GEOPHYSICS, VOL. 86, NO. 5 (SEPTEMBER-OCTOBER 2021); P. WB29–WB49, 15 FIGS., 6 TABLES.
10.1190/GEO2020-0438.1

Hydrogeophysical comparison of hillslope critical zone architecture for different geologic substrates

Andrew D. Parsekian[1], Dario Grana[2], Felipe dos Anjos Neves[2], Mark S. Pleasants[3], Mark Seyfried[4], Bryan G. Moravec[5], Jon Chorover[5], Anthony M. Moraes[2], Natalie Y. Smeltz[2], John H. Westenhoff[2], and Thijs Kelleners[3]

ABSTRACT

The belowground architecture of the critical zone (CZ) consists of soil and rock in various stages of weathering and wetness that acts as a medium for biological growth, mediates chemical reactions, and controls partitioning of hydrologic fluxes. Hydrogeophysical imaging provides unique insights into the geometries and properties of earth materials that are present in the CZ and beyond the reach of direct observation beside sparse wellbores. An improved understanding of CZ architecture can be achieved by leveraging the geophysical measurements of the subsurface. Creating categorical models of the CZ is valuable for driving hydrologic models and comparing belowground architectures between different sites to interpret weathering processes. The CZ architecture is revealed through a novel comparison of hillslopes by applying facies classification in the elastic-electric domain driven by surface-based hydrogeophysical measurements. Three pairs of hillslopes grouped according to common geologic substrates — granite, volcanic extrusive, and glacially altered — are classified by five different hydrofacies classes to reveal the relative wetness and weathering states. The hydrofacies classifications are robust to the choice of initial mean values used in the classification and noncontemporaneous timing of geophysical data acquisition. These results will lead to improved interdisciplinary models of CZ processes at various scales and to an increased ability to predict the hydrologic timing and partitioning. Beyond the hillslope scale, this enhanced capability to compare CZ architecture can also be exploited at the catchment scale with implications for improved understanding of the link between rock weathering, hydrochemical fluxes, and landscape morphology.

INTRODUCTION

The critical zone (CZ) is frequently defined to encompass the near-surface earth system from the top of the plant canopy down to the base of circulating groundwater (Anderson et al., 2008). CZ science aims to reveal the influence of physical, chemical, and biological processes and interactions in the zone of the earth's crust that effectively supports life. The subsurface portion of the CZ is essentially analogous to the weathering zone plus aggregational and degradational processes, that is, the zone between the land surface and unaltered (unweathered) bedrock. There is currently no single unified theoretical framework that can predict the structure of the CZ, nor how the CZ controls hydrologic processes (Riebe et al., 2017).

Manuscript received by the Editor 29 June 2020; revised manuscript received 15 January 2021; published ahead of production 6 February 2021; published online 17 June 2021.
[1]University of Wyoming, Department of Geology and Geophysics and Department of Civil and Architectural Engineering, 1000 East University Avenue, Laramie, Wyoming 82071, USA. E-mail: aparseki@uwyo.edu (corresponding author).
[2]University of Wyoming, Department of Geology and Geophysics, 1000 East University Avenue, Laramie, Wyoming 82071, USA. E-mail: dgrana@uwyo.edu; fneves@uwyo.edu; anthmoraes@gmail.com; nsmeltz@uwyo.edu; jhwestenhoff@gmail.com.
[3]University of Wyoming, Department of Ecosystem Science and Management, 1000 East University Avenue, Laramie, Wyoming 82071, USA. E-mail: mpleasan@uwyo.edu; tkellene@uwyo.edu.
[4]U.S. Department of Agriculture, Agricultural Research Service, 251 East Front Street, Suite 400, Boise, Idaho 83702, USA. E-mail: mark.seyfried@ars.usda.gov.
[5]University of Arizona, Department of Environmental Science, 1177 East 4th Street Tucson, Arizona 85721, USA. E-mail: bmoravec@arizona.edu; chorover@arizona.edu.

The term "CZ architecture" (Leopold et al., 2013; Befus et al., 2011; Orlando et al., 2016; Brantley et al., 2017; Riebe et al., 2017; West et al., 2019; Moravec et al., 2020) has been used as a term roughly analogous to structure. Here, we build upon and clarify this definition and use it to mean "the geometries and material properties associated with the CZ subsurface" with implied relevance to the corresponding subsurface hydrologic properties such as porosity and saturation. The makeup of the CZ subsurface is determined by many factors such as the parent material, age of uplift, glaciation history, geochemical weathering, incision/deposition processes, and bioturbation. Direct observations of the subsurface are generally limited to road cuts, well logs, and soil pits that are labor-intensive and/or have limited spatial coverage. As a result, subsurface structure and properties are often estimated using sparse information. Here, we take a combined hydrogeophysical and rock-physics approach to enable comparisons of subsurface hydrologically relevant properties of CZ architecture between similar and contrasting hillslopes.

Near-surface geophysics has emerged as a powerful tool to image the subsurface and provide information on subsurface properties and geometries that benefit our understanding of the hydrologic process for CZ investigations (Parsekian et al., 2015). The translation from geophysical properties to material properties involves the use of rock-physics relationships. For example, Archie's law may be used to translate resistivity to water content (Binley et al., 2015). When appropriate calibration information is available, these rock-physics relationships can provide the basis for meaningful interpretations (Flinchum et al., 2018a).

In mountain environments, hydrogeophysical measurements are frequently conducted at the hillslope scale (Leopold et al., 2013; Befus et al., 2011; Holbrook et al., 2014; Olyphant et al., 2016). Hillslopes are generally considered the smallest meaningful hydrologic unit because they control the transformation of precipitation inputs into vertical and lateral water fluxes (Loritz et al., 2017). Hydrologic processes in watersheds in sizes up to the mesoscale (<50 km^2) are dominated by hillslope behavior (Robinson et al., 1995). The concept of hillslope hydraulic connectivity is used to explain the relationship between hillslope soil moisture and streamflow generation (McNamara et al., 2005). Understanding the subsurface structure of mountain hillslopes thus has implications for developing conceptual models of hydrologic partitioning that are grounded in understanding the hydraulic connections in the CZ (McIntosh et al., 2017). Furthermore, interrogating the subsurface helps resolve hydraulic properties such as storage and routing, as well as conceptualizing geochemical transformations along flow paths (White et al., 2019). Although incorporating geophysical images into subsurface parameterizations for hydrologic models has been attempted (Binley et al., 2004), unclassified images do not necessarily improve the ability to retrieve hydrologic parameters (Binley et al., 2010). Furthermore, fully distributed parameters from more than one geophysical image type (e.g., seismic and resistivity) are not easily jointly implemented into modeling frameworks without relying on rock-physics models such as Archie's law or Hertz-Mindlin, which are difficult to calibrate across 2D or 3D geophysical images. More specifically, having a simplified categorical definition of subsurface material geometry and properties is useful for hydrologic interpretations such as flow modeling (Farmani et al., 2008; Claes et al., 2020) and is of interest in this study.

Many past studies have used geophysical techniques to analyze subsurface structure in mountain hillslopes (Mills, 1990; Leopold et al., 2013). Befus et al. (2011) use shallow seismic refraction tomography (SSR) transects to determine the depth to saprolite, fractured rock, and unaltered rock for two small watersheds in the Boulder Creek Critical Zone Observatory (CZO). It was found that saprolite started at approximately 1 m depth and fractured rock started at approximately 6 m depth. The unaltered Precambrian igneous and metamorphic rock started at approximately 12 m depth. Leopold et al. (2013) use electrical resistivity tomography (ERT) transects to study subsurface structure in the same watersheds. The estimated depth to fractured bedrock of 4.3–5.8 m by ERT was 0.5–1.5 m shallower compared to the seismic survey. Overall, the results between the seismic and ERT surveys agreed, with both techniques showing high lateral variability. Olyphant et al. (2016) use SSR to classify a volcanic substrate into soil, regolith, and bedrock, interpreting bedrock between 4 and 60 m below the surface.

Although independent geophysical methods may be useful for hydrologic interpretations of the CZ, particular advantages emerge from combining various geophysical measurements that provide complementary information. For example, seismic surveys may provide information about the makeup of the CZ (i.e., soil, saprolite, fractured bedrock, and solid bedrock), while time-lapse ERT measurements can be used to study subsurface moisture dynamics through time (Thayer et al., 2018). Holbrook et al. (2014) use SSR and ERT to study CZ structure across a forested slope and a swampy meadow in the Southern Sierra CZO, east of Fresno, CA. A rock-physics model, linking porosity and bedrock mineralogy to seismic velocity, was used to estimate the spatial distribution of subsurface porosity. Comparison with observed water contents and porosities from hand auger and Geoprobe samples, obtained coincidently with the geophysics data, indicated that the top 3 m of the subsurface was mostly dry whereas saturation appeared at depths >10 m. The results of the seismic and ERT surveys indicated that the depth to unaltered bedrock ranged from 10 to 35 m, with an average depth of 23 m.

Simple, qualitative, and concurrent interpretation of colocated geophysical measurements has proved useful, although it is at least somewhat subjective due to the qualitative choices of the interpreter (Thayer et al., 2018; Kotikian et al., 2019). Joint geophysical inversion optimizes a subsurface model of physical properties (i.e., permittivity, velocity, and resistivity) given two different input geophysical data sets assuming some structural similarity (Gallardo and Meju, 2003), whereas joint rock-physics inversion identifies a subsurface model of material properties (i.e., porosity, saturation, water content, etc.) that satisfies both data sets by using rock-physics relationships (Grana, 2018). It is a nontrivial task to quantitatively link different geophysical data sets through any joint approach, and each has strengths and weaknesses. For our purposes, for the categorizing, delineating, and simplifying of subsurface structures such as hydrofacies (Anderson, 1989) to be useful for modeling studies, the end result of Geoprobe relative material property classes is most desirable; therefore, we use joint classification in the elastic-electric domain for this work. This approach uses independent geophysical inversions as inputs and produces a single classified image.

Statistical and machine learning approaches, such as discriminant analysis and *k*-means clustering, have been applied previously to lithologic and fluid facies classification problems. Clustering methods (Hastie et al., 2002) generally allow classifying a group of samples based on a set of measurements, such as geophysical data or estimated rock and fluid properties. Paasche et al. (2006) propose a fuzzy *c*-

means method for the classification of lithologic facies based on petrophysical parameters estimated from geophysical data. Paasche et al. (2010), Paasche and Eberle (2011), and Sun and Li (2016) apply fuzzy logic methods to various geophysical data sets and compare it to other clustering methods. Bedrosian et al. (2007) use clustering analysis to derive the geologic structure of the subsurface using geophysical data. Statistical methods such as the expectation-maximization (EM) method for multimodal distributions (Hastie et al., 2002) have also been used for facies classification problems in exploration geophysics (Grana and Della Rossa, 2010; Grana et al., 2015) and for zonation problems in near-surface geophysics (Doetsch et al., 2010).

In this work, we adopt the EM method as a semisupervised algorithm, where instead of randomly choosing the prior parameters of the statistical model as in many data science applications, we select them according to the available geologic information and rock-physics reference values, to reduce the nonuniqueness of the classification problem. Compared to deterministic solutions of classification methods, the use of EM for multimodal distributions allows quantifying the uncertainty in the predicted classification that should be considered in the decision-making process. Our focus is on classification of facies based on fluid-relevant properties — that is, hydrofacies, spatially defined hydrologic units — rather than zonation — to discriminate/classify the vertical and lateral variations of fluid (air and water) and lithology (soft and stiff rocks) in the near surface. Joint rock-physics inversion can produce models of material properties that satisfy multiple geophysical data sets; however, to achieve our objective of producing simplified subsurface models that adhere to conceptual models of subsurface hillslope and catchment architecture (e.g., spatial continuity, location of active weathering processes), classification in the elastic-electric domain is required. The proposed implementation of the EM algorithm assumes that the distribution of the geophysical properties is Gaussian within each hydrofacies (Grana et al., 2017; Astic et al., 2020), resulting in a Gaussian mixture model (Grana and Della Rossa, 2010; Grana et al., 2017). The statistical classification does not make use of an explicit rock-physics model that is hard to calibrate. We use the rock-physics model to generate the training data set, and then we classify the elastic and electrical properties estimated from measured geophysical data using the EM clustering method. This approach avoids issues with limited direct "deep" subsurface sampling for calibration, while quantitatively integrating the SSR and ERT data sets. It also allows for reliable, independent geophysical inversions.

This investigation is novel due to the combination of ERT and SSR that has been only used in limited past classification studies (Hachmöller and Paasche, 2013), the application to CZ science, and — most importantly — our study highlights a comparison of geologically grouped classified properties between three pairs of sites. The purpose of this investigation was to reveal the hydrologically relevant material properties and geometries (i.e., the hydrogeophysical properties) of the CZ associated with the three geologic substrates as the context for future hydrologic investigations. We were motivated by (1) the relative lack of multidimensional subsurface property maps available to CZ scientists for hydrologic interpretations, (2) the rarity of controlled studies making comparisons between geographically distinct CZ sites, and (3) a scarcity of examples quantitatively integrating colocated hydrogeophysical data sets particularly in the context of CZ science.

Our objectives were to (1) reveal the relative distribution of subsurface material properties at each hillslope, (2) expose the relative spatial wetness of the substrate at each hillslope, (3) compare the geometries and properties within and across geologic parent materials, and (4) contribute to the growing body of literature demonstrating statistical approaches for joint analysis of geophysical images in the absence of detailed extracted samples. Through these objectives, we address the following questions. (1) What are the hydrogeophysical properties of hillslopes underlain by glacially deposited, weathered granite, and volcanic geologic materials? (2) How is the geologic history linked to modern hydrogeophysical properties of the substrate? (3) What is the spatial variability in the near subsurface of each site in the context of structures, geometries, relative wetness, and materials?

STUDY SITES AND METHODS

Study sites

We selected six representative hillslopes (Table 1) paired across the three characteristic geologic substrates: old granite (GR), young extrusive volcanics (VO), and glacially altered (G). Geophysical transect locations are indicated in Appendix A and in the data set associated with this work (Parsekian et al., 2020). We focused on CZO sites at Reynolds Creek in Idaho and Jemez River Basin in New Mexico and three additional sites in Wyoming. The sites were chosen based on accessibility, the availability of geophysical data sets, and because they are all in snowmelt-dominated catchments. Because each of the sites has been established for several years or more before this study, they have informal names; however, we have used simple geologic/geographical identifiers here to enable easy comparison (Table 1). Most of the hillslopes are generally south-facing except for the GR_ID and VO_ID sites that are northeast-facing. Due to their relatively low elevations, the GR_ID and VO_ID sites receive meltwater inputs from snowdrifts, whereas snowpack at the other four higher elevation sites is more continuous. Additional details on the surface, geographic, and geologic context for each hillslope may be found in Table 1. More information on the hydrologic dynamics of each site can be found in Fullhart et al. (2018), Thayer et al. (2018), and Pleasants et al. (2017) (Gs_WY); Moraes (2019) (Gn_WY); Seyfried et al. (2018) (GR_ID and VO_ID); and Kotikian et al. (2019) (GR_WY); Moravec et al. (2020) and White et al. (2019) (VO_NM). Although the sites have a variable and unknown long-term climate history as well as contrasting vegetation communities, we assume that the geologic substrate is the first-order architectural control for these moderate angle, semiarid hillslopes (Salvucci and Entekhabi, 1995; Onda et al., 2006). Although climate and meteoric water inputs are central controls on hydrology and weathering processes that drive the evolution of subsurface CZ properties and geometries, the timescales over which we know about local climate by direct measurement of approximately 0.1 ka and the timescales that we can reliably predict past local hydroclimate of approximately 10 ka (Liefert and Shuman, 2020) are short in comparison to the timescales of many mineral weathering processes of 100–10,000 ka (White and Brantley, 2003). Although we report modern estimates of hydrologic inputs (Table 1), these are primarily for context and should be recognized to not necessarily represent variability over longer timescales that may be important for weathering processes.

