Received: 30 September 2021 | Revised: 31 July 2022 | Accepted: 21 September 2022

DOI: 10.1002/hyp.14713

SPECIAL ISSUE PAPER

WILEY

Identification of hyporheic extent and functional zonation during seasonal streamflow recession by unsupervised clustering of time-lapse electrical resistivity models

Joel G. Singley[1,2,3] | Kamini Singha[3] | Michael N. Gooseff[2,4] | Ricardo González-Pinzón[5] | Timothy P. Covino[6] | Adam S. Ward[7,8] | Jancoba Dorley[5] | Eve-Lyn S. Hinckley[1,2]

[1]Department of Environmental Studies, University of Colorado, Boulder, Colorado, USA

[2]Institute of Arctic and Alpine Research, Boulder, Colorado, USA

[3]Geology and Geological Engineering, Colorado School of Mines, Golden, Colorado, USA

[4]Civil, Environmental and Architectural Engineering, University of Colorado, Boulder, Colorado, USA

[5]Civil, Construction and Environmental Engineering, University of New Mexico, Albuquerque, New Mexico, USA

[6]Ecosystem Science and Sustainability, Colorado State University, Fort Collins, Colorado, USA

[7]O'Neill School of Public and Environmental Affairs, Indiana University, Bloomington, Indiana, USA

[8]Biological and Ecological Engineering, Oregon State University, Corvallis, Oregon, USA

Correspondence
Joel G. Singley, Department of Biology, Marine Biology, and Environmental Science, Roger Williams University, Bristol, RI, USA.
Email: jsingley@rwu.edu

Present address
Joel G. Singley, Department of Biology, Marine Biology, and Environmental Science, Roger Williams University, Bristol, Rhode Island, USA.

Funding information
National Science Foundation, Grant/Award Numbers: EAR-1642403, EAR-1642402; University of Colorado Boulder; Beverly Sears Graduate Research Grant

Abstract

Hyporheic exchange influences hydrologic transport and water quality through transient storage, which extends solute transit time, and leads to mixing of surface water and groundwater. Despite its importance, estimating the extent and spatiotemporal variability of the hyporheic zone remains challenging due to limitations in assessing the subsurface with discrete point-scale sampling. Analysis of time-lapse electrical resistivity (ER) data from tracer studies has shown potential to ameliorate such limitations. However, its utility in objectively delimiting hyporheic extent and quantifying changes in surface-groundwater exchange has been impeded by reliance on qualitative analysis of hyporheic extent or the use of a priori assumptions about data quality and signal strength. This study applies a novel unsupervised clustering method to time-lapse ER models derived from a benchmark dataset collected throughout baseflow recession in a mountain stream. We demonstrate that unsupervised clustering of inverted ER model time series can delimit hyporheic extent by distinguishing solute transport signals from noisy background inversions and identify functional zones defined by unique transport characteristics. We found that the structure of these zones was stable even as discharge changed by an order of magnitude, likely due to morphological constraints in this steep, narrow valley. Compared to traditional methods utilizing a priori thresholds to delimit hyporheic extent, clustering is robust to unintentional variations in tracer breakthrough curves that are typical of field-based studies. Therefore, clustering of inverted ER models represents a more robust and data-driven functional zonation representation of hyporheic exchange than has been possible with point-scale sampling or transport modelling, which usually assumes a single well-mixed hyporheic zone.

KEYWORDS
electrical resistivity imaging, H. J. Andrews Experimental Forest, hyporheic exchange, hyporheic zone

1 | INTRODUCTION

The exchange and mixing of surface and groundwater in stream corridors exert a strong control on hydrologic transport, biogeochemical reactions, and the existence of ecological refugia (Harvey et al., 2018; Harvey & Bencala, 1993; Lewandowski et al., 2019; Ward, 2016). Despite decades of research, estimating the extent and spatiotemporal variability of the hyporheic zone remains challenging due to the structural heterogeneity of the subsurface and the difficulty of making direct observations beyond a few discrete points (i.e., wells and piezometers; González-Pinzón et al., 2015). Numerous studies have sought to determine how the extent of hyporheic exchange responds to variable hydrologic conditions, typically with respect to the implications for which biogeochemical reactions can occur and their reach-scale significance, but results are often in conflict between sites and few generalizable behaviours have been identified (Ward, 2016). The capability for predictive modelling is similarly limited either by overly simplistic representation of the hyporheic zone as a single well-mixed storage zone (Marion et al., 2003; Wondzell, 2006) or the rarity of sufficient data needed to inform accurate representation of transport heterogeneity at scales beyond individual channel features (Schmadel et al., 2017; Ward et al., 2017). Consequently, advances in observing and modelling the dynamic behaviour of hyporheic exchange will depend on developing data-driven techniques that can constrain the spatiotemporal complexity of hyporheic exchange at functionally meaningful and tractable scales (Magliozzi et al., 2018).

Numerous definitions of the hyporheic zone have been proposed, with specific criteria reflecting the primary discipline of a given study (Gooseff, 2010; Knapp et al., 2017; Tonina & Buffington, 2007; Ward, 2016; White, 1993). A primary challenge in defining the extent of the hyporheic zone originates from the heterogeneity of nested hydrological flow paths that govern both reach-scale hydrologic transport and the significance of biogeochemical reactions (Poole et al., 2008). To incorporate prior studies and promote interdisciplinary synthesis, Ward et al. (2016) suggested that the region encompassing the hyporheic zone must (1) be in the saturated subsurface, (2) include hydrological flow paths that originate from and return to surface water, and (3) interact with the stream water within a specified temporal scale related to hydrologic or biogeochemical processes of interest. While this definition is flexible, it remains practically difficult to simultaneously delineate both the spatial and temporal boundaries implied by this definition in an actual field study.

Interactions between surface water and the hyporheic zone are most often assessed through conservative-solute tracer injections (Harvey et al., 1996; Harvey & Bencala, 1993; Kasahara & Wondzell, 2003; Ward et al., 2019). The resulting solute breakthrough curves (BTCs) reflect the effects of advection, diffusion, and transient-storage processes (both surface and subsurface) that are integrated over space and time (Stream Solute Workshop, 1990). Point-scale subsurface sampling combined with surface-water data is used for inverse model tuning to estimate lumped transport and storage parameters, but the results are often not clearly meaningful (Marion et al., 2003; Wondzell, 2006). Additionally, main-channel BTC observations reflect the convolution of multiple transient-storage zones and their processes, yet these compartments are known to exhibit distinct biogeochemical functions, especially aerobic versus anaerobic processes, making the need to parse their effects important in many studies (Knapp et al., 2018). Despite the computational feasibility of numerically modelling spatially explicit heterogeneity of coupled transport and biogeochemical reactions (e.g., Marzadri et al., 2011), it is rarely possible to overcome issues of equifinality in parameterizing multiple storage-zone models from surface and well BTC observations alone (e.g., Bottacin-Busolin, 2019), especially for reactive solutes (e.g., Kelleher et al., 2019).

The use of time-lapse electrical resistivity (ER) imaging of hyporheic exchange during tracer injections was introduced over a decade ago (Singha et al., 2008; Ward et al., 2010b) as a promising method for characterizing time-varying hyporheic extent. Fundamentally, ER methods utilize surface measurements of electrical potentials from induced current flow to estimate subsurface properties that are sensitive to the introduction and transport of electrically conductive solute tracers (Singha et al., 2008). Since its adaptation to stream tracer studies, ER imaging has been employed to investigate how hyporheic exchange, especially its extent, responds to in-channel downed wood (Doughty et al., 2020), seasonal flow recession (Ward et al., 2012; 2014), flow regulation by dam operation (Cardenas & Markowski, 2011), and structural variations in bedrock boundaries (Rucker et al., 2021).

Successful application of ER for delimiting and quantifying changes in hyporheic exchange has been impeded by reliance on qualitative analysis of hyporheic extent or a priori assumptions about the consequences of data quality and inversion decisions on the final model images when more quantitative analyses are attempted. Prior studies have relied on the evaluation of time-lapse 2D ER model images to visually compare hyporheic extent between times of data collection (e.g., Ward et al., 2010b), largely ignoring well-known sensitivity issues (e.g., Day-Lewis et al., 2005). This approach is flexible but does not allow for robust quantitative comparison or prediction between stream reaches or injections. Alternatively, some studies (e.g., Doughty et al., 2020) have analysed spatially lumped changes in bulk apparent resistivity data, which provide a basis for quantitatively describing temporal variations in exchange but do not include the spatially resolved information inherent in the inverted ER models. Finally, a few studies (e.g., Ward et al., 2010b; 2012) have applied a priori signal thresholds to delimit and estimate hyporheic extent from model images. Unfortunately, the resulting estimates of hyporheic extent are particularly sensitive to the subjectively selected change in resistivity ($\Delta\rho$) threshold (Ward et al., 2010b), so only relative changes, not the actual extents, are meaningful. Application of a standard threshold across multiple datasets (i.e., $|\Delta\rho| > 2.5\%$) does not account for actual differences in either the quality of the data or the relative strength of tracer injection signals compared to noise.

In order to more objectively evaluate hyporheic extent, this study introduces a novel method for analysing inverted ER images based on unsupervised time-series clustering. Our approach simultaneously addresses the challenges of resolving the extent and spatial

heterogeneity of hyporheic exchange during tracer injections. Unsupervised clustering is a data-mining technique in which time series are objectively grouped based on structures within the data rather than a priori assumptions (e.g., Aghabozorgi et al., 2015; Fu, 2011). We apply unsupervised clustering to time-lapse ER models from a benchmark dataset (Ward et al., 2012; 2020) to assess how the extent and heterogeneity of hyporheic connectivity changes during baseflow recession in a headwater stream. Prior work in this steep, highly constrained valley showed that riparian water-table head gradients, the extent of tracer transport into riparian zones, and hyporheic transport timescales are quite stable even as streamflow changes by three orders of magnitude (Voltz et al., 2013, Ward et al., 2016). Yet, other work suggests that hyporheic flow path geometries change during recession (Ward et al., 2017). Thus, this dataset provides opportunity to assess whether the structure of transport heterogeneity—that is the extent and location of more and less advective regions—within the hyporheic zone is responsive to seasonal changes in surface hydrologic forcing. We expect that the influence of constant valley morphology and steep down-valley gradients in this system will result in stable structuring of both cross-sectional hyporheic extent and exchange heterogeneity regardless of changing surface flows.

