities include a qualitative metric based on site observations (e.g. presence of visible flow or channels in the sediment).

Simulated discharge from groundwater models includes both discharge above the waterline and submerged seeps. Another possible limitation of TIR is that it assumes the majority of the discharge occurs above the waterline (i.e. bank seeps), so submerged discharge zones in deeper, faster water may be missed. In river systems, however, discharge is typically weighted toward the water line due to expected highest hydraulic gradients and generally decreases with distance from shore exponentially in homogeneous sediments (Anibas et al., 2011). Also, fines tend to accumulate in deeper water forming a low-K “cap” that inhibits discharge, while fines are actively scoured by wave action along the water line (Rosenberry et al., 2015). We conducted one survey during an extensive reservoir draw down that exposed the majority of the streambed and the seeps we observed were at the pre-draw-down water line. Together these suggest that submerged discharge zones are a relatively minor component of the total groundwater discharge and will not substantially limit the use of TIR for mapping areas of groundwater discharge or evaluating groundwater-flow models in systems similar to our study watershed.

Across all stream sizes, the models predicted groundwater discharge in most model cells with rivers, including where we did not observe preferential bank discharge zones with TIR (Fig. 5). The presence of modeled discharge in areas without TIR-observed discharge zones could be a result of both the limitations of TIR imaging and model deficiencies, which we cannot distinguish here. Higher predicted rates in areas without visible preferential groundwater discharge could reflect widespread spatially diffuse discharge and/or discharge below the water surface that is not adequately mapped with TIR imaging, since it is based on strong thermal anomalies and does not penetrate the water surface. Conversely, the higher rates could indicate areas where the models do not accurately represent the processes driving groundwater discharge. For example, areas with greater local-scale heterogeneity in hydraulic conductivity values and variable depth-to-bedrock may not be well modeled.

4.3. Implications for modeling groundwater discharge at the river-network scale

We found that a common approach to modeling groundwater successfully predicted broad patterns of groundwater discharge, but failed to predict local patterns within headwaters sites. In the 3rd and 5th order rivers, discharge patterns were robust across model variations. Within 1st order streams, only 1 model (*SurfMatZone*) weakly distinguished between areas with and without observed discharge locations, and no model predicted the observed patterns in the 2nd order stream.

This work suggests several important implications for improved modeling of groundwater discharge at the river-network scale. First, modeling efforts should focus on improving predicted discharge patterns in smaller reaches, particularly in areas of glacial till and near-surface bedrock. This might involve finer-resolution data on surficial-material type and thickness, calibrating by zones based on stream order (as was the case for the *SurfMatZone* model), or better representing the bi-directional nature of river-aquifer exchange in headwaters areas. Geophysical methods could be used to better map the thickness and character of sediments in stream corridors (Auken et al., 2017; McLachlan et al., 2017). Second, bedrock hydraulic conductivity values

should be calibrated, not based on established literature values. Although this increased the computational requirements of calibration and requires data on groundwater heads in bedrock, we found that it improved the accuracy of predicted discharge patterns. Third, careful attention should be paid to the conceptualization of river-aquifer exchange. This is particularly true in the smaller streams, where bi-directional exchange between groundwater and surface water can be particularly complex (Payn et al., 2009). Varying the magnitude of the riverbed hydraulic conductivity improved the model fit in some instances, but created unreasonably large fluxes from river to aquifer in other locations. Eliminating losing reaches (limiting river-aquifer exchange to discharge) improved the model fit in some cases, but prevents accurate representation of losing reaches.

4.4. Next steps

This work takes an important step in evaluating spatial patterns of groundwater discharge at the river-network scale and demonstrating the viability of comparing groundwater-flow model predictions to observed groundwater-discharge zone spatial clustering. Improving this approach will involve more spatially expansive field surveys that better cover the extensive headwater stream lengths. For example, at most of the waded (non-boatable) sites, the survey length was less than 800 m; longer surveys in these sites may have highlighted more within-site variation. In this project we were limited to stream reaches that were accessible through wading or paddling. Recent advances in remote sensing of groundwater discharge using TIR and unmanned aerial surveys offer a potential methodology to survey reaches through steep or wetland areas that were inaccessible in our work (Briggs et al., 2019; Harvey et al., 2019), though forested headwater streams will remain challenging for aerial surveys. Extensive spatial surveys have the potential to be used as a numerical calibration target, if a method is developed to convert predicted discharge rates to a probability of observing discharge in the field, possibly incorporating channel and landscape geomorphology (e.g. Dugdale et al., 2015).