The availability of direct subsurface observations varies substantially between sites ranging from only shallow soil pits (GR_ID) to >40 m deep wells with geophysical logs and geochemical

Table 1. Site parameters. After the geologic prefix indicated in the text, the U.S. state abbreviation of each site's location is appended (ID, WY, NM). To discriminate the two G sites that are both in Wyoming, we use an "n" and "s" after the geologic prefix, indicating that the site is located in the north or south region of the state, respectively. Contextual site references and mean-annual precipitation (MAP): [1](Kotikian et al., 2019), [2](Nielson et al., 2015), [3](Moravec et al., 2020), [4](Seyfried et al., 2018), [5](Thayer et al., 2018), and [6](Moraes, 2019). RCCZO is Reynolds Creek Critical Zone Observatory, JRBCZO is the Jemez River Basin Critical Zone Observatory, and ZOB is zero-order Basin. The age column refers to the timing of the exposure or landscape reset.

Site ID	Informal name	Geologic parent material	Age (Ma)	Latitude	Longitude	Elevation (m)	Slope aspect	Vegetation	MAP (mm yr^{-1})
GR_WY	Headquarters[1]	Granite	>55	41.22	−105.39	2570	South-southeast	Rangeland	610
GR_ID	Johnston draw (RCCZO)[2]	Granite	>66	43.12	−116.79	1720	East-northeast	—	620
VO_NM	Mixed conifer ZOB (JRBCZO)[3]	Tuff	1.6–1.1	35.88	−106.54	3015	West-southwest	Fire-impacted forest	700
VO_ID	Upper sheep creek (RCCZO)[4]	Olivine basalt	7.2–3.6	43.12	−116.72	1890	Northeast	Rangeland	480
Gs_WY	No-name[5]	Glacial deposits; gneiss	0.012	41.34	−106.21	2940	South-southwest	Pine forest	860
Gn_WY	Spear-O[6]	Glacial deposits; granite	0.012	44.54	−107.20	2640	Southwest	—	690

characterization (VO_NM). Gs_WY has a 3.1 m borehole and soil pits, and GR_WY and Gn_WY have boreholes <2 m in depth.

Geophysical approach

The key characteristic of this investigation is the joint analysis of two geophysical measurements that respond to distinctly different subsurface physical properties yet can be deployed to have generally similar spatial resolution. The use of SSR and ERT together enables more confident interpretation of the material properties at any given point in the subsurface compared to their isolated use. Each measurement effectively helps to explain the poorly constrained traits of the other. For example, the compressional wave (P-wave) velocity (V_P) of a material only varies 0.030 km s^{-1} in the range of 0%–99% saturation for soft sands (Bachrach and Nur, 1998), whereas the electrical resistivity (ρ) is highly sensitive to saturation (Miller et al., 2008; Hübner et al., 2015; Claes et al., 2019). Having both pieces of information colocated across the subsurface constrains the number of physically plausible spatial distributions of material properties that can be produced. The concept of correlating electrical and acoustic data sets was broadly introduced for near-surface geophysical imaging by Meju et al. (2003), followed by integrated geostatistical applications (Hachmöller and Paasche, 2013), as well as qualitative cointerpretation (Kotikian et al., 2019) on a site-specific basis. Here, we measure the two data sets on the same line in the field, regrid to a rectilinear 0.5 m grid, and spatially colocate the tomographically inverted results for statistical analyses and interpretation. We keep only the grid values that are within the calculated maximum depth of investigation of each measurement. Although these measurements are sensitive to near-surface soil properties, due to the inherent spatial resolution of ERT and SSR and regridding to a 0.5 m mesh, shallow soils may only be represented by the top one or two rows of pixels. We apply the joint classification in the elastic-electric domain to spatially partition the image given the colocated input physical properties. Given the available geophysical data, this leads to a classification naturally driven by the relative wetness and relative stiffness. To achieve geologic plausibility, the classification seeks to group similar material properties. This follows the principle of geologic "facies," or units of rock with similar characteristics. However, given the material properties that we determine are not exclusive to petrographic rock properties, we instead use the term "hydrofacies" for these classifications, a term coined by Anderson (1989) who indicates a spatially defined hydrologic unit with distinct hydrologic properties.

Seismic refraction

Seismic refraction data sets were measured in the summer of several years, between mid-June and mid-August, as indicated in Table 2. Each survey used four 24-channel Geometrics Geode systems with 40 Hz vertical component geophones except for Gn_WY, which used only two Geodes (Table 2). We used a 5.4 kg sledgehammer striking a steel plate to generate the seismic source, with 8–24 stacks at each shot point. We processed the seismic data using the Geogiga software package (Geogiga Technology Corp.) for picking and inversion using a 4:1 horizontal to vertical smoothing ratio (based on the assumption that weathering and depositional processes result in a general trend toward horizontal layering rather than the opposite), a 0.5 m quadrilateral mesh grid size, and a maximum of five iterations. Using these parameters, all of the inversions

converged with a root-mean-square (rms) misfit of 2 ms (Table 2). The depth of investigation was masked for all tomograms using the deepest calculated raypath (see Appendix B). We estimated the sensitivity of the seismic images following the approach of Holbrook et al. (2014), using a suite of 10 starting models for each tomogram (see Appendix B).

ERT

ERT measurements were made on the same day of the same year as the corresponding seismic measurement at each site, except for VO_ID where we used ERT data from 2019 and seismic data collected in 2014 as a part of previous work (Seyfried et al., 2018). The ERT data from the VO_ID site were measured on the same day of 2019 as the seismic data were measured in 2014; therefore, we assume that the subsurface hydrologic conditions were similar during these two seasons (2014 and 2019 were ±5% of MAP based on the accumulated precipitation measured at the Reynolds Creek SNOTEL site during each water year until the date of measurement). Each data set was acquired with a 4PointLight 10w (Lippmann Geophysical Instruments) single-channel resistivity meter, except for Gs_WY, which was measured with a DAS1 (Multiphase Technologies, Sparks, NV) because that site was initiated during an earlier project (Thayer et al., 2018). Previous work demonstrates that different ERT instruments will produce statistically equivalent measurements (Parsekian et al., 2017); therefore, we assume that our data sets are comparable. The electrode spacing was 1 m (except for Gs_WY; 1.5 m) and a Wenner array was used for all measurements. Table 2 summarizes all of the measurement parameters. ERT data were inverted using R2 (Binley, 2015) with a custom error model for each site based on observed data uncertainties following Robinson et al. (1995). The X^2 goodness of fit for each tomogram was between 1.00 and 1.41 (Table 2) indicating a good match between the model and the data. The depth of investigation was calculated using the approach of Oldenburg and Li (1999) (see Appendix B).

Theory for the EM method

To produce material properties based jointly on two geophysical images, we use the EM method for Gaussian mixture models (Hasselblad, 1966; Dempster et al., 1977; Hastie et al., 2002). To begin, we assume a multivariate Gaussian distribution $\mathcal{N}_M(\boldsymbol{d};\boldsymbol{\mu}_{d|k},\boldsymbol{\Sigma}_{d|k})$ of the geophysical attributes $\boldsymbol{d}$ for each hydrofacies k (for $k = 1, \ldots N$), with the mean $\boldsymbol{\mu}_{d|k}$ and the covariance matrix $\boldsymbol{\Sigma}_{d|k}$. Therefore, the probability distribution of $\boldsymbol{d}$ is a Gaussian mixture model of the form

$$f(\boldsymbol{d}) = \sum_{k=1}^{N} \pi_k \mathcal{N}_M(\boldsymbol{d};\boldsymbol{\mu}_{d|k},\boldsymbol{\Sigma}_{d|k}), \tag{1}$$

where π_k are the weights of the components for a Gaussian mixture which correspond, following our assumption, to the probability of each class.

The EM method is an iterative algorithm based on two steps. In the expectation step, given the current configuration of the estimated parameters, we compute the expectation of the likelihood function. In the maximization step, we maximize the expected likelihood and update the parameter configuration. The goal is to estimate the set of hyperparameters $\boldsymbol{\theta}$:

$$\boldsymbol{\theta} = (\pi_1, \ldots, \pi_{N-1}, \boldsymbol{\mu}_{d|1}, \ldots, \boldsymbol{\mu}_{d|N}, \boldsymbol{\Sigma}_{d|1}, \ldots, \boldsymbol{\Sigma}_{d|N}), \tag{2}$$

to maximize the log-likelihood of the M measured data points

$$\begin{aligned}\ell(\boldsymbol{\theta},\boldsymbol{d}) &= \sum_{j=1}^{M}\sum_{k=1}^{N} \log(\pi_k f_{\boldsymbol{d}|k}(\boldsymbol{d}_j)) \\ &= \sum_{j=1}^{M}\sum_{k=1}^{N} \log(\pi_k \mathcal{N}_M(\boldsymbol{d}_j;\boldsymbol{\mu}_{d|k},\boldsymbol{\Sigma}_{d|k})),\end{aligned} \tag{3}$$

with the condition that $\pi_1 + \ldots + \pi_{N-1} = 1 - \pi_N$. To maximize the log-likelihood function $\ell(\boldsymbol{\theta},\boldsymbol{d})$, we define an auxiliary variable $\boldsymbol{\kappa} = [\kappa_1, \ldots, \kappa_{N_F}]^T$, such that, for a given $j \in \{1, \ldots, M\}$, $\kappa_i = 1$ if $\boldsymbol{d}_j$ comes from $f_{\boldsymbol{d}|i}(\boldsymbol{d}_j)$; otherwise, $\kappa_i = 0$ for $i = 1, \ldots, N_F$. If the values of the variable $\boldsymbol{\kappa}$ are known, the maximum likelihood estimates of $\boldsymbol{\mu}_{d|k}$ and $\Sigma_{d|k}$ would be the sample mean and covariance matrix estimated from the data $\boldsymbol{d}_j$ belonging to the kth component (i.e., $\kappa_{k,j} = 1$) and the estimate of π_k would be the proportion of samples such that $\kappa_{k,j} = 1$. Because the values of the variable $\boldsymbol{\kappa}$ are unknown, we substitute the unknown values $\kappa_{k,j}$ with its expected value $\gamma_{k,j}(\boldsymbol{\theta}) = \mathrm{E}(\kappa_{k,j}|\boldsymbol{\theta},\boldsymbol{d}_j) = P(\kappa_k = 1|\boldsymbol{\theta},\boldsymbol{d}_j)$. In the maximization step, the so-defined expectations are used to update the estimates of the parameters. The expectation and maximization

Table 2. Acquisition and analysis parameters. The terms X^2 and rms are the goodness-of-fit parameters for the ERT and SSR inversions, respectively. The terms a and b are the coefficients of the linear error model used in ERT inversion.

	ERT							SSR			
Site	ERT instrument	Electrode spacing (m)	# of electrodes	Date of measurement	a	b	X^2	Geophone spacing (m)	# of geophones	Date of measurement	rms (ms)
GR_WY	Lippmann 4PL	1	64	6/16/2017	0.020	0.032	1.15	0.5	96	6/16/2017	2.4
GR_ID	Lippmann 4PL	1	64	7/24/2019	0.020	0.040	1.00	1	96	7/24/2019	1.5
VO_NM	Lippmann 4PL	1	64	6/18/2019	0.010	0.025	1.15	1	96	6/18/2019	1.7
VO_ID	Lippmann 4PL	1	64	8/17/2019	0.010	0.025	1.41	1	96	8/17/2014	2.2
Gs_WY	MPT DAS1	1.5	56	7/7/2016	0.020	0.040	1.28	1	96	7/7/2016	1.4
Gn_WY	Lippmann 4PL	1	60	7/9/2018	0.020	0.040	1.00	1	48	7/9/2018	1.8

steps are then repeated until convergence to obtain the optimal set of the parameters $\boldsymbol{\theta}_{ML} = \text{argmax}_{\boldsymbol{\theta}} \ell(\boldsymbol{\theta}, \boldsymbol{d})$ that fits the Gaussian mixture model to the data.

The EM method requires an initial guess for the parameters in $\boldsymbol{\theta}$, which is chosen according to the available site-specific geologic information and literature data. Although a "global" set of starting parameters may seem desirable for reducing bias and maximizing portability of the algorithm, in practice the variability in site geophysical and material properties is too large for a global starting point to be effective. A key requirement of this method is the human input to define plausible properties for categorization. The hydrofacies classification is then obtained according to a Bayesian classification method. The posterior distribution of each class, given the geophysical data, $P(k|\boldsymbol{d})$, is then obtained by applying Bayes's rule:

$$P(k|\boldsymbol{d}) = \frac{P(\boldsymbol{d}|k)P(k)}{P(\boldsymbol{d})} = \frac{\mathcal{N}_M(\boldsymbol{d}; \boldsymbol{\mu}_{d|k}, \boldsymbol{\Sigma}_{d|k})\pi_k}{\sum_{h=1}^{M} \pi_h \mathcal{N}_M(\boldsymbol{d}; \boldsymbol{\mu}_{d|h}, \boldsymbol{\Sigma}_{d|h})}, \quad k = 1, \ldots, N. \tag{4}$$

At each location, the predicted classification is then computed as the argument of the maximum of $P(k|\boldsymbol{d})$ for $k = 1, \ldots, N$, that is, the class for which the expression in equation 4 attains its maximum value.

EM hydrofacies classification

The statistical algorithm used for hydrofacies classification is the EM method that identifies the maximum a posteriori (maximum likelihood) estimates of the parameters of interest in complex data sets in which the model depends on an unobserved categorical variable (Hastie et al., 2002). In our application, the method is used to estimate the parameters of a Gaussian mixture model for a geophysical data set $\boldsymbol{d}$ composed of seismic (elastic) and electrical properties, namely, V_P and ρ, estimated from geophysical measurements. Based on this model, we can then compute the probability distribution of hydrofacies conditioned by geophysical data, that is, the probability for each data point belonging to a given hydrofacies and classify that hydrofacies.

We perform the classification on hillslope pairs grouped by geologic history.