In this study, we test that unsupervised clustering of inverted ER model time series from tracer injections can be used to (1) delimit hyporheic extent by distinguishing solute transport signals from noisy background inversions (adjacent hillslopes or at depth), and (2) characterize transport heterogeneity within the hyporheic zone in terms of spatially defined functional zones. Clustering of inverted ER models, therefore, represents a shift towards a data-driven functional zonation representation of hyporheic connectivity, which is akin to other facies frameworks in which complex heterogeneity is simplified by characterizing compartments for which in-group heterogeneity is smaller than between-group differences (e.g., Delforge et al., 2021; Hermes et al., 2020; Hou et al., 2019; Sassen et al., 2012; Wainwright et al., 2014).

2 | METHODS

The principles of ER data collection and its application in stream tracer studies have been described extensively by prior studies (González-Pinzón et al., 2015; McLachlan et al., 2017; Singha et al., 2008; Ward et al., 2010a; 2010b; 2012; 2014). Briefly, ER measurements are sensitive to lithology, porosity, connectivity of pore spaces, pore fluid conductivity, subsurface temperature, and subsurface moisture content. The introduction of an electrically conductive tracer alters pore-fluid conductivity, thereby allowing detection of solute transport through the subsurface (Singha et al., 2008). Data are collected by applying an electric current (I, A) to the ground surface and measuring the resulting potential difference (V) between two locations to calculate the geometry-dependent resistance (R, Ω) by Ohm's Law ($R = V/I$). Then, depending on the arrangement of electrodes, a geometric factor (K) can be calculated for each measurement (see Binley, 2015a) which is used to convert R to apparent electrical bulk resistivity (ρ, Ωm) as $\rho = KR$. Apparent resistivity can, in turn, be converted to apparent electrical bulk conductivity (σ, S/m) as $\sigma = 1/\rho$. We present results for this study in terms of σ, which is more intuitively related to fluid conductivity (σ_{fl}) typically measured in surface water or wells during tracer studies.

2.1 | Injections and ER collection from a benchmark dataset

We use ER survey data and main channel σ_{fl} from tracer studies conducted in a forested second-order stream within the H. J. Andrews Experimental Forest, Oregon (44°13′N, 122°15′W) during the summer of 2010 (Ward et al., 2020). This dataset has been previously used to examine hyporheic connectivity throughout baseflow recession (Ward et al., 2012; 2014) and, therefore, serves as a useful benchmark for comparison of new methods. We focus our analysis on four 48-h tracer tests that were conducted in a 50-m reach of a headwater stream in Watershed 3 (101 ha) for discharges decreasing from 35 to 4 L s^{-1}. All injection solutions contained only sodium chloride (NaCl) as a conservative tracer. A two-week recovery period was observed between injections. ER data were collected using dipole-dipole configurations with an IRIS Syscal Pro (Orleans, France) on lateral transects, each consisting of 12 surface electrodes with ~1 m spacing (Figure 1). The average stacking error on repeat measurements within this dataset was 0.2% across all data while the average reciprocal error (collected for 55 of 323 quadripoles) was 1.3% (Ward et al., 2014). While data were collected from six transects, we consider Transects 2 and 3 as only they have complete data from all four injections. They also represent different channel morphologies, with Transect 2 crossing a more constrained section, while Transect 3 crosses a wider portion of the stream with notably more large boulders.

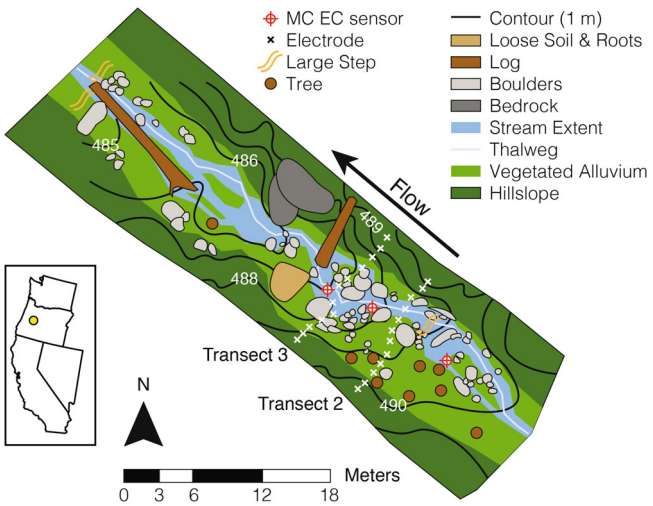

FIGURE 1 Site location and instrumentation map for WS3 in the H. J. Andrews Experimental Forest, located in the Cascade Range of Central Oregon. Main channel electrical conductivity sensors are identified as A–C from upstream to downstream locations in the text. Modified from Ward et al. (2012).

2.2 | ER inversions

We inverted resistance data from surface measurements with R2 (Binley, 2019; v2.7b compiled for Unix), which uses a regularized objective function and weighted least-squares regression approach to model and solve current flow in a quadrilateral finite-element mesh for each transect and injection (Binley, 2015b; Binley & Kemna, 2005). The number of nodes in the inversion meshes for this study ranged from 2394–2698 due to differences in transect widths. Each inversion mesh was generated with 25-cm spacing horizontally and 20-cm spacing vertically down to a relative depth of 6 m, with spacing doubling to each node thereafter. In all cases, we extended the inversion mesh at least 100 m horizontally beyond the outermost electrode locations and to a depth of about 150 m to reduce boundary effects. Surface topography for the inversion mesh was linearly interpolated between surveyed electrode locations. The duration of data collection following the end of each injection varied. Therefore, for comparability between injections, we limited analysis to data collected between 8 h prior to and 96 h after the beginning of each injection – a period for which data was available for each transect during all injections.

For the time-lapse inversions we utilized a difference method wherein the first timestep data and ER model are used as a starting model and target dataset to which subsequent inversions are regularized (Binley, 2015b). Changes in conductivity ($\Delta\sigma$) from the starting model are provided for each period of collection (timestep, hereafter) during the inversion process. For each injection and transect we calculated the diagonal of the resolution matrix as described by Binley and Kemna (2005) to quantify nodal sensitivity within the inversion mesh. We then used that matrix to select nodes with a resolution of at least 1%, meaning that at least 1% of the node's modelled conductivity was independent of adjacent nodes, for subsequent analyses (Binley & Kemna, 2005; Ward et al., 2012). Areas with resolution values <1% were parts of the inverted model for which temporal changes in conductivity could not be meaningfully interpreted.

2.3 | Unsupervised clustering of time-lapse ER models

As an alternative to qualitative assessments of hyporheic extent, spatially lumped analysis, or a priori selection of signal thresholds, we used unsupervised hierarchical clustering of nodal $\Delta\sigma$ time series to identify clusters of nodes for which within-group differences in tracer response are smaller than between-group differences. In so doing, we (1) estimated the spatial arrangement of functional zones, (2) identified characteristic $\Delta\sigma$ BTCs for each cluster, and (3) estimated total hyporheic extent. Applying this method to time-lapse ER models identifies emergent patterns within the model outputs and retains both spatial and temporal information but does not require selection of arbitrary cutoffs in $\Delta\sigma$ to determine where meaningful changes have occurred.

2.3.1 | Clustering of nodal ER time series

From the time-lapse ER models, we calculated a metric describing the dissimilarity between pairs of nodal $\Delta\sigma$ time series (Figure 2a–c). To do so, we calculated the absolute value of Euclidean distances for all

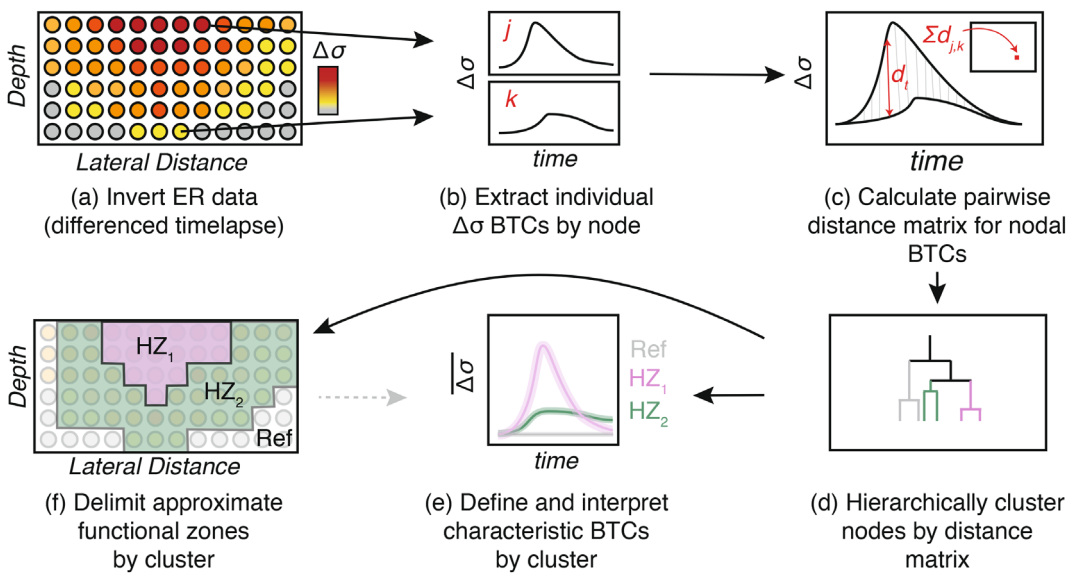

FIGURE 2 Conceptual depiction of unsupervised clustering analysis of time-lapse ER models from tracer injections. At each timestep (a) percent change in modelled conductivity ($\Delta\sigma$) relative to the pre-injection condition is generated, then (b) time series of $\Delta\sigma$ are extracted for each node in the inversion mesh, (c) Euclidean distances are calculated and summed for each pairwise comparison of nodes to construct a dissimilarity matrix, which is then used to (d) hierarchically cluster nodes. The resulting clusters can then be analysed for (e) characteristic BTCs and (f) approximate spatial extent.

pairwise combinations of the $\Delta\sigma$ time series, at each timestep, using the *TSclust* package in *R* (Montero & Vilar, 2014; R Core Team, 2019). The absolute Euclidean distance (d) between the time series for any two nodes (j and k) with $\Delta\sigma_{j,t}$ and $\Delta\sigma_{k,t}$ as their respective percent change in electrical conductivity at timestep t is:

$$d_{j,k}^{t} = \left|\Delta\sigma_{j,t} - \Delta\sigma_{k,t}\right|. \quad (1)$$

We summed these distances for t between 8 h prior to (t_i) and 96 h after (t_f) the injection commenced to give the individual elements of the dissimilarity matrix ($\boldsymbol{D}$):

$$\boldsymbol{D}_{j,k} = \sum_{t_i}^{t_f} d_{j,k}^{t}. \quad (2)$$

We opted to use Euclidean distances to construct a dissimilarity matrix because they represent the simplest distance metric that retains the physical meaning of the time-series values (no unit conversion) and are sensitive to both scaling and synchronicity in structure amongst time series, unlike other metrics used for time-series comparisons (e.g., Aghabozorgi et al., 2015; Mueen & Keogh, 2016). Consequently, the resulting distance $\boldsymbol{D}_{j,k}$ is indicative of the similarity the in both shape and timing of the two time-series in response to the tracer transport.