In addition, although TIR provides an easy method of identifying groundwater-discharge zones, it does not enable quantification of discharge rates or identification of losing reaches. Using TIR in combination with a toolbox of quantification approaches (flow tracers, weirs, vertical temperature profilers, differential gaging, etc.) may provide the best evaluation of groundwater discharge patterns. For this study we compared model results to general spatial patterns of observed groundwater-discharge zones. The use of additional field methodology beyond TIR will also allow comparison to discharge rate, presuming field-measured fluxes can be aggregated at the model-grid-cell scale.

5. Conclusions

We demonstrate an approach to using TIR to assess model-cell scale groundwater discharge patterns predicted by river-network scale groundwater-flow models. The spatial patterns of discharge identified by TIR provided a way to differentiate between the performance of the models and to evaluate the modeled patterns. In the larger (3rd and 5th order) rivers, spatial patterns of discharge were relatively well represented by most models, particularly those where the bedrock hydraulic conductivity was calibrated instead of based on literature values. In smaller streams (1st and 2nd order), most models did not simulate the observed groundwater discharge patterns. More spatially extensive surveys, including remotely-sensed TIR mapping, combined with a toolbox of discharge quantification methods, can further refine this work and provide a spatially distributed groundwater discharge target for numerical calibration. In addition, new approaches are needed for simulating groundwater discharge in headwater streams, and may necessitate refined geologic data.

CRediT authorship contribution statement

J.R. Barclay: Conceptualization, Methodology, Investigation, Data curation, Formal analysis, Writing – original draft, Visualization. **M.A. Briggs:** Conceptualization, Methodology, Investigation, Writing – review & editing. **E.M. Moore:** Investigation, Data curation. **J.J. Starn:** Conceptualization, Methodology, Software. **A.E.H. Hanson:** Investigation. **A.M. Helton:** Conceptualization, Methodology, Investigation, Writing – review & editing, Supervision, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments and Data

We would like to thank Katie Skalak, USGS, and the anonymous reviewers who gave helpful feedback on an early version of this manuscript. We would also like to thank Adam Haynes, Kevin Jackson, and numerous students at the University of Connecticut and interns at the USGS for assistance in the field surveys.

This research was supported by Agriculture and Food Research Initiative Competitive Grant no. 2014–38420–21802 from the USDA National Institute of Food and Agriculture; National Science Foundation Division of Earth Sciences award 1824820; USDA National Institute of Food and Agriculture Hatch projects CONS00938 and 1016783; Connecticut Institute for Water Resources 104b grant 2017CT306B; the US Geological Survey Toxic Substances Hydrology Program; the US Geological Survey 2017 Cooperative Summer Training Program; a Geological Society of America Graduate Student Research Grant (2017) and Student Travel Grants (2016 and 2018); three pre-doctoral fellowships from the University of Connecticut (2016, 2017, and 2018), and a scholarship from the New England Outdoor Writers Association.

TIR images and model results are available in Barclay et al. (2019) (https://doi.org/10.5066/P9EIV8L5), Moore et al. (2020) (https://doi.org/10.5066/P915E8JY) and Barclay et al. (2020b) (https://doi.org/10.5066/P960RSKM).

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

References

Ala-aho, P., Rossi, P.M., Isokangas, E., Kløve, B., 2015. Fully integrated surface–subsurface flow modelling of groundwater–lake interaction in an esker aquifer: model verification with stable isotopes and airborne thermal imaging. J. Hydrol. 522, 391–406. https://doi.org/10.1016/j.jhydrol.2014.12.054.

Anibas, C., Buis, K., Verhoeven, R., Meire, P., Batelaan, O., 2011. A simple thermal mapping method for seasonal spatial patterns of groundwater–surface water interaction. J. Hydrol. 397, 93–104. https://doi.org/10.1016/j.jhydrol.2010.11.036.