The EM approach can be applied to an arbitrary number of classes; however, in this case — aiming to balance simplicity and geologic plausibility — we allow for division in up to five classes as guided by a basic abstraction of the rock-physics parameter space. As a conceptual model, we consider the domain of resistivity and seismic velocity and then split this space into four quadrants: high V_P, high ρ; high V_P, low ρ; low V_P, high ρ; and low V_P, low ρ. Given that resistivity is a log-spaced parameter whereas velocity is linear, the quadrants may not be equal in space. For data sets with highly skewed distributions, a normal score transformation could be applied (Deutsch and Journel, 1992). Given the anticipated material properties, these quadrants then approximately correspond to hydrofacies: stiffer and drier (DSt), stiffer and wetter (WSt), softer and drier (DSo), and softer and wetter (WSo). We assign relative properties, that is, "drier" versus "wetter," because we are not particularly concerned with the quantifiable wetness of a subsurface point; rather, we prioritize the classification. We assume that higher resistivity is generally associated with drier materials and lower resistivity is associated with wetter materials in a spatial sense; however, this is not strictly the case given the heterogeneity of the subsurface material properties (e.g., porosity). The spatial continuity of the facies is not explicitly enforced; rather, it is implicit in the geophysical images. An advantage of the joint rock-physics classification is that the seismic velocities partially reveal this spatial heterogeneity, so spatially variable wetness can be more confidently inferred. The terms "stiffer" and "softer" may at first appear counterintuitive; however, these are typical terms that we retain for compatibility with the rock-physics literature. These terms can also be thought of as representing "less weathered" and "more weathered." The soft/stiff terms also help alleviate ambiguity between granular porosity and fracture porosity, both of which are relevant to defining weathering, but they have very different structural implications. For a given lithology, a stiffer rock is less weathered and retains stiffness due to less mass loss per unit volume — this is likely to be weathered in place, fractured rock, immobile regolith, or even fresh bedrock. A softer rock is more weathered (or has more primary porosity, as in the case of vesicular volcanic rocks), has experienced more mass loss per unit volume, and may be considered "mobile regolith (i.e., weathered rock)" and/or soil. We avoid attempting quantitative transformations to porosity and/or saturation because these parameters are poorly defined without calibration. Our approach enables the classification and evaluation of material and geometric patterns without suffering from the limitations of extremely sparse direct observations. We recognize that the chosen hydrofacies classes (i.e., DSt, WSt, DSo, and WSo) are not associated directly with numerical values of hydrologic properties; however, the use of statistical classification means that the determination of classes based on geophysical input data is quantitative. The fifth category in our classification is "unweathered bedrock," corresponding in principle to the highest velocity and resistivity, but distinct from the DSt class due to the lack of weathering. It is important to recognize that, although we allow these five categories into the EM algorithm, it is not required to make use of all five for a given data set. Although our five a priori selected classes focus on the logical quadrants of the velocity-resistivity elastic-electric space, cases may exist in which more or fewer classes would be appropriate. Although we do not attempt here to explore the sensitivity of our results to the number of allowed classes, we highlight that future applications of this method may wish to do so if there is specific evidence about the substrate to justify it.

Although after EM classification we report the probability of the most likely class at each pixel in each profile, this statistic is of limited utility because it cannot capture information about the probability of each of the other four possible classes at the same locations. Therefore, we also report the information entropy (Shannon, 1948; Mavko and Mukerji 1998; Grana et al., 2012) that contains information about all probabilities at each point in the profile. As a measurement of uncertainty, entropy is appropriate for categorical variables and exceeds standard measures such as variance and covariance. For the vector of five probabilities ($\boldsymbol{c}$) produced by the EM classification, entropy (h) is calculated by

$$h = -\boldsymbol{c}^{\boldsymbol{T}} \log_M \boldsymbol{c} = -\sum_{i=1}^{M} c_i \log_M c_i, \tag{5}$$

where M is the number of facies and the logarithm is calculated in base M because there are M facies classified and this therefore normalizes the range of entropy to be between 0 and 1. Because all five facies

probabilities are contained within $\boldsymbol{c}$, this entropy calculation results in a value of one if there is an equal probability of all facies (i.e., $\mathbf{c}$ = [1/5 1/5 1/5 1/5 1/5]) which would correspond to maximum entropy and therefore maximum uncertainty. In contrast, if one facies has a probability of one, and the other four have probability of zero, this results in a minimum entropy of $h = 0$ and no associated uncertainty.

One of our objectives is to produce classified images of the subsurface that may be useful in driving process-based models. Although the geometries of the CZ architecture retrieved through the classifications are useful direct inputs into models (e.g., hydrologic flow models), our defined classes are relative and alone do not produce specific estimates of material properties. Nonetheless, we envision that this approach leads to flexibility for the end user to estimate and assign properties to these classes based on site knowledge, for example, using generalized resources such as the Rosetta library (Schaap et al., 2001).

Synthetic hillslope

We used a fictitious synthetic hillslope with numerical forward models of ERT and seismic data to demonstrate that the classification method can recover the input hydrofacies. We started by defining the porosity and saturation material properties for each facies, and then for each location in the synthetic subsurface we used Archie's law to predict resistivity and the Hertz–Mindlin theory with the Hashin–Shtrikman elastic bounds (Mindlin, 1949; Hashin and Shtrikman, 1963) to predict seismic velocity. Next, we used the forward modeling capabilities in R2 and PyGIMLi (Rücker et al., 2017) to calculate raw geophysical data with 2% noise (applied to resistances and traveltimes, respectively) based on the synthetic hillslope properties, and we finally inverted the modeled data sets and applied EM classification using identical steps that were used for the field data sets. Please see Appendix C for details on the rock physics and forward modeling.

Validation data

Although the highest quality field validation data including deep borehole logs of resistivity and seismic velocity is only available at VO_NM, we also use shallow borings available at three other sites for comparisons with our classification results. At Gs_WY and GR_WY, borings (3 and 2 m deep, respectively) were installed with a backpack-portable drill rig and geologic descriptions were based on the retrieved core and cuttings. The water level was measured using dataloggers with direct observations for calibration. We used a hand auger to install a 2 m boring at Gn_WY. At Gn_WY and GR_WY where a 5.7 cm diameter PVC casing was available, we used nuclear magnetic resonance (NMR) logging (Dart, Vista Clara, Mukilteo, WY) to measure the water content logs. At GR_WY and Gs_WY, the total depth of the borings was limited by refusal, and, at Gn_WY, the total depth was limited by the available auger stem. Logging NMR uses electromagnetic pulses and a permanent background magnetic field to directly measure the volumetric water content. Details about the measurement can be found in Walsh et al. (2013). At GR_WY, logs were obtained at 0.25 m intervals, and, at Gn_WY, logs were obtained at 0.125 m intervals. In both cases, the data were processed with a moving window filter across three intervals and a regularization factor of 100 using the interpretation software provided by the manufacturer.

RESULTS

Synthetic hillslope

The synthetic hillslope had measurement geometries similar to the field data and a plausible distribution of subsurface hydrofacies (Figure 1a). The result of the EM classification (Figure 1b) shows a distribution of classified facies that is representative of the known input plus some effects of smoothing (i.e., inversion regularization) and resolution (i.e., reduction in the ability to resolve features as a function of depth) that are consistent with the calculated measures of uncertainty (Figure 1c and 1d). For example, the entropy image (Figure 1d) shows that the zones of higher entropy (i.e., more uncertain classification) become larger as a function of depth. A simplistic pixel-wise comparison between the input and classified images (Figure 1a versus Figure 1b) indicates that 45% of the pixels were correctly classified. For comparison, when using synthetic data with no noise, a correct classification rate of 61% is achieved.

Independent interpretation of geophysical data

The seismic structure at the field sites consisted generally of surface-parallel isocontours that increase the velocity as a function of depth across all hillslopes. (Figure 2). The GR_ID hillslope revealed the most lateral V_P variability in comparison to the other hillslopes. The vertical velocity gradient was steepest at Gs_WY (>0.4 km s^{-1} m^{-1}), whereas VO_NM and Gn_WY had low vertical velocity gradients (<0.04 km s^{-1} m^{-1}). The electrical structures at each site vary substantially in resistivity range and geometry (Figure 2). For example, GR_ID and VO_ID feature obvious lateral discontinuities, whereas the other four sites have a "layered" surface-parallel structure. Gs_WY, GR_WY, and GR_ID showed resistivities in excess of 10^3 ohm-m, whereas VO_NM showed resistivities as low as 20 ohm-m. The maximum depth of ERT investigation across the sites was approximately 5–7 m below the surface, depending on the resistivity structure. The deepest seismic raypath was deeper than the ERT sensitive zone at all sites except Gn_WY due to the shorter seismic line used at that site.

Rock-physics and hydrofacies classification

The rock-physics crossplots relating resistivity to seismic velocity are the most straightforward way to interpret similarities and differences in geophysical properties across all sites (Figure 3a–3d). These plots in the "V_P-ρ elastic-electric space" are useful for viewing the variability in ρ and V_P across sites and for evaluating the outcome of the classification algorithm. In general, we observe a substantial range in the overall geometric patterns for each geologic pairing that have distinct geometric patterns. These crossplots show that the classification algorithm is not constrained to grouping in the quadrants that are characteristic of our abstraction of the parameter space, and that classification appears to be driven by the local material properties. The probability associated with the classified facies at each point shows gradual transitions between classes (Figure 3e–3g).

The VO_NM site is the only location where well-described lithology and geophysical logs are available immediately adjacent to our geophysical transects. Therefore, in this case (Figure 3b), we show the corresponding well-log-derived physical properties and the associated lithologic descriptions from Moravec et al. (2020). In this example, the algorithmic classification break between DSo

and WSo falls close to the interpreted change in lithology between soil/saprolite and altered tuff.

We compute the probabilities for each facies conditioned by the geophysical measurements for each location (Figure 3), and we viewed these spatial 2D images (Figure 4, the left column) with associated frequency distributions of classified properties (Figure 4, the right column). The material classifications for each hillslope enable interpretation of heterogeneity and geometries of near-surface structures. Slope-parallel geometries are common to all hillslopes. One metric to judge heterogeneity is the occurrence of a greater number of classified properties in a given image (Figure 4, the right column). For example, Gs_WY has occurrences of all five classes within the image, whereas VO_NM and Gn_WY are dominated by only two classes. Another metric for judging heterogeneity is the lack of lateral continuity of classified zones. By this metric, GR_WY, VO_NM, and Gn_WY exhibit the strongest lateral

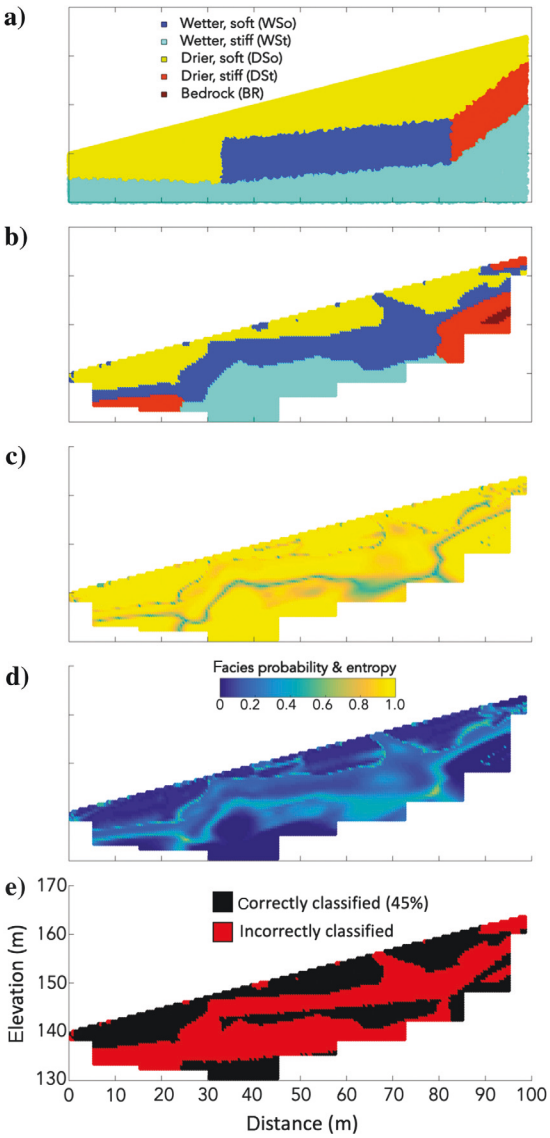

Figure 1. Results of the classification for the synthetic hillslope: (a) input hillslope, (b) classified result, (c) facies probability, (d) entropy, and (e) incorrect versus correct classification.

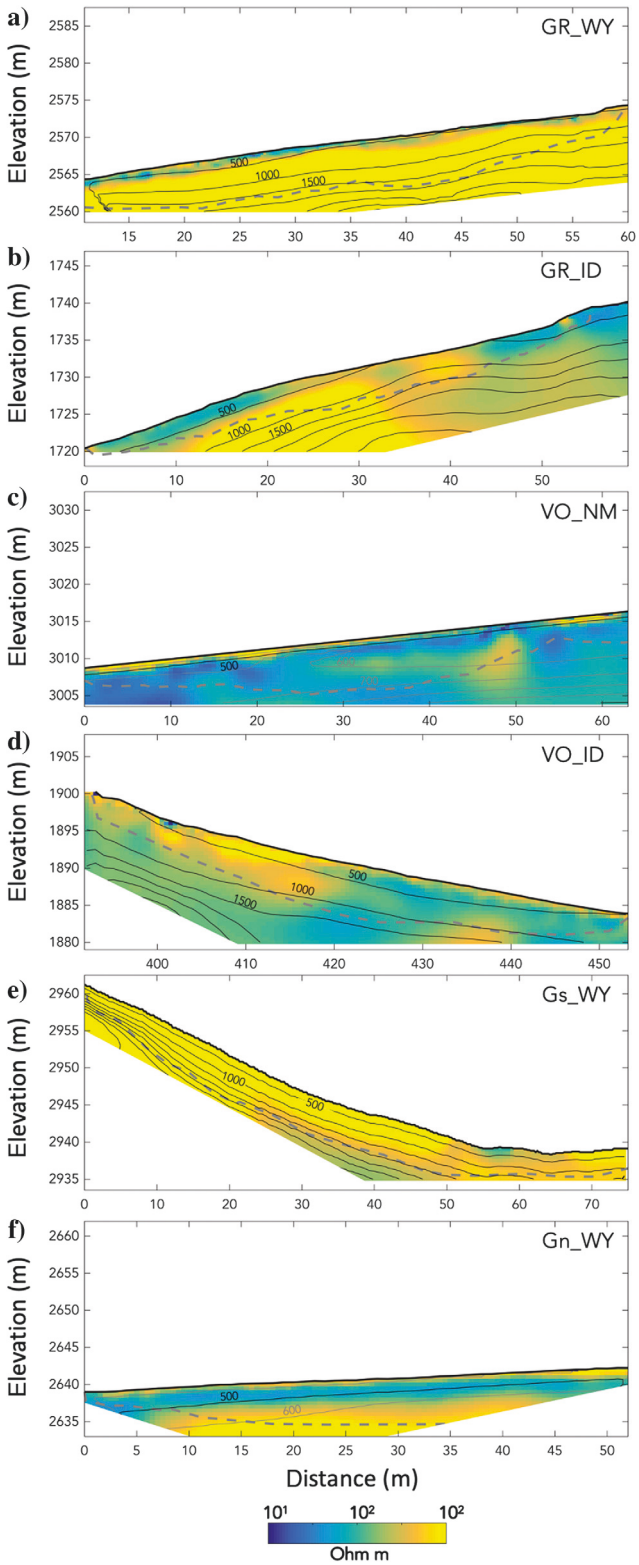

Figure 2. ERT and SSR results. The ERT results are in color, and the seismic results are the solid black contour line overlay. The V_P contours are plotted at 0.5 km s^{-1} intervals except for Gn_WY and VO_NM, which had small vertical velocity gradients; therefore, the 0.1 km s^{-1} contours are also shown in gray. The dashed lines are the maximum depth of the ERT investigation.

continuity and may therefore be judged as more homogeneous. In contrast, VO_ID and Gs_WY have distinct breaks in lateral continuity and are therefore judged as more heterogeneous.