To identify similarities in hyporheic exchange processes within the subsurface, we then applied the built-in agglomerative hierarchical clustering algorithm in *R* (*hclust*; R Core Team, 2019) to $\boldsymbol{D}$ (Figure 2d). Individual nodes are first assigned to their own clusters, then at each subsequent iteration the most similar clusters are merged until a single cluster is formed. Here we use the default complete-linkage method to identify the nearest clusters to be merged at each step. The resulting dendrogram consists of $n-1$ branching events, where n is the number of nodes retained from the ER inversion mesh with resolution >1%. The value of n will vary depending on electrode configuration in the field, data quality, and decisions about the discretization of the differential equations during the inversion process. For this study, n ranged from 300–408.

2.3.2 | Identifying the number of clusters and characteristic BTCs

There is no best approach to selecting the "correct" number of clusters or how to cut a dendrogram (Warren Liao, 2005) and many different cluster-validity indices have been proposed, as summarized in a decadal review by Aghabozorgi et al. (2015) and in application to time-lapse ER models by Delforge et al. (2021) or other environmental phenomena (e.g., Savoy et al., 2019). We pass the dendrogram data through a non-parametric permutation-based test of within- versus between-branch variances (Park et al., 2009) to determine whether each branching event results in the formation of clusters with statistically different responses to the tracer addition. Significant branching events are identified if they satisfy a Bonferroni corrected p-value threshold ($p < 0.05/[n-1]$, where n is the number of nodes). In selecting this approach, we base cluster retention on patterns and structures that exist within the data in a way that allows for an asymmetric combination of non-significant branching events (Park et al., 2009). With this method, any number of clusters that is statistically supported can be retained depending on the end goal or application of the resulting information. We present results for four statistically unique clusters for each transect and injection. This decision was made because objectively parsing the hyporheic zone into a few functional zones signifies an advance beyond representing it as a single well-mixed compartment while not exceeding the complexity represented in widely available, computationally inexpensive multiple transient storage zone models (Briggs et al., 2009; Choi et al., 2000; Kerr et al., 2013; Knapp et al., 2017; Neilson et al., 2010).

After identifying four statistically unique clusters and the membership of individual nodes, we determined the characteristic $\Delta\sigma$ BTC for each cluster by calculating the mean and standard error (SE) of individual nodal $\Delta\sigma$ values within a cluster by timestep (Figure 2e). We calculated the SE instead of the standard deviation as the number of nodes within each cluster can vary largely, with some clusters potentially including fewer than 10 nodes while others may include hundreds.

2.3.3 | Identifying and delimiting extent of hyporheic exchange

Next, we identified which of the retained clusters represent the effective hyporheic zone – that is, which groups of nodes have time series that are reflective of tracer transport at the timescale of interest for a particular injection as informed by the BTC observed in the stream. We used the dendrogram and cluster-wise characteristic BTCs to distinguish clusters comprising the effective hyporheic zone and those that behave as non-responsive "reference" nodes (i.e., "HZ1" and "HZ2" vs. "Ref" in Figure 2e). Specifically, reference nodes lack BTC structure related to the tracer injection. However, reference nodes may exhibit some temporal patterns due to variations in temperature and soil moisture or the spatial smearing of signals through the mesh by the inversion algorithm (Day-Lewis et al., 2005). In contrast, we interpret clusters representing the effective hyporheic zone as exhibiting BTCs with systematic increases in σ (decreases in ρ) from the pre-injection state, which is indicative of conductive solute transport (Singha et al., 2008; Ward et al., 2010b). If such qualitative distinctions between BTC shapes for reference and hyporheic clusters were not obvious, we utilized the branching structure of the dendrogram to inform decision making.

We then qualitatively categorized the clusters comprising the effective hyporheic zone based on the speed and magnitude of their $\Delta\sigma$ BTC as "fast", "moderate" or "slow". These descriptors reflect the relative $\Delta\sigma$ BTC behaviours that indicate differences in advective versus diffusive solute transport amongst the clusters, with "fast" being the most advective (i.e., "HZ1" in Figure 2e). Ward et al. (2010a)

discuss characterization and analysis of BTCs spanning the advective-diffusive continuum and their expression in space throughout a tracer injection. Since these designations denote relative differences amongst nodes within a single transect and for a single injection, we are not required to set a priori criteria (i.e., an exact value or range of median arrival times) that define each category apart from being statistically different from one another (see Section 2.3.2). Thus, our analysis of differences in hyporheic transport is data driven. Finally, we calculated the approximate extent of the clusters comprising the effective hyporheic zone based on the location of nodes within the inversion mesh.

2.4 | Comparison to threshold-based estimates of hyporheic extent

Prior studies (e.g., Ward et al., 2010b; 2012) have estimated hyporheic extent based on a priori selection of $\Delta\rho$ (or $\Delta\sigma$) thresholds. Because the resulting estimates of hyporheic extent are particularly sensitive to the selected threshold (Ward et al., 2010b), only relative changes, not the actual extents, are likely meaningful—ignoring issues with out-of-plane effects (e.g., Bentley & Gharibi, 2004). Therefore, we identified the nodes for each injection and transect for which $\Delta\sigma \geq 2$, 3, 4, 5 and 10% for at least one timestep during the injection to compare patterns across injections as in Ward et al. (2012). We then estimated the hyporheic extent by calculating the total area within the inversion mesh represented by the nodes retained by each of these 5 thresholds. We compared the directionality and magnitude of changes in threshold-based and clustering-based extent estimates in response to seasonally declining streamflow.

Even if the duration of repeat constant-rate injections is nearly identical, it is difficult to perfectly replicate the same magnitude of change in σ_{fl}, especially when surface discharge changes. Also, plateau concentrations are not always steady during constant rate injections. Therefore, to characterize differences in the overall forcing on σ_{fl} over the entire BTC for each injection, we calculated the zeroth temporal moment (μS cm^{-1} h) of the σ_{fl} (μS cm^{-1}) time series from the sensor between the two ER transects during each injection as:

$$M_0^{\sigma,fl} = \int_{t_i}^{t_f} \sigma_{fl} dt. \quad (3)$$

Here t is elapsed time (hours) between the injection start time (t_i) and 96 h (t_f) that constrained our analysis. Since changes in modelled ER are sensitive to σ_{fl}, we anticipated that the magnitude of $M_0^{\sigma,fl}$ would influence threshold-based estimates of hyporheic extent, but clustering may be more robust to these unintended differences between individual injection datasets. To evaluate this expectation, we analysed linear regressions between the estimated hyporheic extent and $M_0^{\sigma,fl}$ for each injection, transect, and method for delineation (i.e., threshold or clustering).

3 | RESULTS AND DISCUSSION

3.1 | Spatial arrangement and transport characteristics of hyporheic clusters

The total number of nodes with resolution ≥1% varied for each set of inverted ER data, with fewer nodes retained (n) for Transect 2 than 3. Over the four injections, n ranged from 300–324 for Transect 2 and 318–408 for Transect 3. In all instances, a permutational test (Park et al., 2009) of inter- versus intra-cluster variances identified >50 significant branching events such that there are more statistically unique clusters than can be individually interpreted. Therefore, we analysed the four clusters resulting from the four highest branching events for each transect and injection (Figure 3), which are significantly different ($p \ll 0.001$) from each other and represent a highly

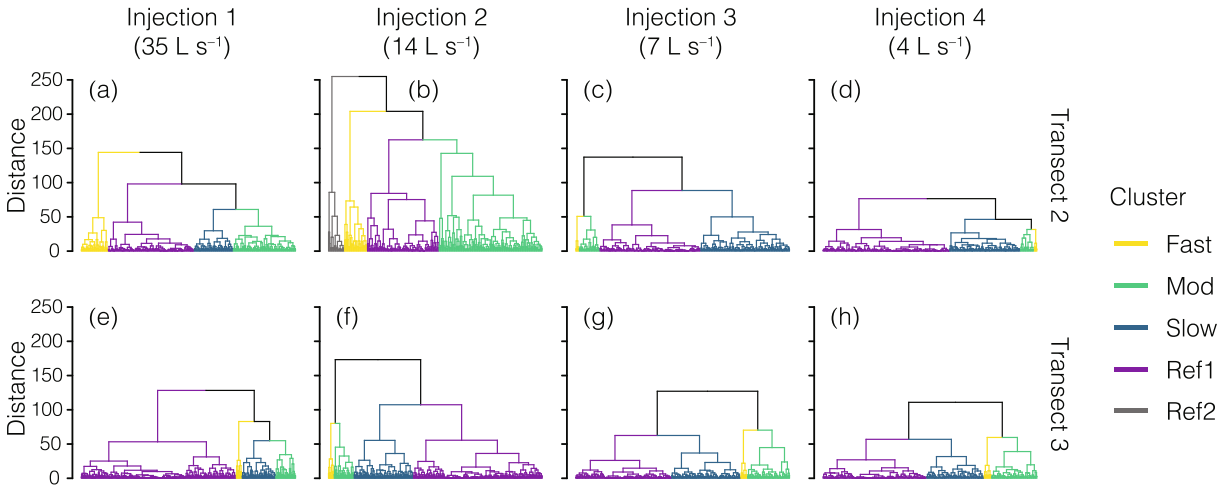

FIGURE 3 Individual injection dendrograms based of hierarchical clustering of nodal $\Delta\sigma$ time series for Transect 2 (a–d) and Transect 3 (e–h). Clusters are labelled by qualitative descriptors of characteristic BTCs (Figure 5). "Ref1" indicates the zone that is unresponsive (no BTC) to the tracer addition. "Ref2" only appears in panel b and is indicative of an inversion anomaly.