Auken, E., Boesen, T., Christiansen, A.V., 2017. Chapter Two - A review of airborne electromagnetic methods with focus on geotechnical and hydrological applications from 2007 to 2017. In: Nielsen, L. (Ed.), Advances in Geophysics. Elsevier, pp. 47–93. https://doi.org/10.1016/bs.agph.2017.10.002.

Baker, E.A., Lautz, L.K., McKenzie, J.M., Aubry-Wake, C., 2019. Improving the accuracy of time-lapse thermal infrared imaging for hydrologic applications. J. Hydrol. 571, 60–70. https://doi.org/10.1016/j.jhydrol.2019.01.053.

Barclay, J.R., Hanson, A.E.H., Briggs, Martin A., Helton, A.M., 2019. Thermal Infrared Images and Direct Temperature Measurements of Groundwater Discharge Zones Throughout the Farmington River Watershed (Connecticut and Massachusetts). U.S. Geological Survey Data Release. https://doi.org/10.5066/P9EIV8L5.

Barclay, J.R., Starn, J.J., Briggs, M.A., Helton, A.M., 2020a. Improved prediction of management-relevant groundwater discharge characteristics throughout river networks. Water Resour. Res. https://doi.org/10.1029/2020WR028027, 56, e2020WR028027.

Barclay, J.R., Starn, J.J., Briggs, M.A., Helton, A.M., 2020b. MODFLOW-NWT and MODPATH groundwater flow models of the Farmington River Watershed. U.S. Geological Survey data release, (Connecticut and Massachusetts). https://doi.org/10.5066/P960RSKM.