Uncertainties in hydrofacies classification

Considering the hydrofacies classification associated with the most likely class at each pixel (Figure 5a–5f) as the probability of the maximum-probability-class closest to one, this indicates that any other classification at that point is unlikely. We note that the entropy statistic is a measurement of the uncertainty in the classification itself and does not account for regularization and spatial resolution of the independently inverted geophysical images themselves. Entropy provides an assessment of the ability of ERT and SSR to discriminate facies under the particular conditions of the selected inversion approach. The zones where probability is substantially <1 (e.g., 0.5–0.6) are at boundaries between spatial zones. These uncertain bands reflect that the algorithm is confident that there are two (or more) different materials adjacent to each other, but the exact spatial position of the boundary is uncertain. Images with high vertical resistivity and/or velocity gradients — and correspondingly more hydrofacies — thus tend to have a higher area of uncertain classification. This characteristic of higher uncertainty is apparent in VO_ID due to lateral heterogeneity toward the bottom of the slope, even though there are only three dominant classes in the image. The VO_NM and Gn_WY sites broadly have the highest hydrofacies probability across the images because they are both laterally continuous and only have two dominant hydrofacies. The entropy at each point follows a pattern similar to the probability, showing higher uncertainty near the boundaries between the facies and lower uncertainty in the middle of the facies (Figure 5g–5l).

DISCUSSION

Comparison of classification against synthetic modeling and boring data

For the synthetic experiment, the recovered hydrofacies classes reproduced the overall known input structure, although the percentage of correctly classified pixels may seem low at 45% (Figure 1). However, the percent of pixels correctly classified is only part of evaluating the result — it is also valuable to make a qualitative assessment of the recovered hydrofacies geometry. In this case, by comparing Figure 1a and 1b, we can see that — with minor exceptions — the overall geometric pattern was reconstructed with DSo on top, WSo in the middle, DSt toward the right edge, and WSt at the bottom. Notable classification errors include the DSt class in the bottom left corner (likely due to the loss of resolution at the bottom of the image, Hermans and Irving, 2017), the excursion of WSo that reaches the surface at approximately 66 m along the line (likely due to regularization in the seismic image), and a small zone classified as bedrock within the DSt zone on the right edge of the image (likely due to regularization of the ERT image). Considering together the percent correctly classified rate and qualitative evaluation of spatial distribution of facies, we conclude that the classified hydrofacies are a reasonable and useful representation of the subsurface structure.

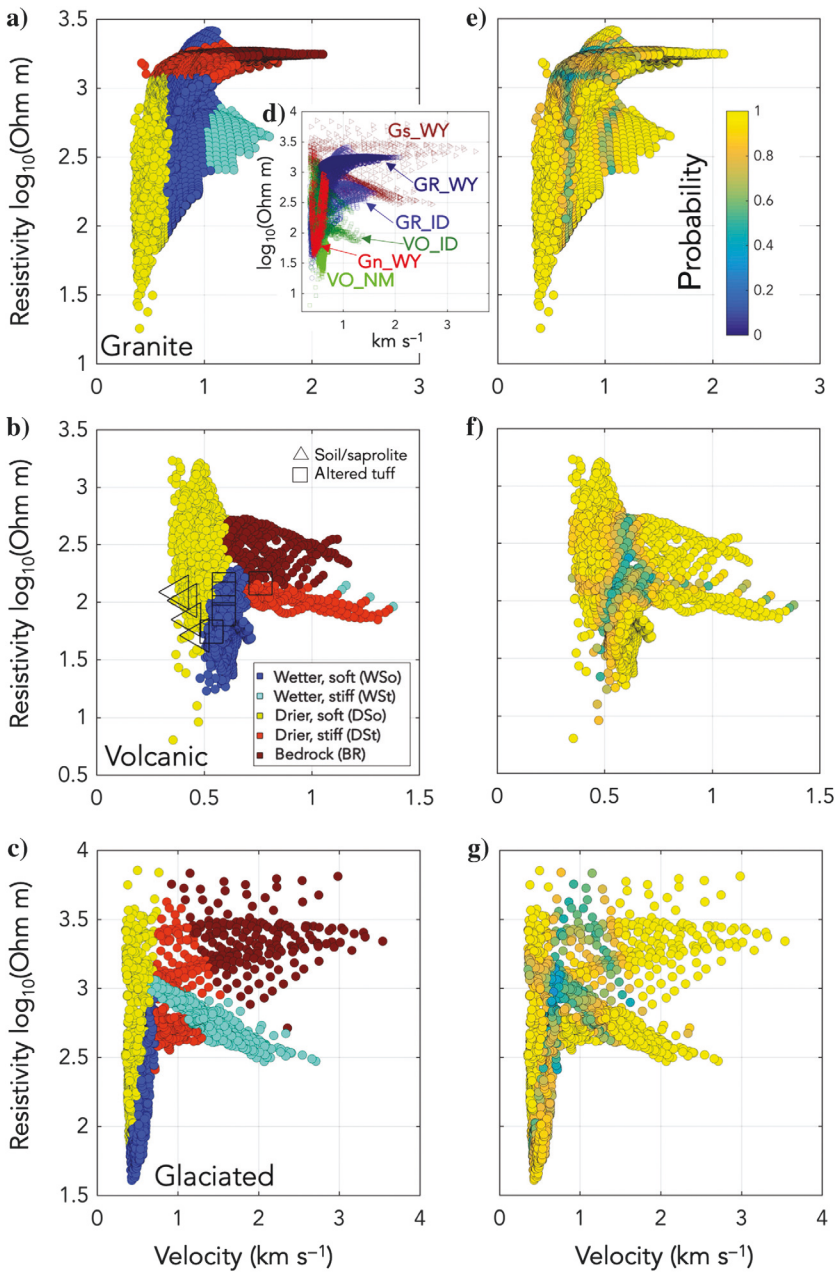

Figure 3. (a–c) Velocity-resistivity plots for each geologic setting. The color of the crossplot indicates the result of the hydrofacies classification. All sites are on the same axes as shown in inset (d). (e–g) The probability of each class displayed in (a–c), respectively. Material interpretations from extracted cores at VO_NM are shown with the associated borehole logged velocity and resistivity (Moravec et al., 2020).

Although limited direct shallow subsurface observations are available at several of our sites (GR_WY rock drill <2 m, Kotikian et al., 2019; Gs_WY rock drill <3 m, Thayer et al., 2018; and hand auger <2 m, Moraes 2019), only VO_NM has detailed deep borehole information — including geophysical logs — immediately adjacent to our transect. With depth control in the geophysical logs and the corresponding description of the geologic materials (borehole 1, Moravec et al., 2020), this is the best opportunity to provide partial validation of the EM classification. The borehole extends to 40 m below the ground surface and crosses our entire geophysical image that has a depth of investigation of 6 m. Within this depth range, we have classified two hydrofacies, DSo and WSo (Figure 4c). If we use the geophysical properties, the geologic interpretations can be overlaid on the rock-physics plot (Figure 3b) to show that the "soil/saprolite" material corresponds with DSo and the "altered tuff" is related to WSo. Importantly, the delineation between the geophysical classifications (i.e., the change from yellow to blue in Figure 3c) coincides with the delineation between geologic materials.

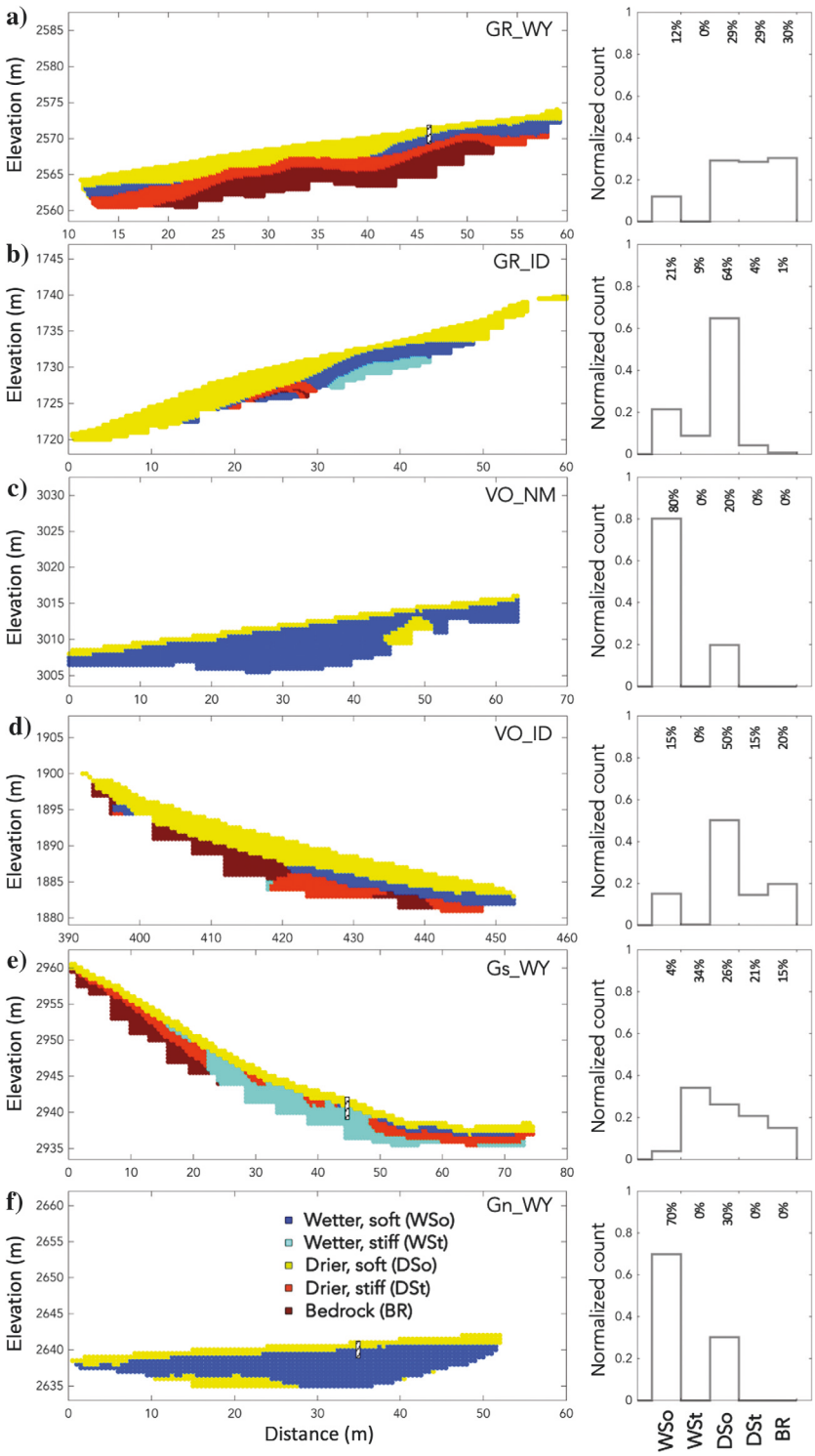

Figure 4. (a–f) Interpreted 2D hydrofacies images (the left column) based on Figure 3 and Table 3 using the EM algorithm. The borehole locations are shown by the striped markers on (a), (e), and (f). The corresponding normalized histograms (the right column) show the relative distribution of hydrofacies in each image.

We consider three shallow boreholes at Gs_WY, GR_WY, and Gn_WY for additional validation (Figure 6). At Gs_WY, only the drill cutting descriptions and the water table depth are available (Figure 6a). The drill cuttings show overburden consistent with glacial deposits from the surface to 2.5 m depth underlain by fractured rock. The measured level of the water table is approximately 0.5 m below the change from a dry classification to a wet classification, and the observed transition to fractured rock is approximately 0.65 m below the change from a soft classification to a stiff classification. At GR_WY (Figure 6b), drilling refusal was encountered at approximately 2 m depth, which corresponds to the top of the DSt classification. Unfortunately, water level logging only began in this well during December 2017 after the geophysical data sets were acquired; however, the position of the water table on the same day in the following year (i.e., 17 June 2018) at a 1.4 m depth corresponds to the transition between the DSo and WSo classes (Figure 6b). The NMR log only covers the vadose zone but confirms values consistent with unsaturated conditions at this site (Kotikian et al., 2019). At Gn_WY, the cuttings log reveals only colluvial silt and sand material below the organic horizon (Figure 6c). Only the deepest interval of the NMR log at this site enters the saturated zone, but this transition is consistent with the change in classification from DSo to WSo. We suggest that future work could acquire additional direct data to validate the classification approach recognizing that we currently have only limited information, particularly at depth.

A categorical view of the CZ: Comparisons across geologic site pairs

One of our primary objectives is to evaluate if there are similarities or differences between sites,

particularly related to geologic substrate and/or geologic history. There are fundamental limitations to making such comparisons across a diverse set of sites because most parameters can only be partially controlled. For example, it is practically impossible to precisely match the elevation, petrology, climate history, aspect, and slope between paired sites. Instead, we took the approach of

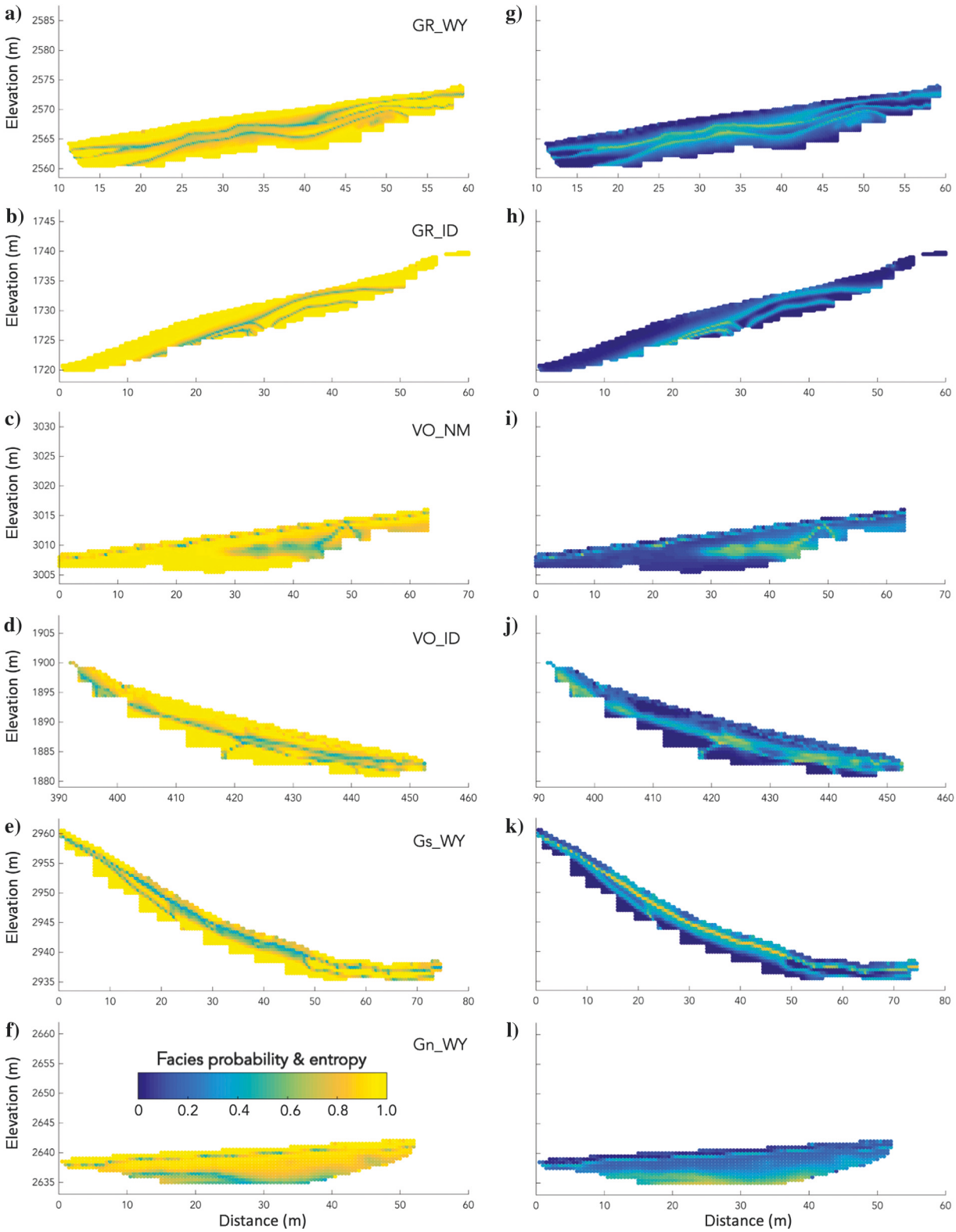

Figure 5. (a–f) Hydrofacies probability. For the classification that was identified as most likely at each pixel, this shows the probability of that classification. (g–l) The entropy calculated for each pixel, corresponding to (a–f), respectively.