conservative evaluation of hyporheic transport heterogeneity. The designation of these four clusters would not change unless the *p*-value thresholds (all $<1.7 \times 10^{-4}$) set by a Bonferroni correction based on the number of values in each test were decreased even further by multiple orders of magnitude. Of course, the delineation of these four clusters would also remain unchanged if the *p*-value threshold was raised or uncorrected (i.e., $p < 0.05$), that would simply identify even more potential clusters with ever fewer nodes in each cluster. While our approach of interpreting the four most unique zones joins smaller clusters formed by subsequent branching events that exhibit significantly different $\Delta\sigma$ responses to the tracer ($p \ll 0.001$), all the within-cluster differences are significantly smaller than those between clusters and no clusters from non-significant branching events are retained. Given the ill-determined nature of inverse modelling from surface ER data and resolution limitations, we chose to avoid retaining large numbers of clusters (>50), wherein the size of some clusters would potentially become too small to generate valid statistical comparisons or meaningful hydrological interpretations. Moreover, evaluation of four clusters (only some of which will represent regions with active exchange) is also well aligned with numerous readily available multiple storage zone models that are frequently and successfully used to represent transport and biogeochemical phenomena in streams (e.g., Briggs et al., 2009; Choi et al., 2000; Kerr et al., 2013; Knapp et al., 2017; Neilson et al., 2010).

In general, the hillslopes to either side of the stream were joined into one cluster with additional clusters forming in a radial pattern within the valley bottom (Figure 4). Based on relative differences in the characteristic $\Delta\sigma$ BTCs (Figure 5), we found that the clusters with the most advective signatures ("fast") occurred nearest to the surface. These regions were ringed by the cluster which exhibited moderately ("mod") advective behaviour, while the least advective ("slow") clusters were located at even greater depths and lateral distances within the subsurface. We again emphasize that these designations are qualitative descriptors based on relative differences for a particular injection and that further quantitative descriptions (or even set definitions) are possible, but beyond the scope of this particular study. This spatial organization matches both conceptual expectations of hyporheic exchange and prior visualizations of ER data from tracer studies in streams (e.g., Doughty et al., 2020; Ward et al., 2010a; 2010b; 2012). However, it is notable that this pattern is neither explicitly defined by nor provided as an input in either the inversion or the cluster-identification algorithms.

While the nodal membership and spatial arrangement of clusters shifted over the four injections, we observed persistent patterns in the organization of functional zones. For Transect 2, the primary pattern is the assignment of two spatially separated regions on either side of the valley bottom to the same cluster (i.e., "fast" cluster for injections 1 and 2, shifting to "slow" for injections 3 and 4). In contrast, a singular radial clustering pattern was exhibited in Transect 3 throughout all four injections. This difference highlights the ability of hierarchical clustering to parse the effective hyporheic area into functional zones with spatial arrangements that reflect connectivity to surface water at a particular transect, but that are not necessarily contiguous. Additionally, the location and extent of certain functional zones (i.e., "fast" for Transect 3) change very little even as flow changes suggesting that this functional zonation approach based on clustering is sensitive to spatial differences in hyporheic exchange that reflect stable physical properties. This result is consistent with prior findings from this site that riparian water table gradients and the extent of tracer transport into riparian zones are quite stable and predominantly down-valley even as streamflow changes by three orders of magnitude throughout summer (Voltz et al., 2013) and reflects the predominant understanding that in this highly constrained valley bottom, advective hyporheic transport is principally influenced by steep

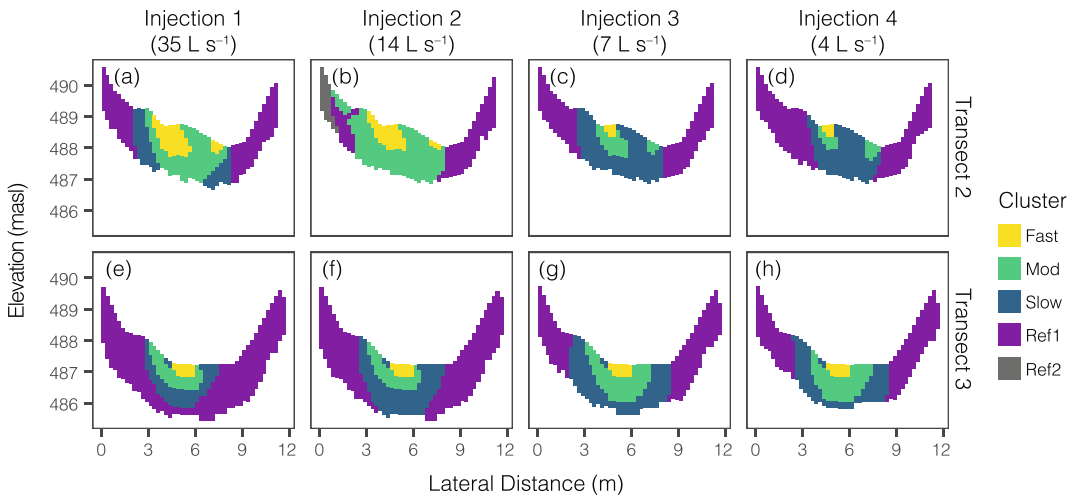

FIGURE 4 Inversion mesh cross-sections with individual node regions coloured by cluster membership for Transect 2 (a–d) and Transect 3 (e–h) across each of the four injections. Vertical relief is exaggerated two-fold. Clusters are labelled by qualitative descriptors of characteristic BTCs (Figure 5). "Ref1" indicates the zone that is unresponsive (no BTC) to the tracer addition. "Ref2" only appears in panel b and is indicative of an inversion anomaly. Stream water surfaces and cross sections are not highlighted due to the shallow nature of the stream relative to riparian vertical relief and depth of ER modelling.

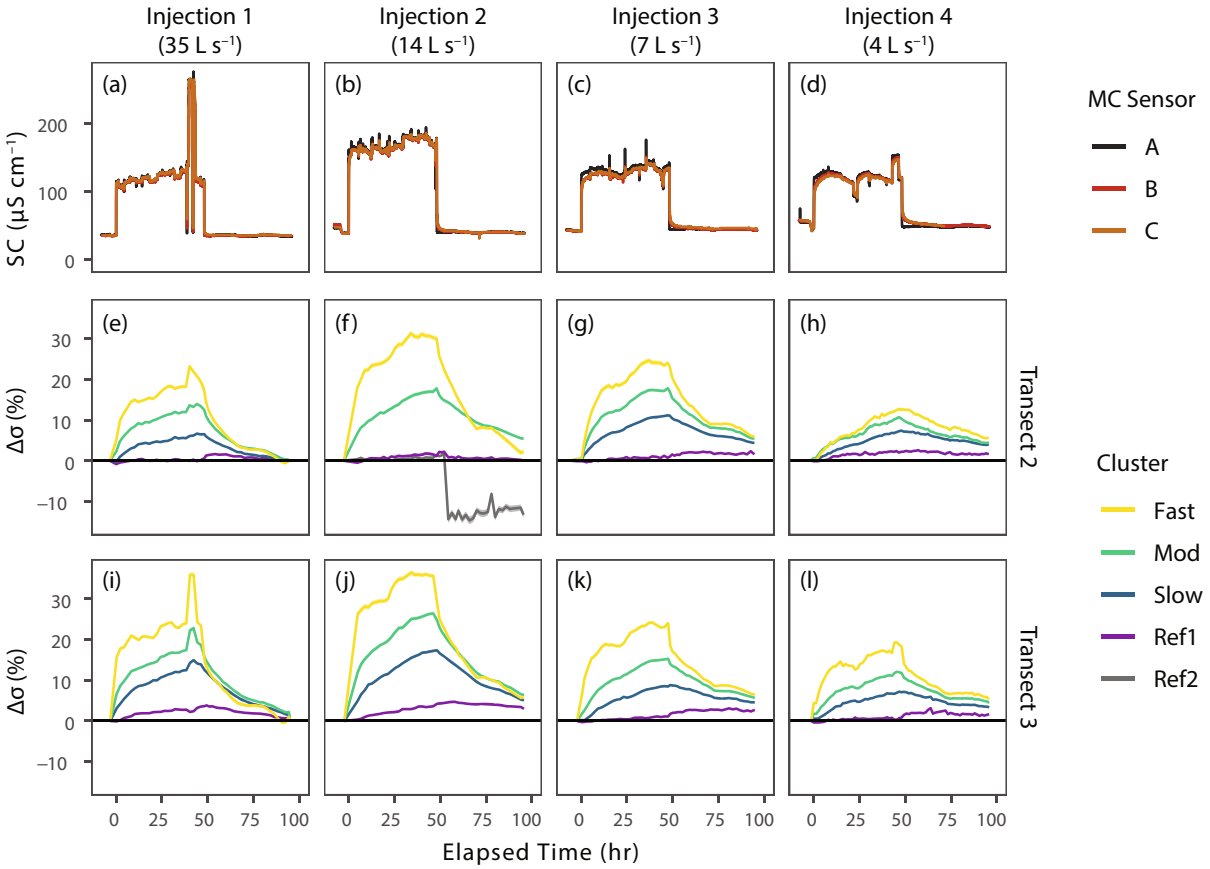

FIGURE 5 Observed specific conductivity BTCs for main channel sensors (a–d) and cluster-wise $\Delta\sigma$ BTCs for Transect 2 (e–h) and Transect 3 (i–l). The $\Delta\sigma$ traces (e–l) are presented as ribbon plots (mean ± *SE*) representing $\Delta\sigma$ values for nodes within each cluster at each timestep. Errors are relatively small compared to line width in most cases. Clusters are labelled by qualitative descriptors of characteristic BTCs. "Ref1" indicates the zone that is unresponsive (no BTC) to the tracer addition. "Ref2" only appears in panel b and is indicative of an inversion anomaly.

down-valley gradients and large morphological features (i.e., Ward et al., 2014, 2017). As expected, this results in stable spatial patterning of functional zonation that is relatively insensitive to multiple orders of magnitude change in streamflow at seasonal timescales. In other systems, hyporheic exchange and riparian water-table head gradients are not seasonally stable (e.g., Burt et al., 2002; Vidon & Hill, 2004; Wroblicky et al., 1998) and event fluctuations in surface and groundwater conditions alter hyporheic exchange at shorter timescales (e.g., Malzone et al., 2016; Ward et al., 2013; Wondzell & Swanson, 1996; Zimmer & Lautz, 2013). However, the clustering method does not rely on assumptions about such behaviour and would, therefore, be responsive to site-specific dynamics.