Barlow, P.M., Leake, S.A., 2012. Streamflow Depletion By Wells–Understanding and Managing the Effects of Groundwater Pumping On Streamflow (Circular 1376). U.S. Geological Survey, Reston, VA. https://doi.org/10.3133/cir1376.
Barnes, R.T., Butman, D.E., Wilson, H.F., Raymond, P.A., 2018. Riverine export of aged carbon driven by flow path depth and residence time. Environ. Sci. Technol. 52, 1028–1035. https://doi.org/10.1021/acs.est.7b04717.
Befus, K.M., Kroeger, K.D., Sampson, D.W., Schwab, W.C., 2017. Continuous and Optimized 3-Arcsecond Elevation Model for United States East and West Coasts. U.S. Geological Survey Data Release. https://dx.doi.org/10.5066/F7W37TGK.
Briggs, M.A., Dawson, C.B., Holmquist-Johnson, C.L., Williams, K.H., Lane, J.W., 2019. Efficient hydrogeological characterization of remote stream corridors using drones. Hydrol. Process 33, 316–319. https://doi.org/10.1002/hyp.13332.
Briggs, M.A., Hare, D.K., 2018. Explicit consideration of preferential groundwater discharges as surface water ecosystem control points. Hydrol. Process 32, 2435–2440. https://doi.org/10.1002/hyp.13178.
Briggs, M.A., Hare, D.K., Boutt, D.F., Davenport, G., Lane, J.W., 2016. Thermal infrared video details multiscale groundwater discharge to surface water through macropores and peat pipes. Hydrol. Process. 30, 2510–2511. https://doi.org/10.1002/hyp.10722.
Briggs, M.A., Lane, J.W., Snyder, C.D., White, E.A., Johnson, Z.C., Nelms, D.L., Hitt, N.P., 2018. Shallow bedrock limits groundwater seepage-based headwater climate refugia. Limnologica 68, 142–156. https://doi.org/10.1016/j.limno.2017.02.005.
Caissie, D., Luce, C.H., 2017. Quantifying streambed advection and conduction heat fluxes. Water Resour. Res. 53, 1595–1624. https://doi.org/10.1002/2016WR019813.
Cartwright, I., Miller, M.P., 2021. Temporal and spatial variations in river specific conductivity: implications for understanding sources of river water and hydrograph separations. J. Hydrol. 593, 125895 https://doi.org/10.1016/j.jhydrol.2020.125895.
Chen, D., Shen, H., Hu, M., Wang, J., Zhang, Y., Dahlgren, R.A., 2018. Chapter Five - Legacy nutrient dynamics at the watershed scale: principles, modeling, and implications, in: Sparks, D.L. (Ed.), Advances in Agronomy. Academic Press, pp. 237–313. https://doi.org/10.1016/bs.agron.2018.01.005.
Danielescu, S., MacQuarrie, K.T.B., Faux, R.N., 2009. The integration of thermal infrared imaging, discharge measurements and numerical simulation to quantify the relative contributions of freshwater inflows to small estuaries in Atlantic Canada. Hydrol. Process 23, 2847–2859. https://doi.org/10.1002/hyp.7383.
Dent, C.L., Grimm, N.B., Fisher, S.G., 2001. Multiscale effects of surface–subsurface exchange on stream water nutrient concentrations. J. North Am. Benthol. Soc. 20, 162–181. https://doi.org/10.2307/1468313.
Domenico, P., Schwartz, F., 1997. Physical and Chemical Hydrogeology, 2nd Edition.
Dugdale, S.J., Bergeron, N.E., St-Hilaire, A., 2015. Spatial distribution of thermal refuges analysed in relation to riverscape hydromorphology using airborne thermal infrared imagery. Remote Sens. Environ. 160, 43–55. https://doi.org/10.1016/j.rse.2014.12.021.
Feinstein, D.T., Hunt, R.J., Reeves, H.W., 2010. Regional Groundwater-Flow Model of the Lake Michigan Basin in Support of Great Lakes Basin Water Availability and Use Studies (Report No. 2010–5109). U.S. Geological Survey Scientific Investigations Report, Reston, VA. https://doi.org/10.3133/sir20105109 (Report No. 2010–5109), Scientific Investigations ReportURL: http://pubs.er.usgs.gov/publication/sir20105109.
Freeze, R.A., Witherspoon, P.A., 1967. Theoretical analysis of regional groundwater flow: 2. Effect of water-table configuration and subsurface permeability variation. Water Resour. Res. 3, 623–634. https://doi.org/10.1029/WR003i002p00623.
Glaser, B., Klaus, J., Frei, S., Frentress, J., Pfister, L., Hopp, L., 2016. On the value of surface saturated area dynamics mapped with thermal infrared imagery for modeling the hillslope-riparian-stream continuum. Water Resour. Res. 52, 8317–8342. https://doi.org/10.1002/2015WR018414.