selecting categorically similar sites (e.g., old granite) and similar hillslopes (i.e., moderate slopes that have soil mantles). In addition to the geologic and environmental parameter controls, the geophysical measurements depend in part on the physical properties of the subsurface; therefore, the measurement volume cannot be exactly controlled by the user. However, we are fortunate that the depth of investigation of the shallower measurement (ERT) across all six sites is relatively similar (6.5 ± 1.0 m; Figure 2). When considering all sites together, we see that there is a bias toward physical properties in the upper-left quadrant (Figure 3d). This is expected because few near-surface materials with dominant electrolytic electrical conduction would have fast velocities and low resistivities (Knight and Endres, 2005). Ore bodies that are electronic conductors may be an exception to this (Collett and Katsube, 1973; Yang and Emerson, 1997), but such materials are rare, requiring particular formational and preservational geologic processes to have occurred (Groves et al., 2005). There is also a clear clustering toward low velocities across the entire range of resistivity (Figure 3d) that is explained by the shallow depth range explored with relatively little gravimetric compaction to drive density controls on seismic velocity (Bachrach and Nur, 1998). The broad range in resistivities is controlled by expected variations in saturations due to variable infiltration and preferential flow (Miller et al., 2008; Hübner et al., 2015; Claes et al., 2019).

The two volcanic sites exhibit a wide range of resistivity, but they have generally slow seismic velocities of <1 km s^{-1} (Figure 3b). Although this similarity may not be anticipated because of the different volcanic rock types (Table 1), it is broadly consistent with extrusive parent material of a fine-grained nature and potential for syngenetic fracture networks. Additionally, these sites have little material that is of fast velocity and high resistivity, likely owing to the high initial porosity and relatively rapid weathering (Steindlberger, 2004). The incipient vesicular porosity (and possible incipient fracture networks) of the rhyolite and basalt likely also contribute to these lower seismic velocities. The presence of low seismic velocities within the top 5 m of the subsurface in VO_NM is consistent with previous results at nearby sites within the Valles Caldera where intact bedrock was interpreted up to 60 m deep (Olyphant et al., 2016).

A logical prediction would be that our two glacially altered sites would be the most similar in character because they have the most straightforward geologic history: a presumed landscape-reset at the end of the last glaciation approximately 11 ka. However, we see that these sites (Figure 3d) are quite different. We suggest that the difference in properties between the parent materials (Gs_WY = gneiss, Gn_WY = granite) may drive this in part because we would expect the less durable metamorphic rocks (Borrelli et al., 2014) to allow intensified glacial alteration. Evidence of this is available in the borehole data (Figure 6a and 6c) where the near-surface substrate at Gs_WY is primarily gravel and cobbles presumed to be colluvial (Thayer et al., 2018) in contrast to the silts and sands observed at Gn_WY. Additionally, it is possible that precise site selection relative to past glacial dynamics is important in these systems, therefore confounding our site selection. We also note that the clear "V" shape of the Gn_WY rock-physics plot (Figures 3d and 8c) bears a resemblance to the results of Meju et al. (2003) that is attributed to a connection of velocity and resistivity to porosity in the near surface (Rudman et al., 1976). In contrast, Gallardo and Meju (2003) attribute this V-shape pattern in the elastic-electric space to water saturation or to a divide between consolidated and unconsolidated materials, a condition that seems unlikely at Gn_WY given the overall slow velocity structure. Although these sites exhibit different patterns in the elastic-electric space, they nonetheless are compatible in the classification (Figure 3c) suggesting that the thicker layer of unconsolidated materials at Gn_WY within the sensitive depth results in an absence of data points in the upper-right quadrant of the elastic-electric space at Gn_WY.

In most cases, EM classification produced images of CZ architecture that are consistent with conceptual models of the geometries expected for these hydrofacies. For example, at the 40 m position on the GR_ID image (Figure 4b), there is DSo on top, underlain by WSo, and finally WSt at the bottom: this is logical because there is dryer material on top of wetter material and the softer material at the surface transitions to stiffer material at depth (i.e., more weathered over less weathered) as would be expected in a weathering granite profile (Flinchum et al., 2018b). In other cases, there are proximal facies that are perhaps not immediately intuitive. For example, at the location of the shallow boring at GR_WY (Figure 4a), there is DSo on top, underlain by WSo, then DSt, and finally bedrock at the bottom. It

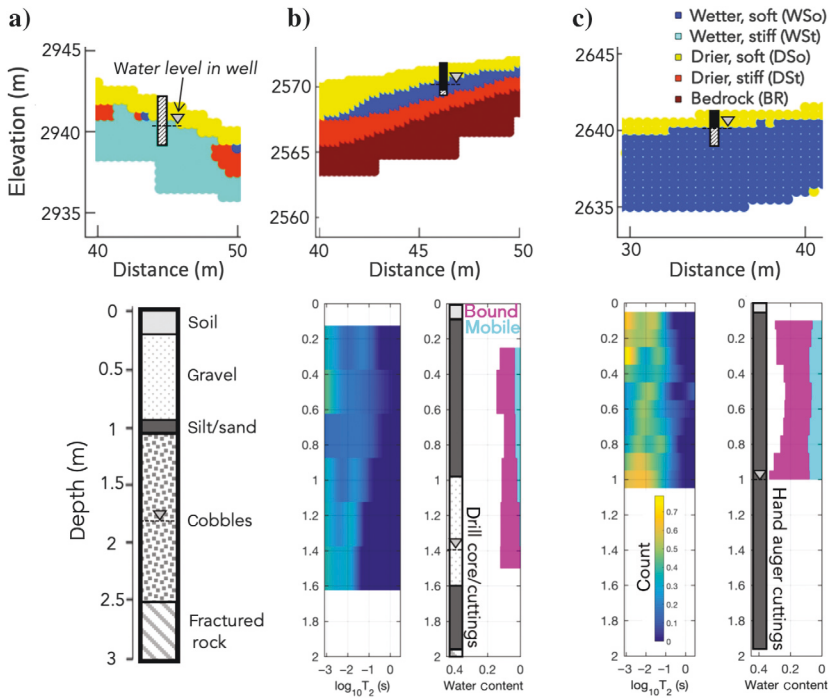

Figure 6. Borehole data from (a) Gs_WY, (b) GR_WY, and (c) Gn_WY. The top row shows an inset of the classified results for each hillslope with the borehole location. The solid portion of the boreholes indicates coverage by the borehole NMR logs, whereas the striped portion indicates the total depth. The bottom row shows the corresponding cuttings descriptions, T_2 relaxation time distributions (left), and water content logs (right). The bound and mobile water is determined using the 33 ms cutoff time from the T_2 relaxation time distributions.

may be expected that the material immediately below WSo should also be classified as wet. However, this classified result is consistent with prior observations at the site that suggest that perched water is present, that water flows laterally over the low-porosity bedrock interface, and that deep percolation of water from the unsaturated zone may not occur until later in the season than when the geophysical data was acquired (Kotikian et al., 2019). A similar arrangement of facies is present at VO_ID (Figure 4d), and we anticipate that the same explanation is valid there.

As we have discussed, one method to judge spatial heterogeneity is based on the frequency distributions of the classified hydrofacies (Figure 4, right column). This evaluation of spatial heterogeneity does not explicitly account for how classes are either grouped or scattered throughout the section (Figure 4, left column). However, the nature of the geophysical images and classification method results in the spatial continuity of the hydrofacies; therefore, the distribution of classes is not an independent feature of the results.

As previously stated, our comparison is primarily based on hillslopes with categorically similar substrates; however, here we provide a basic comparison of the slope aspect and modern precipitation contrasts within our site group. The general expectation is that south-facing hillslopes will be drier and have sparser vegetation than comparable north-facing slopes in the northern hemisphere (Pelletier et al., 2018); consequently, north-facing hillslopes will experience more weathering. In the two granitic hillslopes, GR_WY and GR_ID that are south–southeast facing and east–northeast facing, respectively, the DSo classification is common (Figure 4a and 4b). The relatively shallow depth to bedrock at GR_WY may indicate reduced weathering due to the south–southeast orientation. In the two volcanic hillslopes, VO_NM and VO_ID that are west–southwest facing and northeast facing, respectively, the distribution of classes is different with VO_NM being relatively homogeneous WSo and VO_ID having a heterogeneous distribution of classes (Figure 4c and 4d). The two formerly glaciated sites both have similar southwest aspect; however, Gs_WY receives more precipitation annually (Table 1). Given the recent quaternary age of the glacial materials that make up the substrate, we do not attribute the differences in classification to porosity creation through chemical weathering; rather, their position on the hillslope may be more important in this case. Because the time since the landscape reset at all sites is not precisely known (Table 1) and the climate history in this period is different for all sites, we cannot make quantitative comparisons related to geologic age. However, we do observe that — even with a very large range in site pair ages — there is nonetheless the possibility to observe all classes at any geologic grouping, suggesting that incidental factors such as the hillslope location within the catchment and the depth of investigation play an important role in how the CZ subsurface is interpreted.

Variations and limitations of the EM classification

The hydrofacies classification can be performed using different multivariate techniques, such as discriminant analysis, *k*-means clustering, or machine learning approaches (Hastie et al., 2002). The EM method is a probabilistic iterative approach. Hence, the solution is a probability distribution of the hydrofacies conditioned by the geophysical data. A potential limitation is that the method does not include a spatial correlation model (Delforge et al., 2021) and each data point is classified independently from the adjacent data. Markov chain models could be integrated in the hydrofacies classification as in Eidsvik et al. (2004) and Lindberg and Grana (2015), but application in 2D and 3D is challenging due to the large number of constraints introduced by the spatial model.

Although our classification scheme cannot be entirely unsupervised, we propose that user determination of the prior means may produce a less subjective result than an entirely visual cointerpretation of geophysical images (Thayer et al., 2018) due to the statistical grouping of pixels based on quantitative parameters in the EM. Given the need for "expert" guided selection of the prior means, here we assess how changing the priors affects the classification result. The prior means (vector μ with length equal to the number of classes), or "initial guesses," are the parameters in which the prior geologic knowledge from on-site expert assessment or literature can be used; that is, these are a way to impose a physical meaning on the classification. To achieve the classifications presented in Figures 4 and 5, we used a set of starting values estimated for each site based on general knowledge of local conditions. Nonetheless, there is intersection of the hydrofacies in the crossplots (Figure 3) due to the inherent noise and spatial resolution limitations (i.e., regularization of the inversions) of the geophysical images. Determining unambiguous universal starting values or achieving identical starting values by different expert judges is currently not possible because easily quantifiable site parameters are similarly subject to interpretation and variability. For example, although the bedrock velocity may be consistent and reported in the literature for all sites, the definition of velocity and resistivity for "wetter, softer" materials from sparse extracted samples at a site may vary unpredictably depending on where the samples were taken. One of the advantages of this classification approach is that calibrating rock-physics models is not required, and this advantage would be somewhat degraded if large sets of samples were required to assign starting points for each class. Therefore, we ask: how much does the overall classification result depend on the set of prior means (starting values) particularly if physically unlikely starting values are used? The result of varying μ by $\pm 10\%$ and $\pm 20\%$ (relative to Table 3) in Figure 7 details an example for the Gs_WY hillslope. Using $\pm 10\%$ starting values has little effect on the resulting hydrofacies classification. Using starting points $\pm 20\%$ yields a similar overall structure but changes the frequency and location of the hydrofacies. Although there is some leeway in selecting starting values for each class and obtaining a robust result, this analysis highlights the importance of using physically plausible values for the prior means.

As discussed above, it is also possible to decrease or increase the number of prior means used in the EM classification. In this work, we chose five as the simplest logical framework obtained by breaking the rock-physics crossplot into high versus low quadrants for parameters that reflect conceptual stiffness and dryness conditions and by adding a bedrock classification. However, additional classes could be justified for particular site conditions, as long as the geophysical properties could be readily defined. For example, scree and wetland soils that are commonly found in the vicinity of each of these sites (though not within the lines presented here) would likely have distinct enough properties to justify inclusion as independent classes. However, increasing the number of prior means used in the algorithm may necessitate a more accurate selection of their starting values as the classifications are likely to be reduced to smaller regions within the rock-physics crossplot. Finally, it would also be possible to set up this problem as an unsupervised classification where classes are matched with geologic explanation after the classification process — such an approach may be useful if no a priori geologic information is available.