As noted in prior studies of this watershed (e.g., Ward et al., 2012), the main-channel fluid-conductivity BTCs from each injection reached a plateau quickly (<1 h) and returned to background levels over slightly longer, but still fairly rapid time spans (<2 h) after the injection ended (Figure 5a–d). In contrast, characteristic $\Delta\sigma$ BTCs for each cluster show a much more gradual shift from background conditions and most do not reach a fully plateaued state (Figure 5e–l), especially for injection 4 at the lowest flows. While many clusters show an initially rapid response to the end of the injection, the rate at which characteristic $\Delta\sigma$ BTCs return to their pre-injection state generally slows after a few hours, and most do not return to the initial state even 48 h later. These behaviours reflect the sensitivity of ER to low concentrations of solute that are retained in and slowly released from relatively immobile pore spaces and diluted below detection levels of in-stream fluid electrical conductivity sensors (Singha et al., 2008; Ward et al., 2010a). Notably, the "fast" clusters load and unload with tracer the most rapidly and exhibit the greatest change in σ, reflecting more advective transport and greater dominance by mobile domains. In contrast, the "slow" clusters show the smallest and most gradual σ responses, indicative of less advective transport and more immobile pore spaces. Clustering of inverted ER data, therefore, categorizes portions of the subsurface in terms of BTC behaviour that emerges from distinctive combinations of transport phenomena (i.e., advection vs. diffusion vs. transient storage) and relative density of mobile versus immobile domains within the subsurface.

By providing a means to quantify how these behaviours differ at each timestep for spatially defined sets of nodes, clustering improves upon approaches that either lump all surface ER data together (i.e., bulk apparent-conductivity time series; Doughty et al., 2020) or characterize spatial trends within modelled cross-sections for only a small subset of times (i.e., $\Delta\rho$ cross-sectional images at a few selected

timesteps; Ward et al., 2012). Thus, this application illustrates how functional zones defined by their transport characteristics can be identified and approximately mapped in space by hierarchical clustering of inverted ER data.

We interpreted reference clusters across injections and transects that did not exhibit an obvious BTC response to the tracer addition ("Ref1"; Figure 5) and were mostly located within adjacent hillslopes above the streambed where exchange is unlikely (Figure 4). Generally, this reference cluster was the largest across all injections and exhibited mean $\Delta\sigma$ traces that were flat and near zero (< ±5%) as would be expected. Notably, however, for Transect 2 injection 2, cluster "Ref2" showed a flat $\Delta\sigma$ response during the injection but had a large negative $\Delta\sigma$ step change in the post-injection period. Such a negative $\Delta\sigma$ response is likely an artefact of the second-derivative smoothing incorporated into the inversion process. Detection of this artefact, along with the notably different spatial arrangement of functional zones, demonstrates that clustering may be a useful tool in identifying datasets and inversions with errors than can impact subsequent analyses. Altogether, these examples demonstrate that hierarchical clustering is useful for identifying spatial organization of distinct tracer transport signals in the subsurface, distinguishing unresponsive nodes from those in the effective hyporheic zone and detecting inversion-process artefacts.

While main-channel fluid conductivity BTCs vary between injections (Figure 5a–d), some notable patterns reflective of seasonally evolving subsurface solute transport emerge across the injections. Specifically, cluster-wise $\Delta\sigma$ BTC shapes (particularly the "fast" clusters) shift temporally towards slower loading and weaker plateauing. That change in the "fast" cluster behaviour is most apparent for Transect 2 and to lesser extent for Transect 3. This difference between transects is suggestive of subsurface advective transport declining more substantially for Transect 2 as surface flow (35 L s^{-1} at injection 1 to 4 L s^{-1} during injection 4). Temporal moment analysis on nodal BTCs for this dataset by Ward et al. (2014) similarly reported larger changes in first-arrival time, mean-arrival time, and skewness for Transect 2 than Transect 3. That analysis required lumping data for all nodes identified as part of the effective hyporheic zone, while our analysis parses this into spatially defined functional zones.

3.2 | Comparison of methods for estimating effective hyporheic extent

We compared the total effective hyporheic extent amongst injections with different streamflow rates estimated by hierarchical clustering with $\Delta\sigma$ thresholds (Figure 6). We limit analysis of relative differences in extent to a single significant digit due to the imprecise nature of models resulting from inversion of field data. Based on the clustering method, we found that total effective hyporheic extent for Transect 2 decreased by ~40% (10 to 6 m^2) between the highest and lowest flows. In contrast, for Transect 3 cluster-based analysis resulted in a ~30% increase (6 to 8 m^2). For Transect 2, threshold-based estimates of hyporheic extent were relatively stable for flows of 35 and 14 L s^{-1}, with decreasing extents observed as flow fell from 14 to 7 and then 4 L s^{-1}, except for $\Delta\sigma \geq 2\%$. For Transect 3, estimated hyporheic extent increased as flow fell from 35 to 14 L s^{-1} then declined thereafter.

Interestingly, the cluster-based estimates of total effective hyporheic extent are relatively more stable at the three lowest streamflow conditions and most align with estimates from different threshold values depending on the injection and transect. For the lowest flow conditions, cluster-based extent estimates for both Transect 2 and Transect 3 are very similar to those generated by a $\Delta\sigma$ threshold of 5%, while this shifts to 10% for the 14 L s^{-1} injection. At the highest-flow condition, the cluster-based extent estimate for Transect 2 was again more similar to a 5% threshold while Transect 3 was more similar to that based on a 10% threshold. Thus, applying a single $\Delta\sigma$ threshold across multiple datasets does not replicate the way in which clustering distinguishes reference from responsive signals to delimit hyporheic extent, likely because clustering is a more robust method than threshold-based analysis of hyporheic extent when comparing inverse model results amongst injections.

We also compared the estimated hyporheic extent for each method against the integral of the main channel conductivity BTCs ($M_0^{\sigma,\mathrm{fl}}$) of each injection. We found that, for $\Delta\sigma$ threshold-based methods, the estimated hyporheic extent had a strong positive linear correlation with $M_0^{\sigma,\mathrm{fl}}$ while clustering was less directionally sensitive to $M_0^{\sigma,\mathrm{fl}}$ (Figure 7). In conjunction with the relations between discharge and hyporheic extent by method (Figure 6), this analysis demonstrates

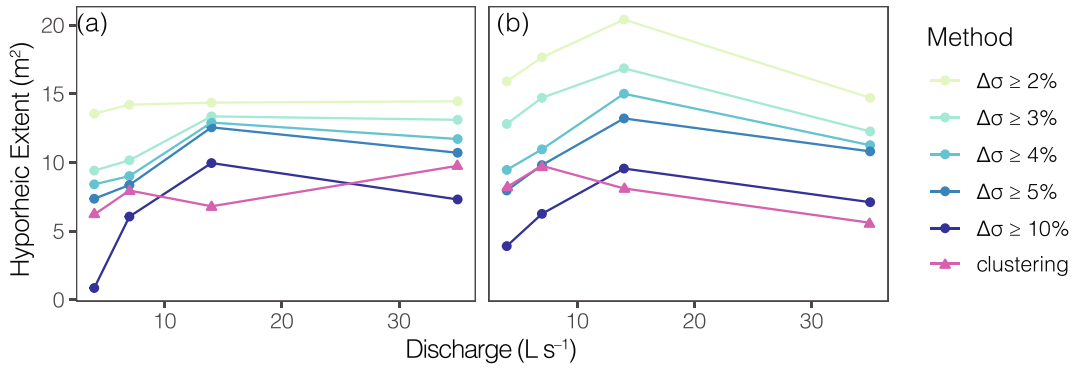

FIGURE 6 Total effective hyporheic extent by method versus surface discharge for (a) Transect 2 and (b) Transect 3.

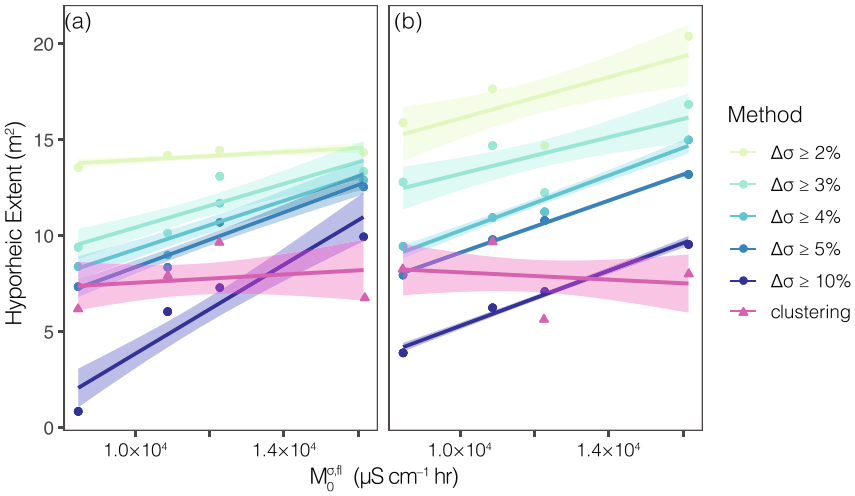

FIGURE 7 Estimated hyporheic extent by method versus the integral of main channel conductivity ($M_0^{\sigma,fl}$) during each injection for (a) Transect 2 and (b) Transect 3. Transparent ribbons depict 95% confidence intervals on linear regressions.

that changes in the hyporheic estimate obtained by threshold analyses are more sensitive to differences in the tracer injection than to changes in exchange resulting from an order of magnitude variation in discharge. Conversely, cluster-based estimates of hyporheic extent are more directionally related to differences in discharge between injections than to $M_0^{\sigma,fl}$. This is a very important finding, especially given the near impossibility of performing multiple injections at different discharge conditions with exactly the same BTC integral. Beyond methodological considerations, the relatively consistent extents across injections resulting from clustering analysis aligns with observations from wells that exchange and water table gradients in this system are relatively stable throughout baseflow recession (Voltz et al., 2013). Clustering, therefore, provides a method for estimating the spatial extent of hyporheic exchange from time-lapse ER models that is more robust to unintended variations in BTCs that occur during field-based tracer experiments.