Haitjema, H.M., Mitchell-Bruker, S., 2005. Are water tables a subdued replica of the topography? Groundwater 43, 781–786. https://doi.org/10.1111/j.1745-6584.2005.00090.x.
Handcock, R.N., Torgersen, C.E., Cherkauer, K.A., Gillespie, A.R., Tockner, K., Faux, R.N., Tan, J., 2012. Thermal infrared remote sensing of water temperature in riverine landscapes. Fluvial Remote Sensing For Science and Management. John Wiley & Sons, Ltd, pp. 85–113. https://doi.org/10.1002/9781119940791.ch5.
Harbaugh, A.W., 2005. MODFLOW-2005, The U.S. Geological Survey Modular Ground-Water Model—The Ground-Water Flow Process. U.S. Geological Survey Techniques and Methods 6-A16. https://doi.org/10.3133/tm6A16.
Hare, D.K., Briggs, M.A., Rosenberry, D.O., Boutt, D.F., Lane, J.W., 2015. A comparison of thermal infrared to fiber-optic distributed temperature sensing for evaluation of groundwater discharge to surface water. J. Hydrol. 530, 153–166. https://doi.org/10.1016/j.jhydrol.2015.09.059.
Hare, D.K., Helton, A.M., Johnson, Z.C., Lane, J.W., Briggs, M.A., 2021. Continental-scale analysis of shallow and deep groundwater contributions to streams. Nat. Commun. 12, 1450. https://doi.org/10.1038/s41467-021-21651-0.
Harvey, M., Hare, D., Hackman, A., Davenport, G., Haynes, A., Helton, A., Lane, W.J., Briggs, A.M., 2019. Evaluation of stream and wetland restoration using UAS-based thermal infrared mapping. Water 11. https://doi.org/10.3390/w11081568.
Herberich, E., Sikorski, J., Hothorn, T., 2010. A robust procedure for comparing multiple means under heteroscedasticity in unbalanced designs. PLS ONE 5, e9788. https://doi.org/10.1371/journal.pone.0009788.
Homer, C.G., Dewitz, J.A., Yang, L., Jin, S., Danielson, P., Xian, G., Coulston, J., Herold, N.D., Wickham, J., Megown, K., 2015. Completion of the 2011 National Land Cover Database for the conterminous United States-Representing a decade of land cover change information. Photogramm. Eng. Remote Sens. 81, 345–354.
Irvine, D.J., Briggs, M.A., Cartwright, I., Scruggs, C.R., Lautz, L.K., 2016. Improved vertical streambed flux estimation using multiple diurnal temperature methods in series. Groundwater 53, 73–80. https://doi.org/10.1111/gwat.12436 n/a-n/a.
Jeannot, B., Weill, S., Eschbach, D., Schmitt, L., Delay, F., 2019. Assessing the effect of flood restoration on surface–subsurface interactions in Rohrschollen Island (Upper Rhine river – France) using integrated hydrological modeling and thermal infrared imaging. Hydrol. Earth Syst. Sci. 23, 239–254. https://doi.org/10.5194/hess-23-239-2019.
Johnson, Z.C., Johnson, B.G., Briggs, M.A., Devine, W.D., Snyder, C.D., Hitt, N.P., Hare, D.K., Minkova, T.V., 2020. Paired air-water annual temperature patterns reveal hydrogeological controls on stream thermal regimes at watershed to continental scales. J. Hydrol. 587, 124929 https://doi.org/10.1016/j.jhydrol.2020.124929.
Kalbus, E., Reinstorf, F., Schirmer, M., 2006. Measuring methods for groundwater ? Surface water interactions: a review. Hydrol. Earth Syst. Sci. Discussions 10, 873–887.
Kilpatrick, F.A., Cobb, E.D., 1985. Measurement of Discharge Using Tracers (Report No. 03-A16), Techniques of Water-Resources Investigations. U.S. Geological Survey, Reston, VA. https://doi.org/10.3133/twri03A16.
Kolbe, T., de Dreuzy, J.-.R., Abbott, B.W., Aquilina, L., Babey, T., Green, C.T., Fleckenstein, J.H., Labasque, T., Laverman, A.M., Marçais, J., Peiffer, S., Thomas, Z., Pinay, G., 2019. Stratification of reactivity determines nitrate removal in groundwater. Proc Natl. Acad. Sci. USA, 201816892. https://doi.org/10.1073/pnas.1816892116.
Kurylyk, B.L., MacQuarrie, K.T.B., Caissie, D., McKenzie, J.M., 2015. Shallow groundwater thermal sensitivity to climate change and land cover disturbances: derivation of analytical expressions and implications for stream temperature modeling. Hydrol. Earth Syst. Sci. 19, 2469–2489. https://doi.org/10.5194/hess-19-2469-2015.