Limitations associated with data acquisition

Here, we address how the timing of data acquisition may impact the classification result. A fundamental limitation to geophysical measurement — particularly active-source seismic refraction — is that acquiring data at different sites on the same day requires multiple teams of people and multiple instruments. Frequently, such duplication is not available. In our case, owing to the substantial geographic and climatic separation of sites, careful control of the timing of measurements is not always possible. Furthermore, we sought to use existing published data sets to enhance opportunities for comparison. Therefore, we aimed to acquire generally "mid-summer" data sets, although we realize that other hydrotemporal markers — such as the start of the water year or a set number of days after snow-off — also have scientific value. Importantly, we adhered to the requirement of ERT and seismic data acquisition on the same day of the year to bolster the assumptions of the underlying rock-physics models (Meju et al., 2003). Nonetheless, we ask: what is the effect on the classification if cotemporal data sets are not available? Saturation is a second-order effect on seismic velocity at most saturations (Bachrach and Nur, 1998); therefore, it is plausible that the two data sets do not need to be measured under identical hydrologic conditions. We provide examples for VO_ID and Gn_WY (Figure 8) showing the difference between the cotemporal data (the same data from Figures 4 and 5) and using ERT data from October 1 of the same year as the seismic data (the seismic acquisition dates in Table 2). For these examples, we used only a "local" classification restricted to single sites (rather than the geologically grouped classification as elsewhere in the manuscript). First, we can clearly see that there is only a small shift in the registration of the two rock-physics plots relative to one another (Figure 8a and 8c) — the overall shape and character of the point clouds are clearly preserved. The frequency of classes at VO_ID (Figure 8b) is similar in both cases, except the "drier, stiffer" (DSt) class is essentially absent in the result from different dates, whereas it is the second most frequent class in the cotemporal data. This may be explained by the inconsistencies of using ERT and SSR from different dates and wetness conditions, resulting in some misclassification, although this could also be viewed as a positive point because we would expect the classification to change depending on the hydrologic input variations throughout the year. At Gn_WY (Figure 8d), we find that there is little difference in the classifications, with only a slight increase in the DSo category and a slight decrease in the WSo and WSt categories. Given the snowmelt-dominated hydrology at Gn_WY, we could generally expect the subsurface to continue drying

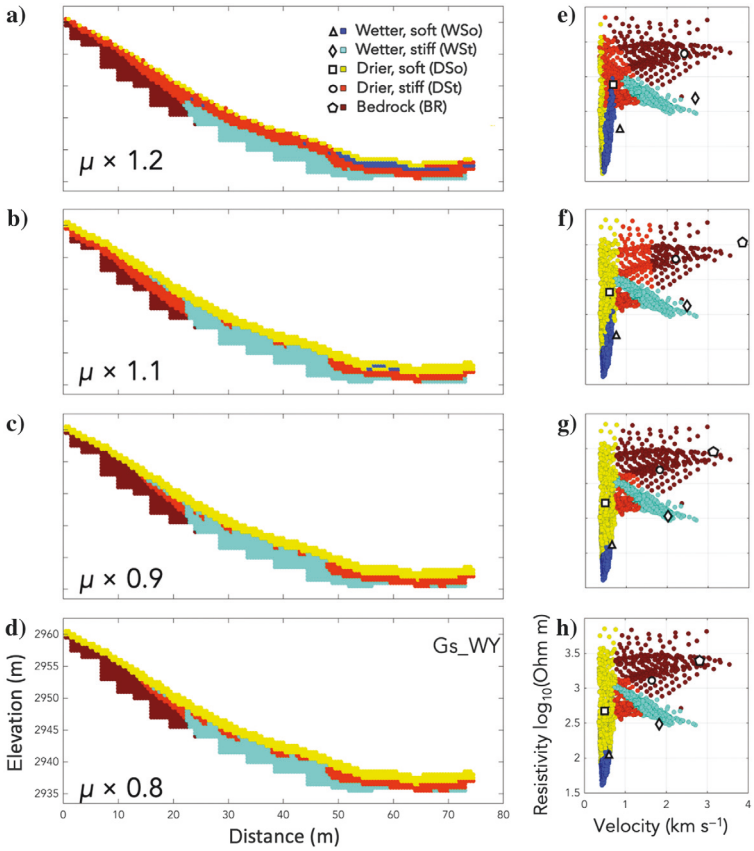

Figure 7. Effect of varying the input means. Using Gs_WY as an example, (a–d) show the effect of using a set of input means (i.e., the five V_P and ρ pairs for a site from Table 3) that are ±10% and ±20% from the values in Table 3 and Figure 4; (e–h) show the corresponding plots in the elastic-electric space with the initial values plotted as the white symbols.

Table 3. Input means for EM classification. The V_P = P-wave seismic velocity. The term ρ is resistivity.

	Glaciated		Granitic		Volcanic	
	V_P, (km s^{-1})	ρ, $\log_{10}$ (Ohm m)	V_P, (km s^{-1})	ρ, $\log_{10}$ (Ohm m)	V_P, (km s^{-1})	ρ, $\log_{10}$ (Ohm m)
Wetter, soft	0.70	2.18	0.90	1.75	0.50	1.00
Drier, soft	0.55	2.80	0.50	2.50	0.50	3.10
Wetter, stiff	2.25	2.60	2.00	2.50	2.00	3.50
Drier, stiff	2.00	3.25	1.75	3.00	2.00	2.50
Bedrock	3.50	3.50	3.50	3.50	3.00	3.50

from midsummer through October when precipitation is low (Moraes, 2019). This could explain the slight differences in frequency distributions between the time-coincident and different season measurements. In general, it seems that using noncoincident geophysical measurements do not have a substantial impact on the resulting hydrofacies classifications.

A limitation of our hydrofacies classification approach is that known material units are not necessarily explicitly resolved either due to the number of allowed classes or the resolution of the measurement. For example, in the Idaho sites VO_ID and GR_ID, the upper portion of the hillslopes has a layer of loess of variable thickness (Seyfried et al., 2018), at least 0.5 m in places based on our test pit observations. At GR_ID (Figure 2b), this is material at the surface with $V_P < 0.5$ km s^{-1} and $\rho < 10^2$ Ohm m. At VO_ID (Figure 2d), this is material with $V_P < 0.5$ km s^{-1} and $\rho > 10^{2.5}$ Ohm m, implying that the loess is drier at VO_ID than GR_ID. However, it is difficult to capture both of these wetness conditions for the same material type — particularly for a unit that is relatively thin and approaches the spatial resolution of SSR and ERT — and the loess is not explicitly resolved in our classification (Figure 4b and 4d).

Although we generally assume that V_P changes little over most of the range of saturation (Bachrach and Nur, 1998) and is therefore primarily sensitive to porosity and ρ primarily responds to wetness, we recognize that ρ is also controlled by the porosity and clay content (Archie, 1942; Waxman and Smits, 1968). Through the joint classification in the elastic-electric domain, the information from seismic velocity provides information on porosity to the classification although the clay content remains unknown. Although we cannot explicitly constrain for clay given the available data, this could possibly be addressed through inclusion of clay estimation into the classification by induced polarization measurement during ERT acquisition (Slater and Lesmes, 2002).

A final consideration and potential limitation is the uncertainty and regularization in the inverted geophysical images. As with all smoothness-constrained geophysical inversions, sharp boundaries are not well resolved, and this can cause overlapping between classes (Hermans and Irving, 2017). In our ERT inversion results, the regularization is optimized based on the measured data uncertainties (Binley, 2015), but due to the inherent smoothness of the images and loss of resolution with depth, some misclassification is unavoidable.

Comparing statistical classification against other subsurface CZ delineation methods

Delineating zones of subsurface materials has been a key aim of past hydrogeophysical efforts to examine CZ architecture. For example, Leopold et al. (2013) use ERT imaging in conjunction with geologic interpretation to map material categories of stratified slope deposits, saprolite, weathered bedrock, and bedrock, for alpine and montane basins in the Boulder Creek CZO. These CZ substrate descriptions are relatively similar to those used by Befus et al. (2011) and Flinchum et al. (2018b) in conjunction with seismic refraction imaging to define disaggregated material, saprolite and compacted sediment, weathered bedrock, and fresh bedrock, in glacially modified, low-relief mountain upland, and deep bedrock gorges. Although these are valuable interpretations, the CZ substrate descriptions are distinct from the hydrofacies we use, primarily because we are able to produce quantitative analysis of categorical variables between measured geophysical parameters and classifications, whereas the CZ substrate descriptions are rooted in subjective expert interpretations. Flinchum et al. (2018b) produce subsurface water capacity quantifications similar to our hydrofacies interpretations, again based primarily on seismic imaging with extensive calibration. These CZ substrate descriptions (Befus et al., 2011; Leopold et al., 2013) and water holding capacity maps (Flinchum et al., 2018b) are developed from a single physical parameter; therefore, nonuniqueness may influence the interpretations, a factor that can be mitigated by using contrasting physical properties as in our EM classification. One advantage of a single-physics approach is that data can be acquired more quickly over larger areas, as evidenced by the longer aggregate line length used in the example studies (6–9 km) in comparison to ours (approximately 0.4 km each of ERT and seismic).

It is not uncommon to acquire colocated seismic and resistivity data, and qualitative cointerpretation has been successfully used to synthesize these data sets. Seyfried et al. (2018) report colocated seismic and airborne electromagnetic resistivity images along our VO_ID line and interpreted weathering and hydrologic parameters, although no structural map was produced. Holbrook et al. (2014) produce a cointerpreted 2D section using CZ substrate classifications (i.e., dry and wet regolith and bedrock) for a site in the Southern Sierra CZO where zone interpretations were made based on the particular strengths of each measurement (e.g., vertical stratification from seismic, wetness from resistivity). Thayer et al. (2018) measure seismic and ERT data along our Gs_WY line to produce a qualitative cointerpretation map of CZ substrate classes following the categories used by Holbrook et al. (2014) with a zone geometry that bears resemblance to the classification determined from EM (Figure 4e). Thayer et al. (2018) predict interfaces slightly deeper

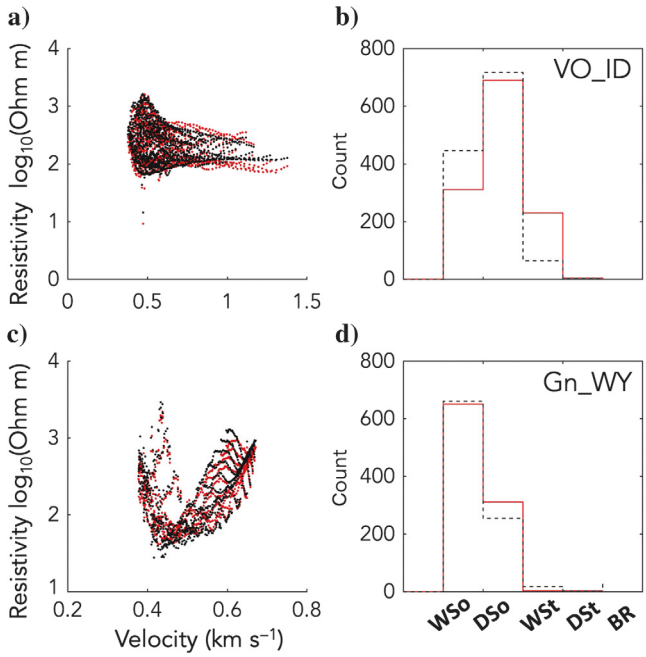

Figure 8. Effect of using ERT data from a different date than the seismic measurement. Using VO_ID and Gn_WY as examples, this shows how (a and c) the rock-physics plots change when using an ERT data set from a different season than the coincident seismic/ERT midsummer measurement. RED, time-coincident midsummer data; BLACK, midsummer seismic with October ERT. (b and d) The resulting changes to the hydrofacies classification shown, respectively, as frequency distributions.

including fresh bedrock 2 m deeper than our EM approach (3 m, Figure 4e). Joint geophysical inversion results have also been classified. However, these efforts were apparently based on grouping semicontinuous trends in the resistivity-velocity space under conditions in which the relationship between these parameters is simpler than we observe (Gallardo and Meju, 2003).

CONCLUSION

We have demonstrated the classification of CZ subsurface materials into hydrofacies using joint classification in the elastic-electric domain of electrical and seismic geophysical images. We classified the hydrogeophysical properties of hillslopes underlain by glacially deposited, weathered granite and volcanic geologic materials into relative stiffness and wetness groups and found four of the five classes, including bedrock, in at least one of each pairing. This approach yields spatial distributions of the relative material properties informed simultaneously by the strength of each individual geophysical method with minimum bias from qualitative human interpretation. Although user input is needed to define a basic starting point for the classification, we find that the results are robust as long as reasonable starting values are chosen based on field observations and background site information. Similarly, we find that seismic and electrical data acquired coincidently are ideal for the classification approach based on assumptions of the underlying physics; however, data sets acquired at different times of the year still resolve similar classifications. Although the geophysical imaging length and depth dimensions were similar at all of our comparison sites, we nonetheless found substantial differences in geometries between sites with a similar geologic history as well as across sites with unrelated substrates. One volcanic site and one glaciated site were characterized only by soft classes. The classified images revealed the spatial variability in the near subsurface of each site primarily in the form of surface parallel layering with wetter classes mostly concentrated toward the deeper portions of the profiles. Our results suggest that the geologic history is linked to modern hydrogeophysical properties of the substrate based on inferred weathering or sedimentation profiles, but this would not be directly identifiable exclusively using the classified geophysical images.

ACKNOWLEDGMENTS

This work was supported by NSF grant #1818550. We thank A. Miller and A. Killebrew for their assistance with geophysical data collection and field logistics. Data associated with this manuscript may be accessed through HydroShare, Parsekian et al. (2020).

DATA AND MATERIALS AVAILABILITY

Data associated with this research are available and can be accessed via the following URL: http://www.hydroshare.org/resource/775a3bd3d8674bc38ceb160b902e5056.

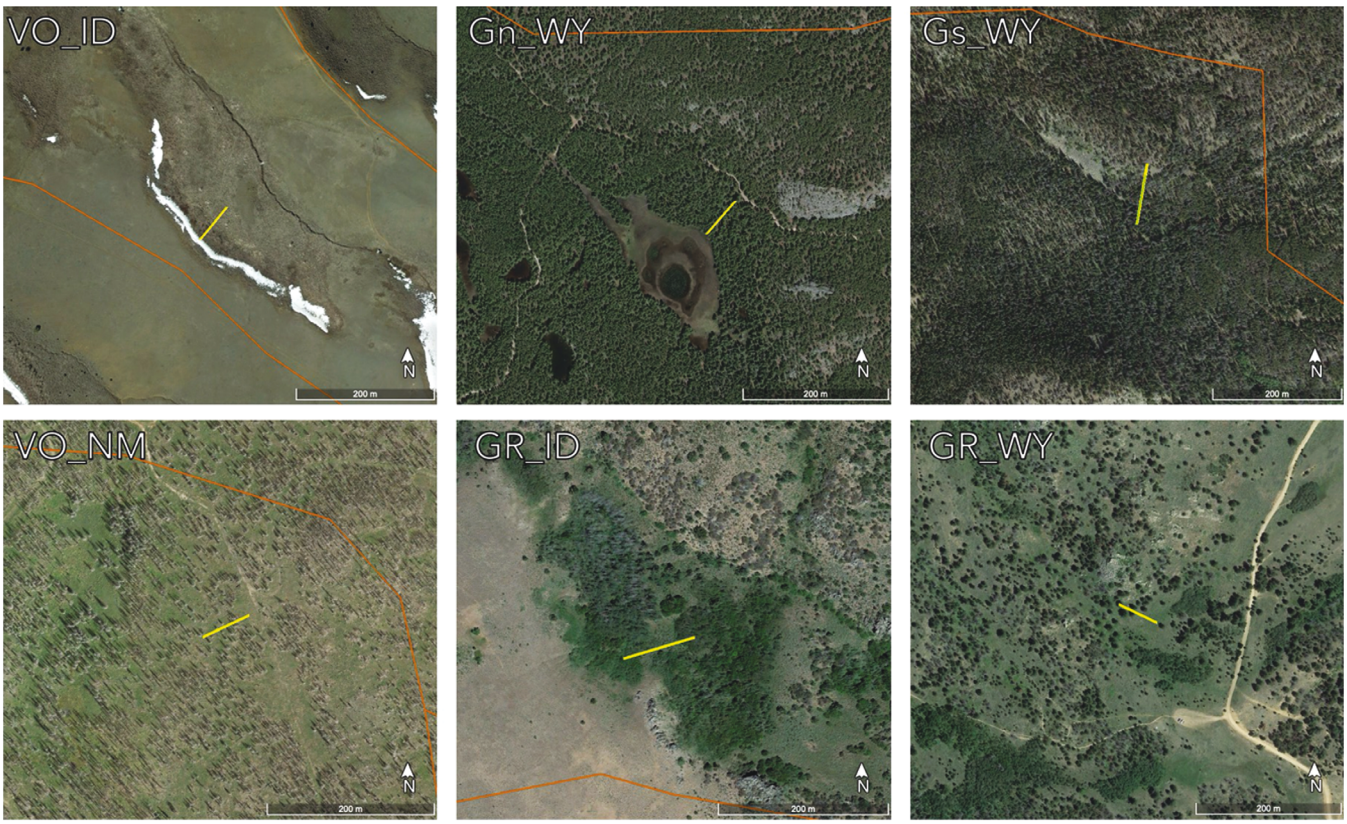

Figure A-1. Locations of the geophysical transects (yellow) on hillslopes. Catchment outlines within the view extent (United States Environmental Protection Agency, 2020) are shown in orange.