Finally, it is notable that the total hyporheic extents that we estimate (as well as the depth to which nodes are retained) are smaller than those reported for prior analysis of the surface ER data (Ward et al., 2012; 2014). This highlights an additional sensitivity not just to the method used to delimit hyporheic extent, but also to small differences in the inversion mesh, models used for regularization, and inversion settings as the underlying data are the same. We tested clustering on those original inversions and, just as with this study, the lower boundaries of clusters interpreted as representing the effective hyporheic zone were all defined by resolution limitations. In other words, the reference clusters did not extend completely and contiguously beneath the hyporheic clusters due to insufficient resolution beyond the depth of probable tracer penetration. The spatial structure and arrangement of clusters within the hyporheic zone, however, are not obviously subject to this issue.

Due to the ill-determined nature of the inverse problem and resolution limitations, neither the ER models nor our calculated hyporheic extents represent a precise quantification of the system. Rather, they are simply estimates based on a smoothed approximation of subsurface properties, from which we more objectively, but imprecisely, map patterns via hierarchical clustering.

We thus emphasize that it is essential to recognize that such issues with insufficient resolution at depth, combined with the smoothing inherent in the inversions, means that the extent estimates should be cautiously interpreted.

3.3 | Potential extensions

While the analyses we introduce are useful for characterizing transport heterogeneity within the effective hyporheic zone, they also have the potential to improve model representation of hyporheic processes. For instance, utilizing our proposed method to identify the extent and characteristic BTCs for regions within the hyporheic zone could be used to parameterize cross-sectional areas prior to inversely tuning exchange coefficients in reactive transport models. This possibility is especially important as physical parameters such as the extent of hyporheic exchange cannot be uniquely determined solely by observing surface-water BTCs (Bottacin-Busolin, 2019). Similar progress has been achieved using the hydrologic facies frameworks to parameterize reduced-complexity models based on sediment property observations (Hou et al., 2019).

The utility of clustering could also be further extended through changes to the inversion process. Most obviously, for streams that have a large enough cross-sectional area, the stream itself could be included as a specified domain in the inversion mesh and subsequent clustering. The resulting model components representing surface water conductivity could be directly compared to field data from in-stream conductivity sensors, rather than comparing observed surface water conductivities to modelled bulk subsurface conductivity as in this study (Figure 5). Post-inversion clustering could also employ constrained inversion techniques beyond standard regularization with smoothing as used here. For example, using Ensemble Kalman Inversions (e.g., Tso et al., 2021) would allow for improved uncertainty quantification and the identification of sharp zone boundaries where regularization is not applied. This may improve both estimation of total hyporheic extent and delineation of boundaries within the hyporheic zone. Identifying zones of interest may also be

accomplished by extending alternative inversion techniques that can solve for specific, albeit simple, shapes reflective of hydrologic processes and limited artefacts using geometric moments (e.g., Pidlisecky et al., 2011). Future implementation of alternative inversion techniques must be carefully guided by the ability of the practitioner to evaluate the specific suitability, limitations, and need for prior information (or conceptualization) about a particular site or their study purpose. The method we have introduced here focuses on applications for hydrologic scientists who may not be pushing the envelope in terms of geophysical inversion methods.

Our approach may also be extended to identify the region over which point-scale sampling (i.e., from wells, piezometers or mini-point samplers) may provide representative information. This information is potentially most useful in the context of reactive-tracer studies, wherein metrics of reactivity or the balance of transport and reaction timescales from particular points could be extrapolated in space based on ER-informed functional-zone mapping. Alternatively, such functional-zone mapping could be used to develop testable hypotheses about the spatial structure of biogeochemical activity (i.e., occurrence of redox reactions, relative reaction completion, etc.) or microbial diversity that could then be tested by point sampling. For instance, denitrification in the hyporheic zone plays an important role in the attenuation of excess nutrient fluxes by streams (Gomez-Velez et al., 2015) and techniques exist to predict net hyporheic dynamics (e.g., Zarnetske et al., 2012), but not where denitrification occurs throughout a particular study site without costly and intensive point-scale sampling (e.g., Harvey et al., 2013). The weaker advective transport that characterizes "slow" zones is likely to be associated with slower flow paths (longer transit time) and oxygen depletion, such that this zonation might be used to generate site-specific predictions of where denitrification is more probable. Such extensions would represent a major development in linking heterogeneity of coupled transport and biogeochemical processes occurring at the scale of a few metres to their aggregate significance over entire reaches (e.g., Harvey et al., 2018).

Beyond identifying functional zones representing distinct transport and connectivity signals within the hyporheic zone, there are also potential extensions for this method that could advance synthesis across time (injections) and space (transects or sites). It is possible to develop methods that form clusters across merged datasets or match clusters across datasets through post-hoc comparisons so that the persistence and spatial evolution of functional zones can be investigated. The greatest challenge to this approach will be in determining how to normalize data given differences in forcing from separate solute injections. It is difficult, if not impossible, to perform multiple tracer additions that generate BTCs with the same relative change in stream water fluid conductivities, as is apparent even in this dataset (Figure 5a–d). Importantly, this new clustering approach is less sensitive (though not entirely immune) to such variations than traditional threshold-based analyses of ER models (Figure 7), thereby offering a method that is better suited to the inevitable complexity of conducting field-based tracer injections. Determining how to quantitatively handle such differences amongst injections will be necessary to differentiate changes due to the tracer input itself or subsurface transport processes when examining clustering between datasets.

Another intriguing potential extension exists around supervised clustering or similar machine-learning techniques in which cluster characteristics are defined based on an initial training dataset containing time-lapse ER and more easily obtained ancillary measurements to allow for prediction elsewhere. Such methods have been applied to classify and then predict spatiotemporal evolution of other hydrologic patterns such as seasonal soil moisture (Hermes et al., 2020) and hydrologically homogeneous regions within catchments (Nadoushani et al., 2018) based on topographic indices, but not, to our knowledge, for hyporheic exchange. For instance, exploring whether high-resolution topography data from within the river corridor (rather than the whole catchment) could be used to predict hyporheic zonation patterns beyond discrete transects is an intriguing possibility. Doing so could support reduced-complexity modelling that still represents spatial variations in functionally distinct zones at finer resolution along stream reaches than is currently possible. The primary challenge to this extension will be in determining which combination of metrics are obtainable over entire reaches (at least compared to discrete ER transects) and can provide predictive power of subsurface functional zonation.

4 | CONCLUSIONS

With the goal of developing a more objective approach to evaluating hyporheic extent, we present a method to analyse inverted ER models using unsupervised hierarchical clustering to delimit the extent of hyporheic exchange and to characterize functional zones with distinct transport behaviours within the subsurface. We used this method to show that total hyporheic extent and the spatial structure of heterogeneity of exchange were predominantly stable throughout seasonal baseflow recession ($4 < Q < 35$ L s^{-1}) for adjacent (<10 m longitudinally separated) transects in a highly constrained, steep mountain stream. While prior research at this site showed that the variability of hyporheic transport processes increases with declining streamflow (i.e., Ward et al., 2014) and individual flow path geometries shift (Ward et al., 2017), our findings reveal that such changes occur within the larger-scale context of relatively unchanging cross-sectional patterning of functional zonation, defined by relative transport characteristics. This result suggests that for this site, stable morphological and alluvial characteristics set a characteristic pattern of hyporheic exchange heterogeneity, while forcing from surface water changes result in smaller within zone changes to transport during baseflow recession. We also found that the application of a single signal threshold to delimit hyporheic extent across ER datasets cannot replicate statistically supported parsing of active hyporheic and inactive reference regions in the subsurface. While clustering does not overcome the inevitable issues of blurring in inverse models of the subsurface, it provides a more objective approach to distinguishing where and to what degree stream tracers may be exchanged with the subsurface from geophysical datasets. Clustering also helps distinguish the spatial

structure of zones with distinctive combinations of transport phenomena (i.e., advection vs. diffusion vs. transient storage) and relative density of mobile versus immobile domains within the subsurface, as well as how these structures persist or change temporally. To our knowledge, this represents the first application of machine learning to statistically classify spatial patterning of hyporheic exchange during tracer studies. Additionally, this approach has the potential to inform data-driven reduced-complexity modelling that could address known shortfalls of representing the hyporheic zone as a single well-mixed compartment.

ACKNOWLEDGEMENTS

This work was supported by funding from the National Science Foundation (EAR-1642402 and EAR-1642403). Additional support for J.G.S. was provided by the University of Colorado Boulder, including a Beverly Sears Graduate Research Grant. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

DATA AVAILABILITY STATEMENT

Field data including site surveys, electrical resistivity measurements, fluid conductivity data are available via HydroShare (Ward et al., 2020). An example script for reproducing cluster analysis of time-lapse electrical resistivity models is publicly available via Zenodo (Singley, 2022).

ORCID

Joel G. Singley https://orcid.org/0000-0002-7906-8491
Kamini Singha https://orcid.org/0000-0002-0605-3774
Michael N. Gooseff https://orcid.org/0000-0003-4322-8315
Ricardo González-Pinzón https://orcid.org/0000-0001-9387-6885
Timothy P. Covino https://orcid.org/0000-0001-7218-4927
Adam S. Ward https://orcid.org/0000-0002-6376-0061
Jancoba Dorley https://orcid.org/0000-0001-7841-2177
Eve-Lyn S. Hinckley https://orcid.org/0000-0002-7081-0530

REFERENCES

Aghabozorgi, S., Seyed Shirkhorshidi, A., Ying Wah, T., Shirkhorshidi, A. S., & Wah, Y. (2015). Time-series clustering – A decade review. *Information Systems*, *53*, 16–38. https://doi.org/10.1016/j.is.2015.04.007

Bentley, L. R., & Gharibi, M. (2004). Two-and three-dimensional electrical resistivity imaging at a heterogeneous remediation site. *Geophysics*, *69*(3), 674–680.