Lane, J.W., Briggs, M.A., Maurya, P.K., White, E.A., Pedersen, J.B., Auken, E., Terry, N., Minsley, B., Kress, W., LeBlanc, D.R., Adams, R., Johnson, C.D., 2020. Characterizing the diverse hydrogeology underlying rivers and estuaries using new floating transient electromagnetic methodology. Sci. Total Environ. 740, 140074 https://doi.org/10.1016/j.scitotenv.2020.140074.
Masterson, J.P., Pope, J.P., Fienen, M.N., Monti Jr., J., Nardi, M.R., Finkelstein, J.S., 2016. Documentation of a Groundwater Flow Model Developed to Assess Groundwater Availability in the Northern Atlantic Coastal Plain Aquifer System From Long Island, New York, to North Carolina (Report No. 2016–5076). U.S. Geological Survey, Scientific Investigations Report, Reston, VA. https://doi.org/10.3133/sir20165076.
Matheswaran, K., Blemmer, M., Rosbjerg, D., Boegh, E., 2014. Seasonal variations in groundwater upwelling zones in a Danish lowland stream analyzed using Distributed Temperature Sensing (DTS). Hydrol. Process 28, 1422–1435. https://doi.org/10.1002/hyp.9690.
McCallum, J.L., Cook, P.G., Berhane, D., Rumpf, C., McMahon, G.A., 2012. Quantifying groundwater flows to streams using differential flow gaugings and water chemistry. J. Hydrol. 416–417, 118–132. https://doi.org/10.1016/j.jhydrol.2011.11.040.
McLachlan, P.J., Chambers, J.E., Uhlemann, S.S., Binley, A., 2017. Geophysical characterisation of the groundwater–surface water interface. Adv. Water Resour. 109, 302–319. https://doi.org/10.1016/j.advwatres.2017.09.016.
Miller, M.P., Tesoriero, A.J., Hood, K., Terziotti, S., Wolock, D.M., 2017. Estimating discharge and nonpoint source nitrate loading to streams from three end-member pathways using high-frequency water quality data. Water Resour. Res. 53, 10201–10216. https://doi.org/10.1002/2017WR021654.
Moore, E.M., Jackson, K.J., Haynes, A.B., Helton, A.M., Briggs, M.A., 2020. Thermal Infrared Images of Groundwater Discharge Zones in the Farmington and Housatonic River Watersheds. Connecticut and Massachusetts, 2019U.S. Geological Survey data release.
Mundy, E., Gleeson, T., Roberts, M., Baraer, M., McKenzie, J.M., 2017. Thermal imagery of groundwater seeps: possibilities and limitations. Groundwater 55, 160–170. https://doi.org/10.1111/gwat.12451.
Niswonger, R.G., Panday, S., Ibaraki, M., 2011. MODFLOW-NWT, A Newton Formulation for MODFLOW-2005 (Report No. 6-A37), Techniques and Methods. U.S. Geological Survey, Reston, VA. https://doi.org/10.3133/tm6A37. URL: http://pubs.er.usgs.gov/publication/tm6A37.
Olcott, P.G., 1995. Connecticut, Maine, Massachusetts, New Hampshire, New York, Rhode Island, Vermont, HA 730-M. Groundwater Atlas of the US. United States Geological Survey. U.S. Geological Survey, Reston, VA. http://pubs.usgs.gov/ha/ha730.
Pandey, P., Gleeson, T., Baraer, M., 2013. Toward quantifying discrete groundwater discharge from frozen seepage faces using thermal infrared images. Geophys. Res. Lett. 40, 123–127. https://doi.org/10.1029/2012GL054315.
Payn, R.A., Gooseff, M.N., McGlynn, B.L., Bencala, K.E., Wondzell, S.M., 2009. Channel water balance and exchange with subsurface flow along a mountain headwater stream in Montana, United States. Water Resour. Res. 45 https://doi.org/10.1029/2008WR007644.
PRISM Climate Group, 2012. 30yr mean temperature (PRISM_tmean_30yr_normal_800mM2_annual_bil.bil). Oregon State University. http://prism.oregonstate.edu.
R Core Team, 2019. R: a Language and Environment For Statistical Computing. R Foundation for Statistical Computing. http://www.R-project.org.
Rosenberry, D.O., 2008. A seepage meter designed for use in flowing water. J. Hydrol. 359, 118–130. https://doi.org/10.1016/j.jhydrol.2008.06.029.
Rosenberry, D.O., Lewandowski, J., Meinikmann, K., Nützmann, G., 2015. Groundwater - the disregarded component in lake water and nutrient budgets. Part 1: effects of groundwater on hydrology. Hydrol. Process. 29, 2895–2921. https://doi.org/10.1002/hyp.10403.