APPENDIX A

STUDY SITE LOCATIONS

The geophysical transects are located on hillslopes within snowmelt-dominated headwaters catchments in Wyoming, Idaho, and New Mexico (Figure A-1). Sites in Idaho and New Mexico are associated with the Reynolds Creek Critical Zone Observatory (RCCZO) and the Jemez River Basin Critical Zone Observatory (JRBCZO), respectively.

APPENDIX B

SENSITIVITY ANALYSIS FOR GEOPHYSICAL DATA SETS

All geophysical images have parameter uncertainty that varies spatially. This must be accounted for when analyzing geophysical results, particularly when selecting which regions of an image can be assigned high confidence. Seismic refractions are often masked below the depth of the deepest calculated raypath (Befus et al., 2011; Figure B-1). However, an image-wide sensitivity also can be estimated by varying the gradient starting models used in the inversion (Holbrook et al., 2014). We completed this sensitivity analysis for each of our transects using a range of gradients from 0.3–0.5 km s^{-1} to 0.3–5 km s^{-1}, producing a sensitivity map as shown in the example in Figure B-2a. We summarize the sensitivity as a function of velocity for all seismic data sets as shown by the trendlines in Figure B-2b, summarized in Table B-1. Histograms of velocity-normalized standard deviation are shown in Figure B-2c. We only used portions of the seismic image within the sensitive zone for EM analysis. All sites had a similar sensitivity with a peak approximately 0.08 in the histograms (Figure B-2c), except VO_ID, which had higher uncertainty peaking at approximately 0.25.

The standard approach for calculating the depth of investigation for ERT images is to use the sensitivity map calculation approach of Oldenburg and Li (1999). For each of our images, we define the base of the sensitive zone at the suggested 0.2 limit, which is likely conservative. The example sensitivity map shown in Figure B-3a shows that useful measurements are limited to approximately 5 m below the surface in the deepest areas, as expected given the somewhat convex-upward geometry of the hillslope that reduces the depth of investigation. Although the measured resistance and uncertainty are linked (Slater et al., 2000), sensitivity and resistivity are not necessarily related given that the instrument setup geometry and distribution of subsurface materials contribute to the calculated sensitivity. Nonetheless, we find plotting the sensitivity as a function of resistivity (Figure B-3b, summarized in Table B-1) to be an efficient way of summarizing the results for all data sets, in conjunction with normalized histograms (Figure B-3c).

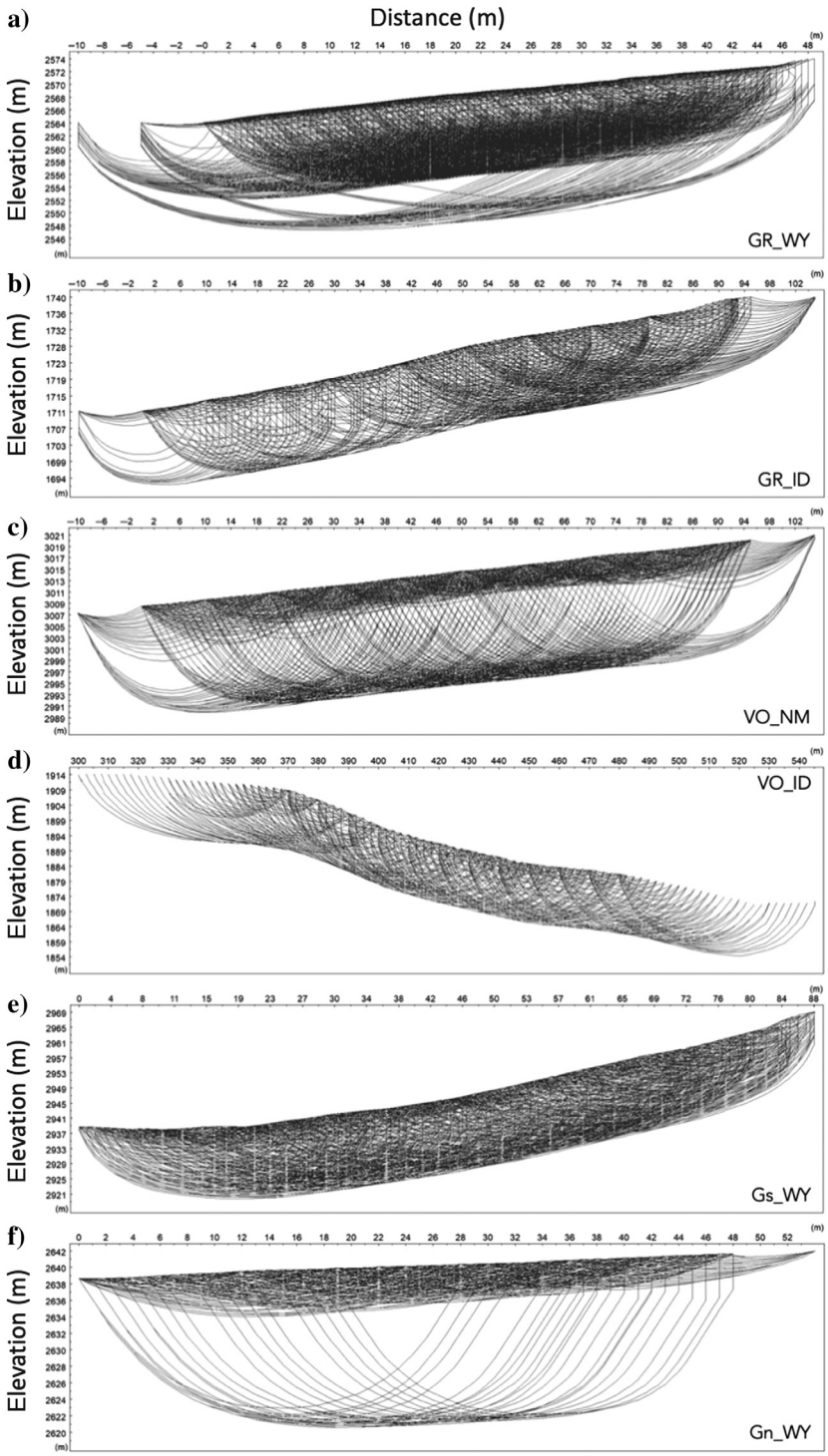

Figure B-1. Ray tracing for each seismic refraction tomogram.

APPENDIX C

SYNTHETIC HILLSLOPE

The fictitious hillslope was not intended to replicate any of our field sites, but rather to have a realistic distribution of facies and geometry. We defined the porosity and saturation material properties for each of the five hydrofacies classes with plausible values of

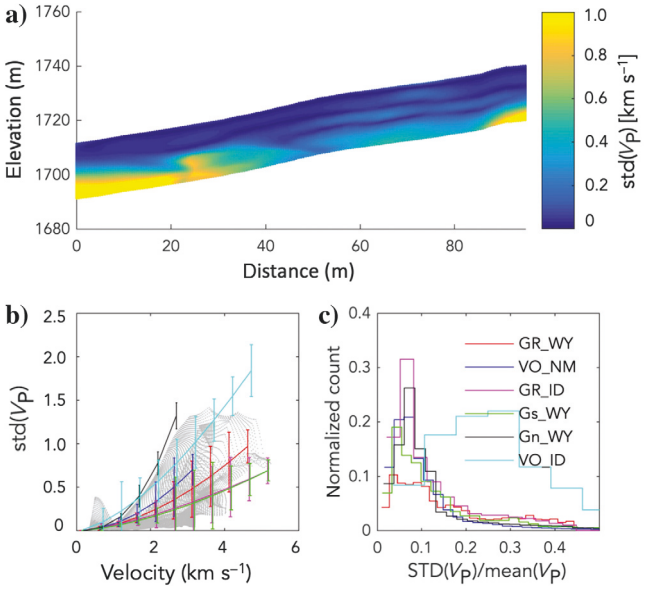

Figure B-2. Sensitivity analysis of the seismic results. Using 10 different starting models for each data set with gradients ranging from 0.33/0.5 to 0.33/5 km s^{-1}. An example of the standard deviation (std) of the resulting set of inversions is shown for GR_ID in (a). The velocity versus standard deviation relationships for all sites are shown in (b) with the points from GR_ID shown in gray as an example. Each data set is fitted with a power-law function (Table B-1), and uncertainty bars are shown based on bin averaging. (c) The histogram of the standard deviation divided by the mean velocity for all sites.

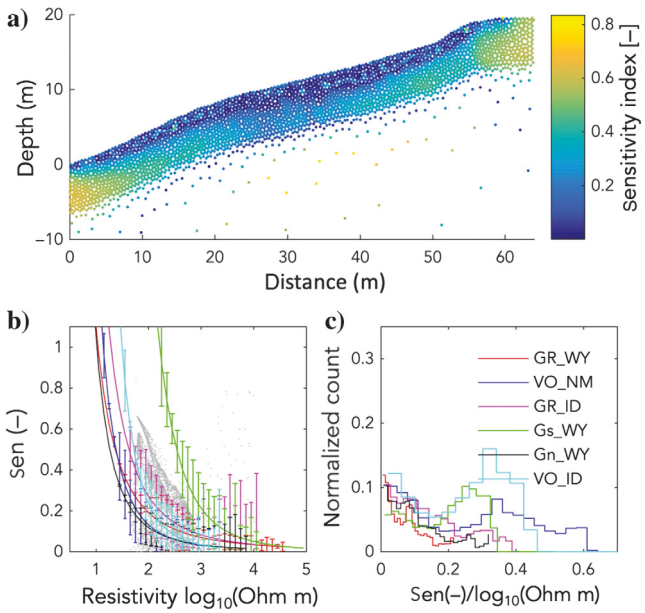

Figure B-3. Sensitivity analysis of the ERT results using the method of Oldenburg and Li (1999). An example of the calculated sensitivity map is shown for GR_ID in (a) with the colorbar indicating the calculated sensitivity value for the centroid of each mesh element. The resistivity versus sensitivity relationship for all sites is shown in (b) with the points from GR_ID shown in gray as an example. Each data set is fit with a power-law function (Table B-1) and uncertainty bars are shown based on bin averaging. (c) The histogram of the sensitivity divided by the mean resistivity for all sites.

Table B-1. Model parameters for the trendlines in Figure B-2b and B-3b using $y = \exp(a)x^b$.

	ERT		SSR	
	a	*b*	*a*	*b*
GR_WY	−2.5	0.1	1.8	−8.3
VO_NM	−3.6	0.5	2.0	−9.2
GR_ID	−2.9	0.7	1.5	−6.2
Gs_WY	−5.1	4.1	1.6	−7.1
Gn_WY	−3.1	0.0	2.6	−13.2
VO_ID	−4.9	2.0	1.5	−5.6

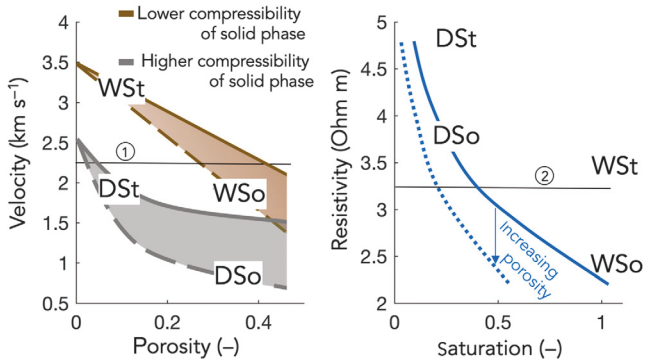

Figure C-1. The hydrofacies properties used in the synthetic hillslope plotted within the V_P/porosity and ρ/saturation space. The conceptualized effects of variations in porosity and saturation within the rock-physics framework are annotated (note: The diagrams are for illustration purposes only, and they do not correspond with the hydrofacies classes used in the synthetic hillslope). The horizontal lines labeled "1" and "2" indicate examples of possible ambiguity for the given observed geophysical properties that justify a classification approach where sites are grouped according to similar geologic substrates.

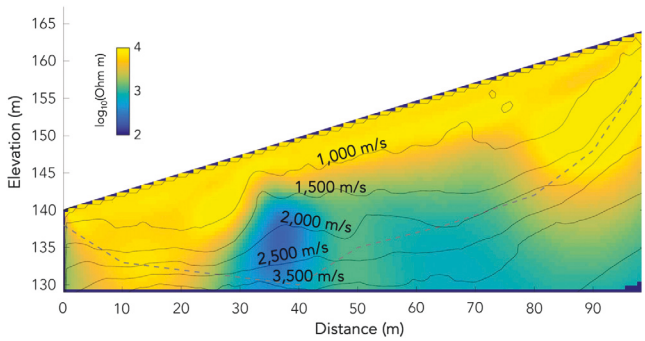

Figure C-2. Synthetic ERT and SSR results. The ERT results are in color, and the seismic results are the solid black contour line overlay. The V_P contours plotted at 0.5 km s^{-1} intervals. The dashed line is the maximum depth of the ERT investigation.

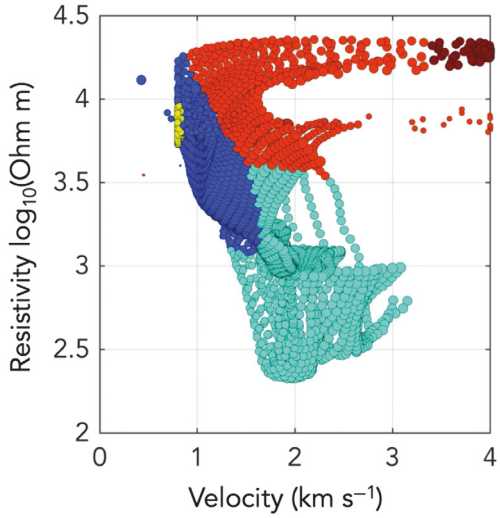

Figure C-3. Velocity-resistivity plot for the synthetic data set. The color of the crossplot indicates the result of the hydrofacies classification. The V_P = P-wave seismic velocity. The term ρ is resistivity.

Table C-1. Parameters input to the hydrofacies for the synthetic hillslope.

	Porosity	Saturation	Velocity, km s^{-1}	Resistivity, $\log_{10}$ (ohm m)
DSo	0.3	0.2	0.83	3.9
DSt	0.05	0.2	1.92	4.9
WSo	0.3	1	1.85	2.5
WSt	0.05	1	3.13	3.5

Table C-2. Input parameters for the rock-physics equations used to predict the velocities and resistivities listed in Table C-1.

Symbol	Value	Units	Definition
R_w	100	Ohm m	Pore-water resistivity
a	0.6	—	Tortuosity constant
m	1.3	—	Cementation exponent
n	2	—	Saturation exponent
K	33	GPa	Bulk moduli of mineral
G	33	GPa	Shear moduli of mineral
ρ_m	2.65	g/cm^3	Mineral density
K_w	2.15	GPa	Bulk moduli of pore water
ρ_w	1	g/cm^3	Density of pore water
ϕ_c	0.6	—	Critical porosity
C	4	—	Coordination number
P	0.00014	GPa	Effective pressure

low and high porosity and saturation (Table C-1 and Figure C-1). These material properties were converted to geophysical properties using Archie's law to predict resistivity and the Hertz–Mindlin theory with Hashin–Shtrikman elastic bounds (Mindlin, 1949; Hashin and Shtrikman, 1963) to predict seismic velocity (Table C-1). Full details related to how we applied these models can be found in Flinchum et al. (2018a). The parameters used in the rock-physics equations are shown in Table C-2. The results of the geophysical forward modeling based on these input properties are shown in Figure C-2. The collocated points on the transect are plotted in the elastic-electric space with final classifications in Figure C-3.