Binley, A. (2015a). DC electrical methods. In *Treatise on geophysics*. (pp. 233–259). Elsevier.

Binley, A. M. (2015b). Tools and techniques: electrical methods. In *Treatise on geophysics* (Vol. 11, Second ed., pp. 233–259). Elsevier Inc.. https://doi.org/10.1016/B978-0-444-53802-4.00192-5

Binley, A. M. (2019). R2: Summary. Retrieved from http://www.es.lancs.ac.uk/people/amb/Freeware/R2/R2_readme.pdf.

Binley, A. M., & Kemna, A. (2005). Electrical methods. In Y. Rubin & S. S. Hubbard (Eds.), *Hydrogeophysics* (pp. 129–156). Springer.

Bottacin-Busolin, A. (2019). Modeling the effect of hyporheic mixing on stream solute transport. *Water Resources Research*, *55*, 9995–10011. https://doi.org/10.1029/2019WR025697

Briggs, M. A., Gooseff, M. N., Arp, C. D., & Baker, M. A. (2009). A method for estimating surface transient storage parameters for streams with concurrent hyporheic storage. *Water Resources Research*, *45*(4). https://doi.org/10.1029/2008WR006959

Burt, T. P., Pinay, G., Matheson, F. E., Haycock, N. E., Butturini, A., Clement, J. C., Danielescu, S., Dowrick, D. J., Hefting, M. M., Hillbricht-Ilkowska, A., & Maitre, V. (2002). Water table fluctuations in the riparian zone: Comparative results from a pan-European experiment. *Journal of Hydrology*, *265*(1–4), 129–148.

Cardenas, M. B., & Markowski, M. S. (2011). Geoelectrical imaging of hyporheic exchange and mixing of river water and groundwater in a large regulated river. *Environmental Science and Technology*, *45*(4), 1407–1411. https://doi.org/10.1021/es103438a

Choi, J., Harvey, J. W., & Conklin, M. H. (2000). Characterizing multiple timescales of stream and storage zone interaction that affect solute fate and transport in streams. *Water Resources Research*, *36*(6), 1511–1518.

Day-Lewis, F. D., Singha, K., & Binley, A. (2005). The application of petrophysical models to radar and electrical resistivity tomograms: Resolution-dependent limitations. *Journal of Geophysical Research*, *110*, B08206. https://doi.org/10.1029/2004JB003569

Delforge, D., Watlet, A., Kaufmann, O., Van Camp, M., & Vanclooster, M. (2021). Time-series clustering approaches for subsurface zonation and hydrofacies detection using a real time-lapse electrical resistivity dataset. *Journal of Applied Geophysics*, *184*, 104203.

Doughty, M., Sawyer, A. H., Wohl, E., & Singha, K. (2020). Mapping increases in hyporheic exchange from channel-spanning logjams. *Journal of Hydrology*, *587*, 124931. https://doi.org/https://doi.org/10.1016/j.jhydrol.2020.124931

Fu, T. C. (2011). A review on time series data mining. *Engineering Applications of Artificial Intelligence*, *24*, 164–181. https://doi.org/10.1016/j.engappai.2010.09.007

Gomez-Velez, J., Harvey, J., Cardenas, M., & Kiel, B. (2015). Denitrification in the Mississippi River network controlled by flow through river bedforms. *Nature Geoscience*, *8*, 941–945. https://doi.org/10.1038/ngeo2567

González-Pinzón, R., Ward, A. S., Hatch, C. E., Wlostowski, A. N., Singha, K., Gooseff, M. N., Haggerty, R., Harvey, J. W., Cirpka, O. A., & Brock, J. T. (2015). A field comparison of multiple techniques to quantify groundwater–surface-water interactions. *Freshwater Science*, *34*(1), 139–160.

Gooseff, M. N. (2010). Defining hyporheic zones - advancing our conceptual and operational definitions of where stream water and groundwater meet. *Geography Compass*, *4*, 945–955. https://doi.org/10.1111/j.1749-8198.2010.00364.x

Harvey, J. W., & Bencala, K. E. (1993). The effect of streambed topography on surface-subsurface water exchange in mountain catchments. *Water Resources Research*, *29*(1), 89–98. https://doi.org/10.1029/92WR01960

Harvey, J. W., Böhlke, J. K., Voytek, M. A., Scott, D., & Tobias, C. R. (2013). Hyporheic zone denitrification: Controls on effective reaction depth and contribution to whole-stream mass balance. *Water Resources Research*, *49*, 6298–6316. https://doi.org/10.1002/wrcr.20492

Harvey, J. W., Gomez-Velez, J. D., Schmadel, N. M., Scott, D. T., Boyer, E. W., Alexander, R. B., Eng, K., Golden, H., Kettner, A., Konrad, C., Moore, R., Pizzuto, J., Schwarz, G., Soulsby, C., & Choi, J. (2018). How hydrologic connectivity regulates water quality in river corridors. *Journal of the American Water Resources Association*, *55*, 369–381. https://doi.org/10.1111/1752-1688.12691

Harvey, J. W., Wagner, B. J., & Bencala, K. E. (1996). Evaluating the reliability of the stream tracer approach to characterize stream-subsurface water exchange. *Water Resources Research*, *32*(8), 2441–2451. https://doi.org/10.1029/96WR01268

Hermes, A. L., Wainwright, H. M., Wigmore, O., Falco, N., Molotch, N. P., & Hinckley, E.-L. S. (2020). From patch to catchment: A statistical framework to identify and map soil moisture patterns across

complex alpine terrain. *Frontiers in Water*, *2*. https://doi.org/10.3389/frwa.2020.578602

Hou, Z., Scheibe, T. D., Murray, C. J., Perkins, W. A., Arntzen, E. V., Ren, H., Mackley, R. D., & Richmond, M. C. (2019). Identification and mapping of riverbed sediment facies in the Columbia River through integration of field observations and numerical simulations. *Hydrological Processes*, *33*(8), 1245–1259. https://doi.org/10.1002/hyp.13396

Kasahara, T., & Wondzell, S. M. (2003). Geomorphic controls on hyporheic exchange flow in mountain streams. *Water Resources Research*, *39*(1), SBH 3-1–SBH 3-14. https://doi.org/10.1029/2002wr001386

Kelleher, C., Ward, A., Knapp, J. L. A., Blaen, P. J., Kurz, M. J., Drummond, J. D., Zarnetske, J. P., Hannah, D. M., Mendoza-Lera, C., Schmadel, N. M., Datry, T., Lewandowski, J., Milner, A. M., & Krause, S. (2019). Exploring tracer information and model framework trade-offs to improve estimation of stream transient storage processes. *Water Resources Research*, *55*, 3481–3501. https://doi.org/10.1029/2018WR023585

Kerr, P. C., Gooseff, M. N., & Bolster, D. (2013). The significance of model structure in one-dimensional stream solute transport models with multiple transient storage zones–competing vs. nested arrangements. *Journal of Hydrology*, *497*, 133–144.

Knapp, J. L. A., González-Pinzón, R., & Haggerty, R. (2018). The resazurin-resorufin system: Insights from a decade of "smart" tracer development for hydrologic applications. *Water Resources Research*, *54*(9), 6877–6889. https://doi.org/10.1029/2018WR023103

Knapp, J. L., González-Pinzón, R., Drummond, J. D., Larsen, L. G., Cirpka, O. A., & Harvey, J. W. (2017). Tracer-based characterization of hyporheic exchange and benthic biolayers in streams. *Water Resources Research*, *53*(2), 1575–1594.

Lewandowski, J., Arnon, S., Banks, E., Batelaan, O., Betterle, A., Broecker, T., Coll, C., Drummond, J., Gaona Garcia, J., Galloway, J., Gomez-Velez, J., Grabowski, R., Herzog, S., Hinkelmann, R., Höhne, A., Hollender, J., Horn, M., Jaeger, A., Krause, S., … Wu, L. (2019). Is the hyporheic zone relevant beyond the scientific community? *Water (Switzerland)*, *11*(11), 2230. https://doi.org/10.3390/w11112230

Magliozzi, C., Grabowski, R. C., Packman, A. I., & Krause, S. (2018). Toward a conceptual framework of hyporheic exchange across spatial scales. *Hydrology and Earth System Sciences*, *22*(12), 6163–6185. https://doi.org/10.5194/hess-22-6163-2018

Malzone, J. M., Lowry, C. S., & Ward, A. S. (2016). Response of the hyporheic zone to transient groundwater fluctuations on the annual and storm event time scales. *Water Resources Research*, *52*, 5301–5321. https://doi.org/10.1002/2015WR018056

Marion, A., Zaramella, M., & Packman, A. I. (2003). Parameter estimation of the transient storage model for stream–subsurface exchange. *Journal of Environmental Engineering*, *129*(5), 456–463. https://doi.org/10.1061/(asce)0733-9372(2003)129:5(456)

Marzadri, A., Tonina, D., & Bellin, A. (2011). A semianalytical three-dimensional process-based model for hyporheic nitrogen dynamics in gravel bed rivers. *Water Resources Research*, *47*, W11518. https://doi.org/10.1029/2011WR010583

McLachlan, P. J., Chambers, J. E., Uhlemann, S. S., & Binley, A. M. (2017). Geophysical characterisation of the groundwater–surface water interface. *Advances in Water Resources*, *109*, 302–319. https://doi.org/10.1016/J.ADVWATRES.2017.09.016

Montero, P., & Vilar, J. A. (2014). TSclust: An R package for time series clustering. *Journal of Statistical Software2*, *62*(1), 1–43.

Mueen, A., & Keogh, E. (2016). Extracting optimal performance from dynamic time warping. *Proceedings of the ACM SIGKDD international conference on knowledge discovery and data mining*:Vol. 13. (pp. 2129–2130). Association for Computing Machinery. https://doi.org/10.1145/2939672.2945383.