Rosenberry, D.O., Pitlick, J., 2009. Local-scale variability of seepage and hydraulic conductivity in a shallow gravel-bed river. Hydrol. Process. 23, 3306–3318. https://doi.org/10.1002/hyp.7433.

Sanford, W.E., 2011. Calibration of models using groundwater age. Hydrogeol. J. 19, 13–16. https://doi.org/10.1007/s10040-010-0637-6.

Sanford, W.E., Pope, J.P., Selnick, D.L., Stumvoll, R.F., 2012. Simulation of Groundwater Flow in the Shallow Aquifer System of the Delmarva Peninsula, Maryland and Delaware (Open-File Report 2012–1140. U.S. Geological Survey, Reston, VA. https://doi.org/10.3133/ofr20121140. https://doi.org/10.3133/ofr20121140.

Schuetz, T., Weiler, M., 2011. Quantification of localized groundwater inflow into streams using ground-based infrared thermography. Geophys. Res. Lett. 38 https://doi.org/10.1029/2010GL046198.

Soller, D.R., Packard, P.H., Garrity, C.P., 2012. Database for USGS Map I-1970 - Map Showing the Thickness and Character of Quaternary Sediments in the Glaciated United States east of the Rocky Mountains (Report No. 656). U.S. Geological Survey Data Series. https://doi.org/10.3133/ds656.

Starn, J.J., Belitz, K., 2018. Regionalization of groundwater residence time using metamodeling. Water Resour. Res. 54 https://doi.org/10.1029/2017WR021531.

Starn, J.J., Brown, C.J., 2007. Simulations of Ground-Water Flow and Residence Time near Woodbury, Connecticut (Report No. 2007–5210). U.S. Geological Survey Scientific Investigations Report, Reston, VA. https://doi.org/10.3133/sir20075210. https://doi.org/10.3133/sir20075210.

Toran, L., Nyquist, J., Rosenberry, D., Gagliano, M., Mitchell, N., Mikochik, J., 2015. Geophysical and hydrologic studies of lake seepage variability. Groundwater 53, 841–850. https://doi.org/10.1111/gwat.12309.

Torgersen, C.E., Ebersole, J.L., Keenan, D.M., 2012. Primer for identifying cold-water refuges to protect and restore thermal diversity in riverine landscapes. Environmental Protection Agency EPA.

Torgersen, C.E., Faux, R.N., McIntosh, B.A., Poage, N.J., Norton, D.J., 2001. Airborne thermal remote sensing for water temperature assessment in rivers and streams. Remote Sens. Environ. 76, 386–398. https://doi.org/10.1016/S0034-4257(01)00186-9.

U.S. Geological Survey, 2020. National Water Information System Data Available on the World Wide Web (Water Data For the Nation) [WWW Document]. Water Data For the Nation. U.S. Geological Survey, Reston, VA. URL. http://waterdata.usgs.gov/nwis/ (accessed 2.7.20).

U.S. Geological Survey, 2007. Geologic Maps of US States [WWW Document]. Mineral Resources On-Line Spatial DataURL. http://mrdata.usgs.gov/geology/state/ (accessed 4.15.15).

Vaccaro, J.J., 2011. River-Aquifer Exchanges in the Yakima River Basin, Washington (Report No. 2011–5026). U.S. Geological Survey Scientific Investigations Report, Reston, VA. https://doi.org/10.3133/sir20115026.

Welter, D.E., White, J.T., Hunt, R.J., Doherty, J.E., 2015. Approaches in Highly Parameterized Inversion—PEST++ Version 3, A PARAmeter ESTimation and Uncertainty Analysis Software Suite Optimized for Large Environmental Models (Report No. 7-C12). U.S. Geological Survey Techniques and Methods, Reston, VA. https://doi.org/10.3133/tm7C12. https://doi.org/10.3133/tm7C12.

Wherry, S.A., Tesoriero, A.J., Terziotti, S., 2021. Factors affecting nitrate concentrations in stream base flow. Environ. Sci. Technol. 55, 902–911. https://doi.org/10.1021/acs.est.0c02495.

Winter, T.C., Harvey, J.W., Franke, O.L., Alley, W.M., 1998. Ground Water and Surface Water; A Single Resource (Report No. 1139). U.S. Geological Survey Circular, Reston, VA. https://doi.org/10.3133/cir1139.

Xie, Y., Cook, P.G., Shanafield, M., Simmons, C.T., Zheng, C., 2016. Uncertainty of natural tracer methods for quantifying river–aquifer interaction in a large river. J. Hydrol. 535, 135–147. https://doi.org/10.1016/j.jhydrol.2016.01.071.

Yager, R.M., Kappel, W.M., Plummer, N., 2007. Halite Brine in the Onondaga Trough near Syracuse, New York: Characterization and Simulation of Variable-Density Flow (Report No. 2007–5058). U.S. Geological Survey Scientific Investigations Report, Reston, VA. https://doi.org/10.3133/sir20075058.

Yager, R.M., Southworth, S.C., Voss, C.I., 2008. Simulation of Ground-Water Flow in the Shenandoah Valley, Virginia and West Virginia, Using Variable-Direction Anisotropy in Hydraulic Conductivity to Represent Bedrock Structure (Report No. 2008–5002). U.S. Geological Survey Scientific Investigations Report, Reston, VA. https://doi.org/10.3133/sir20085002.