Nadoushani, S. S. M., Dehghanian, N., & Saghafian, B. (2018). A fuzzy hybrid clustering method for identifying hydrologic homogeneous regions. *Journal of Hydroinformatics*, *20*(6), 1367–1386. https://doi.org/10.2166/hydro.2018.004

Neilson, B. T., Chapra, S. C., Stevens, D. K., & Bandaragoda, C. (2010). Two-zone transient storage modeling using temperature and solute data with multiobjective calibration: 1. Temperature. *Water Resources Research*, *46*(12). https://doi.org/10.1029/2009WR008756

Park, P. J., Manjourides, J., Bonetti, M., & Pagano, M. (2009). A permutation test for determining significance of clusters with applications to spatial and gene expression data. *Computational Statistics and Data Analysis*, *53*(12), 4290–4300. https://doi.org/10.1016/j.csda.2009.05.031

Pidlisecky, A., Singha, K., & Day-Lewis, F. D. (2011). A distribution-based parameterization for improved tomographic imaging of solute plumes. *Geophysical Journal International*, *187*(1), 214–224. https://doi.org/10.1111/j.1365-246X.2011.05131.x

Poole, G. C., O'daniel, S. J., Jones, K. L., Woessner, W. W., Bernhardt, E. S., Helton, A. M., Standord, J. A., Boer, B. R., & Beechie, T. J. (2008). Hydrologic spiralling: The role of multiple interactive flow paths in stream ecosystems. *River Research and Applications*, *24*(7), 1018–1031.

R Core Team. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Rucker, D. F., Tsai, C. H., Carroll, K. C., Brooks, S., Pierce, E. M., Ulery, A., & Derolph, C. (2021). Bedrock architecture, soil texture, and hyporheic zone characterization combining electrical resistivity and induced polarization imaging. *Journal of Applied Geophysics*, *188*, 104306. https://doi.org/10.1016/j.jappgeo.2021.104306

Sassen, D. S., Hubbard, S. S., Bea, S. A., Chen, J., Spycher, N., & Denham, M. E. (2012). Reactive facies: An approach for parameterizing field-scale reactive transport models using geophysical methods. *Water Resources Research*, *48*(10), 10526. https://doi.org/10.1029/2011WR011047

Savoy, P., Appling, A. P., Heffernan, J. B., Stets, E. G., Read, J. S., Harvey, J. W., & Bernhardt, E. S. (2019). Metabolic rhythms in flowing waters: An approach for classifying river productivity regimes. *Limnology and Oceanography*, *64*(5), 1835–1851. https://doi.org/10.1002/lno.11154

Schmadel, N. M., Ward, A. S., & Wondzell, S. M. (2017). Hydrologic controls on hyporheic exchange in a headwater mountain stream. *Water Resources Research*, *53*(7), 6260–6278. https://doi.org/10.1002/2017WR020576

Singha, K., Pidlisecky, A., Day-Lewis, F. D., & Gooseff, M. N. (2008). Electrical characterization of non-Fickian transport in groundwater and hyporheic systems. *Water Resources Research*, *46*(4), 7. https://doi.org/10.1029/2008WR007048

Singley, J. G. (2022). *Example R scripts for unsupervised clustering of time-lapse electrical resistivity models from stream tracer injections*. Zenodo. https://doi.org/10.5281/zenodo.6945970

Stream Solute Workshop. (1990). Concepts and methods for assessing solute dynamics in stream ecosystems. *Journal of the North American Benthological Society*, *9*(2), 95–119. https://doi.org/10.2307/1467445

Tonina, D., & Buffington, J. M. (2007). Hyporheic exchange in gravel bed rivers with pool-riffle morphology: Laboratory experiments and three-dimensional modeling. *Water Resources Research*, *43*(1). https://doi.org/10.1029/2005WR004328

Tso, C.-H. M., Iglesias, M., Wilkinson, P., Kuras, O., Chambers, J., & Binley, A. (2021). Efficient multiscale imaging of subsurface resistivity with uncertainty quantification using ensemble Kalman inversion. *Geophysical Journal International*, *225*(2), 887–905. https://doi.org/10.1093/gji/ggab013

Vidon, P. G., & Hill, A. R. (2004). Landscape controls on the hydrology of stream riparian zones. *Journal of Hydrology*, *292*(1–4), 210–228.

Voltz, T., Gooseff, M., Ward, A. S., Singha, K., Fitzgerald, M., & Wagener, T. (2013). Riparian hydraulic gradient and stream-groundwater exchange dynamics in steep headwater valleys. *Journal of Geophysical Research: Earth Surface*, *118*(2), 953–969.

Wainwright, H. M., Chen, J., Sassen, D. S., & Hubbard, S. S. (2014). Bayesian hierarchical approach and geophysical data sets for estimation of

reactive facies over plume scales. *Water Resources Research*, *50*(6), 4564–4584. https://doi.org/10.1002/2013WR013842

Ward, A. S. (2016). The evolution and state of interdisciplinary hyporheic research. *Wiley Interdisciplinary Reviews: Water*, *3*(1), 83–103. https://doi.org/10.1002/wat2.1120

Ward, A. S., Fitzgerald, M., Gooseff, M. N., Voltz, T. J., Binley, A. M., & Singha, K. (2012). Hydrologic and geomorphic controls on hyporheic exchange during base flow recession in a headwater mountain stream. *Water Resources Research*, *48*(4), 4513. https://doi.org/10.1029/2011WR011461

Ward, A. S., Gooseff, M. N., & Singha, K. (2010a). Characterizing hyporheic transport processes - interpretation of electrical geophysical data in coupled stream-hyporheic zone systems during solute tracer studies. *Advances in Water Resources*, *33*(11), 1320–1330. https://doi.org/10.1016/j.advwatres.2010.05.008

Ward, A. S., Gooseff, M. N., & Singha, K. (2010b). Imaging hyporheic zone solute transport using electrical resistivity. *Hydrological Processes*, *24*(7), 948–953. https://doi.org/10.1002/hyp.7672

Ward, A. S., Gooseff, M. N., Fitzgerald, M., Voltz, T. J., & Singha, K. (2014). Spatially distributed characterization of hyporheic solute transport during baseflow recession in a headwater mountain stream using electrical geophysical imaging. *Journal of Hydrology*, *517*, 362–377. https://doi.org/10.1016/j.jhydrol.2014.05.036

Ward, A. S., Gooseff, M. N., Voltz, T. J., Fitzgerald, M., Singha, K., & Zarnetske, J. P. (2013). How does rapidly changing discharge during storm events affect transient storage and channel water balance in a headwater mountain stream? *Water Resources Research*, *49*, 5473–5486. https://doi.org/10.1002/wrcr.20434

Ward, A. S., Schmadel, N. M., Wondzell, S. M., Gooseff, M. N., & Singha, K. (2017). Dynamic hyporheic and riparian flow path geometry through base flow recession in two headwater mountain stream corridors. *Water Resources Research*, *53*(5), 3988–4003. https://doi.org/10.1002/2016WR019875

Ward, A. S., Schmadel, N. M., Wondzell, S. M., Harman, C. J., Gooseff, M. N., & Singha, K. (2016). Hydrogeomorphic controls on hyporheic and riparian transport in two headwater mountain streams during base flow recession. *Water Resources Research*, *52*, 1479–1497. https://doi.org/10.1002/2015WR018225

Ward, A. S., Singha, K., & Gooseff, M. N. (2020). *Hyporheic exchange studies in H.J. Andrews watersheds 01 and 03, summer 2010*. HydroShare. https://doi.org/10.4211/hs.8207c26f492e49f0be33e7a2427ccfea

Ward, A. S., Wondzell, S. M., Schmadel, N. M., Herzog, S., Zarnetske, J. P., Baranov, V., Blaen, P. J., Brekenfeld, N., Chu, R., Derelle, R., Drummond, J., Fleckenstein, J. H., Garayburu-Caruso, V., Graham, E., Hannah, D., Harman, C. J., Hixson, J., Knapp, J. L. A., Krause, S., … Wisnoski, N. I. (2019). Spatial and temporal variation in river corridor exchange across a 5th-order mountain stream network. *Hydrology and Earth System Sciences*, *23*(12), 5199–5225. https://doi.org/10.5194/hess-23-5199-2019

Warren Liao, T. (2005). Clustering of time series data - a survey. *Pattern Recognition*, *38*(11), 1857–1874. https://doi.org/10.1016/j.patcog.2005.01.025

White, D. S. (1993). Perspectives on defining and delineating hyporheic zones. *Journal of the North American Benthological Society*, *12*(1), 61–69.

Wondzell, S. M. (2006). Effect of morphology and discharge on hyporheic exchange flows in two small streams in the Cascade Mountains of Oregon, USA. *Hydrological Processes*, *20*(2), 267–287. https://doi.org/10.1002/hyp.5902

Wondzell, S. M., & Swanson, F. J. (1996). Seasonal and storm dynamics of the hyporheic zone of a 4 th-order mountain stream. 1: Hydrologic processes. *Journal of the North American Benthological Society*, *15*(1), 3–19.

Wroblicky, G. J., Campana, M. E., Valett, H. M., & Dahm, C. N. (1998). Seasonal variation in surface-subsurface water exchange and lateral hyporheic area of two stream-aquifer systems. *Water Resources Research*, *34*(3), 317–328.

Zarnetske, J. P., Haggerty, R., Wondzell, S. M., Bokil, V. A., & González-Pinzón, R. (2012). Coupled transport and reaction kinetics control the nitrate source-sink function of hyporheic zones. *Water Resources Research*, *48*(11). https://doi.org/10.1029/2012WR011894

Zimmer, M., & Lautz, L. K. (2013). Temporal and spatial response of hyporheic zone geochemistry to a storm event. *Hydrological Processes*, *28*(4), 2324–2337.

How to cite this article: Singley, J. G., Singha, K., Gooseff, M. N., González-Pinzón, R., Covino, T. P., Ward, A. S., Dorley, J., & Hinckley, E.-L. S. (2022). Identification of hyporheic extent and functional zonation during seasonal streamflow recession by unsupervised clustering of time-lapse electrical resistivity models. *Hydrological Processes*, *36*(10), e14713. https://doi.org/10.1002/hyp.14713