REFERENCES

Anderson, M. P., 1989, Hydrogeologic facies models to delineate large-scale spatial trends in glacial and glaciofluvial sediments: Geological Society of America Bulletin, **101**, 501–511, doi: 10.1130/0016-7606(1989)101<0501:HFMTDL>2.3.CO;2.

Anderson, S. P., R. C. Bales, and C. J. Duffy, 2008, Critical zone observatories: Building a network to advance interdisciplinary study of Earth surface processes: Mineralogical Magazine, **72**, 7–10, doi: 10.1180/minmag.2008.072.1.7.

Archie, G. E., 1942, The electrical resistivity log as an aid in determining some reservoir characteristics: Transactions of the AIME, **146**, 54–62.

Astic, T., L. J. Heagy, and D. W. Oldenburg, 2020, Petrophysically and geologically guided multi-physics inversion using a dynamic Gaussian mixture model: Geophysical Journal International, **224**, 40–68, doi: 10.1093/gji/ggaa378.

Bachrach, R., and A. Nur, 1998, High-resolution shallow-seismic experiments in sand — Part 1: Water table, fluid flow, and saturation: Geophysics, **63**, 1225–1233, doi: 10.1190/1.1444423.

Bedrosian, P. A., N. Maercklin, U. Weckmann, Y. Bartov, T. Ryberg, and O. Ritter, 2007, Lithology-derived structure classification from the joint interpretation of magnetotelluric and seismic models: Geophysical Journal International, **170**, 737–748, doi: 10.1111/j.1365-246X.2007.03440.x.

Befus, K. M., A. F. Sheehan, M. Leopold, S. P. Anderson, and R. S. Anderson, 2011, Seismic constraints on critical zone architecture, Boulder Creek watershed, Front Range, Colorado: Vadose Zone Journal, **10**, 915–927, doi: 10.2136/vzj2010.0108.

Binley, A., 2015, Tools and techniques: Electrical methods, *in* G. Schubert, ed., Treatise on geophysics, 2nd ed.: Elsevier, Resources in the Near-Surface Earth, Vol. **11**.

Binley, A., G. Cassiani, and R. Deiana, 2010, Hydrogeophysics: Opportunities and challenges: Bollettino di Geofisica Teorica ed Applicata, **51**, 267–284.

Binley, A., G. Cassiani, and P. Winship, 2004, Characterization of heterogeneity in unsaturated sandstone using borehole logs and cross-borehole tomography, *in* J. S. Bridge and D. W. Hyndman, eds., Aquifer characterization: Society for Sedimentary Geology.

Borrelli, L., F. Perri, S. Critelli, and G. Gullà, 2014, Corrigendum to "characterization of granitoid and gneissic weathering profiles of the Mucone River basin (Calabria, Southern Italy)": Catena, **113**, 325–340, doi: 10.1016/j.catena.2013.08.014.

Brantley, S. L., M. I. Lebedeva, V. N. Balashov, K. Singha, P. L. Sullivan, and G. Stinchcomb, 2017, Toward a conceptual model relating chemical reaction fronts to water flow paths in hills: Geomorphology, **277**, 100–117, doi: 10.1016/j.geomorph.2016.09.027.

Claes, N., G. Paige, D. Grana, and A. D. Parsekian, 2020, Parameterization of a hydrologic model with geophysical data to simulate observed subsurface return flow paths: Vadose Zone Journal, **19**, e20024, doi: 10.1002/vzj2.20024.

Claes, N., G. B. Paige, and A. D. Parsekian, 2019, Uniform and lateral preferential flows under flood irrigation at field scale: Hydrological Processes, **33**, 2131–2147, doi: 10.1002/hyp.13461.

Collett, L. S., and T. J. Katsube, 1973, Electrical parameters of rocks in developing geophysical techniques: Geophysics, **38**, 76–91, doi: 10.1190/1.1440336.

Delforge, D., A. Watlet, O. Kaufmann, M. Van Camp, and M. Vanclooster, 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, doi: 10.1016/j.jappgeo.2020.104203.

Dempster, A. P., N. M. Laird, and D. B. Rubin, 1977, Maximum likelihood from incomplete data via the EM algorithm: Journal of the Royal Statistical Society: Series B (Methodological), **39**, 1–22.

Deutsch, C. V., and A. G. Journel, 1992, Geostatistical software library and user's guide: University Press.

Doetsch, J., N. Linde, I. Coscia, S. A. Greenhalgh, and A. G. Green, 2010, Zonation for 3D aquifer characterization based on joint inversions of multimethod crosshole geophysical data: Geophysics, **75**, no. 6, G53–G64, doi: 10.1190/1.3496476.

Eidsvik, J., T. Mukerji, and P. Switzer, 2004, Estimation of geological attributes from a well log: An application of hidden Markov chains: Mathematical Geology, **36**, 379–397, doi: 10.1023/B:MATG.0000028443.75501.d9.

Farmani, M. B., N. O. Kitterød, and H. Keers, 2008, Inverse modeling of unsaturated flow parameters using dynamic geological structure conditioned by GPR tomography: Water Resources Research, **44**, W08401, doi: 10.1029/2007WR006251.

Flinchum, B. A., W. S. Holbrook, D. Grana, A. D. Parsekian, B. J. Carr, J. L. Hayes, and J. Jiao, 2018b, Estimating the water holding capacity of the critical zone using near-surface geophysics: Hydrological Processes, **32**, 3308–3326, doi: 10.1002/hyp.13260.

Flinchum, B. A., W. S. Holbrook, D. Rempe, S. Moon, C. S. Riebe, B. J. Carr, J. L. Hayes, J. St Clair, and M. P. Peters, 2018a, Critical zone structure under a granite ridge inferred from drilling and three-dimensional seismic refraction data: Journal of Geophysical Research: Earth Surface, **123**, 1317–1343, doi: 10.1029/2017JF004280.

Fullhart, A. T., T. J. Kelleners, D. G. Chandler, J. P. McNamara, and M. S. Seyfried, 2018, Water flow modeling with dry bulk density optimization to determine hydraulic properties in mountain soils: Soil Science Society of America Journal, **82**, 31–44, doi: 10.2136/sssaj2017.06.0196.

Gallardo, L. A., and M. A. Meju, 2003, Characterization of heterogeneous near-surface materials by joint 2D inversion of DC resistivity and seismic data: Geophysical Research Letters, **30**, 1658, doi: 10.1029/2003GL017370.

Grana, D., 2018, Joint facies and reservoir properties inversion: Geophysics, **83**, no. 3, M15–M24, doi: 10.1190/geo2017-0670.1.

Grana, D., and E. Della Rossa, 2010, Probabilistic petrophysical-properties estimation integrating statistical rock physics with seismic inversion: Geophysics, **75**, no. 3, O21–O37, doi: 10.1190/1.3386676.

Grana, D., T. Fjeldstad, and H. Omre, 2017, Bayesian Gaussian mixture linear inversion for geophysical inverse problems: Mathematical Geosciences, **49**, 493–515, doi: 10.1007/s11004-016-9671-9.

Grana, D., M. Pirrone, and T. Mukerji, 2012, Quantitative log interpretation and uncertainty propagation of petrophysical properties and facies classification from rock-physics modeling and formation evaluation analysis: Geophysics, **77**, no. 3, WA45–WA63, doi: 10.1190/geo2011-0272.1.

Grana, D., K. Schlanser, and E. Campbell-Stone, 2015, Petroelastic and geomechanical classification of lithologic facies in the Marcellus Shale: Interpretation, **3**, no. 1, SA51–SA63, doi: 10.1190/INT-2014-0047.1.

Groves, D. I., R. M. Vielreicher, R. J. Goldfarb, and K. C. Condie, 2005, Controls on the heterogeneous distribution of mineral deposits through time, *in* I. McDonald, D. A. Boyce, A. J. Butler, I. B. Herrington, and R. J. Polya, Geological Society Special Publication, **248**, 71–101.

Hachmöller, B., and H. Paasche, 2013, Integration of surface-based tomographic models for zonation and multimodel guided extrapolation of sparsely known petrophysical parameters: Geophysics, **78**, no. 4, EN43–EN53, doi: 10.1190/geo2012-0417.1.

Hashin, Z., and S. Shtrikman, 1963, A variational approach to the theory of the elastic behaviour of multiphase materials: Journal of the Mechanics and Physics of Solids, **11**, 127–140, doi: 10.1016/0022-5096(63)90060-7.

Hasselblad, V., 1966, Estimation of parameters for a mixture of normal distributions: Technometrics, **8**, 431–444, doi: 10.1080/00401706.1966.10490375.

Hastie, T., R. Tibshirani, and J. Friedman, 2002, The elements of statistical learning: Springer.

Hermans, T., and J. Irving, 2017, Facies discrimination with electrical resistivity tomography using a probabilistic methodology: Effect of sensitivity and regularisation: Near Surface Geophysics, **15**, 13–25, doi: 10.3997/1873-0604.2016047.

Holbrook, W. S., C. S. Riebe, M. Elwaseif, J. L. Hayes, K. Basler-Reeder, D. L. Harry, A. Malazian, A. Dosseto, P. C. Hartsough, and J. W. Hopmans, 2014, Geophysical constraints on deep weathering and water storage potential in the Southern Sierra Critical Zone Observatory: Earth Surface Processes and Landforms, **39**, 366–380, doi: 10.1002/esp.3502.

Hübner, R., K. Heller, T. Günther, and A. Kleber, 2015, Monitoring hillslope moisture dynamics with surface ERT for enhancing spatial significance of hydrometric point measurements: Hydrology and Earth System Sciences, **19**, 225–240.

Knight, R. J., and A. L. Endres, 2005, An introduction to rock physics principles for near-surface geophysics, *in* D. K. Butler, ed., Near-surface geophysics: SEG, 31–70.

Kotikian, M., A. D. Parsekian, G. Paige, and A. Carey, 2019, Observing heterogeneous unsaturated flow at the hillslope scale using time-lapse electrical resistivity tomography: Vadose Zone Journal, **18**, 1–16, doi: 10.2136/vzj2018.07.0138.

Leopold, M., J. Völkel, J. Huber, and D. Dethier, 2013, Subsurface architecture of the Boulder Creek Critical Zone Observatory from electrical resistivity tomography: Earth Surface Processes and Landforms, **38**, 1417–1431.

Liefert, D. T., and B. N. Shuman, 2020, Pervasive desiccation of North American lakes during the Late Quaternary: Geophysical Research Letters, **47**, e2019GL086412.

Lindberg, D. V., and D. Grana, 2015, Petro-elastic log-facies classification using the expectation–maximization algorithm and hidden Markov models: Mathematical Geosciences, **47**, 719–752, doi: 10.1007/s11004-015-9604-z.

Loritz, R., S. K. Hassler, C. Jackisch, N. Allrogen, L. van Schaik, J. Wienhofer, and E. Zehe, 2017, Picturing and modeling catchments by representative hillslopes: Hydrology and Earth System Science, **21**, 1225–1249, doi: 10.5194/hess-21-1225-2017.

Mavko, G., and T. Mukerji, 1998, A rock physics strategy for quantifying uncertainty in common hydrocarbon indicators: Geophysics, **63**, 1997–2008, doi: 10.1190/1.1444493.

McIntosh, J. C., C. Schaumberg, J. Perdrial, A. Harpold, A. Vázquez-Ortega, C. Rasmussen, D. Vinson, X. Zapata-Rios, P. D. Brooks, T. Meixner, and J. Pelletier, 2017, Geochemical evolution of the Critical Zone across variable time scales informs concentration-discharge relationships: Jemez River Basin Critical Zone Observatory: Water Resources Research, **53**, 4169–4196, doi: 10.1002/2016WR019712.

McNamara, J. P., D. Chandler, M. Seyfried, and S. Achet, 2005, Soil moisture states, lateral flow, and streamflow generation in a semi-arid, snowmelt-driven catchment: Hydrological Processes, **19**, 4023–4038, doi: 10.1002/hyp.5869.

Meju, M. A., L. A. Gallardo, and A. K. Mohamed, 2003, Evidence for correlation of electrical resistivity and seismic velocity in heterogeneous near-surface materials: Geophysical Research Letters, **30**, 1373, doi: 10.1029/2002GL016048.

Miller, C. R., P. S. Routh, T. R. Brosten, and J. P. McNamara, 2008, Application of time-lapse ERT imaging to watershed characterization: Geophysics, **73**, no. 3, G7–G17, doi: 10.1190/1.2907156.

Mills, H. H., 1990, Thickness and character of regolith on mountain slopes in the vicinity of Mountain Lake, Virginia, as indicated by seismic refraction, and implications for hillslope evolution: Geomorphology, **3**, 143–157, doi: 10.1016/0169-555X(90)90042-O.

Mindlin, R. D., 1949, Compliance of elastic bodies in contact: Journal of Applied Mechanics, **16**, 259–268.

Moraes, T., 2019, Monitoring non-uniform infiltration of snow melt using time-lapse electrical resistivity tomography: M.S. thesis, University of Wyoming.

Moravec, B. G., A. M. White, R. A. Root, A. Sanchez, Y. Olshansky, B. K. Paras, B. Carr, J. McIntosh, J. D. Pelletier, C. Rasmussen, and W. S. Holbrook, 2020, Resolving deep critical zone architecture in complex volcanic terrain: Journal of Geophysical Research, Earth Surface, **125**, 1–24, doi: 10.1029/2019JF005189.

Nielson, T., J. H. Bradford, and W. S. Holbrook, 2015, Geophysical investigation of differences in weathering depths between the north and south facing slopes of a small catchment in the Reynolds Creek Critical Zone Observatory: AGU Fall Meeting Abstracts.

Oldenburg, D. W., and Y. Li, 1999, Estimating depth of investigation in DC resistivity and IP surveys: Geophysics, **64**, 403–416, doi: 10.1190/1.1444545.

Olyphant, J., J. D. Pelletier, and R. Johnson, 2016, Topographic correlations with soil and regolith thickness from shallow-seismic refraction constraints across Upland Hillslopes in the Valles Caldera, New Mexico: Earth Surface Processes and Landforms, **41**, 1684–1696, doi: 10.1002/esp.3941.

Onda, Y., M. Tsujimura, J. Fujihara, and J. Ito, 2006, Runoff generation mechanisms in high-relief mountainous watersheds with different underlying geology: Journal of Hydrology, **331**, 659–673, doi: 10.1016/j.jhydrol.2006.06.009.

Orlando, J., X. Comas, S. A. Hynek, H. L. Buss, and S. L. Brantley, 2016, Architecture of the deep critical zone in the Río Icacos watershed (Luquillo Critical Zone Observatory, Puerto Rico) inferred from drilling and ground penetrating radar (GPR): Earth Surface Processes and Landforms, **41**, 1826–1840, doi: 10.1002/esp.3948.

Paasche, H., and D. Eberle, 2011, Automated compilation of pseudo-lithology maps from geophysical data sets: A comparison of Gustafson-Kessel and fuzzy *c*-means cluster algorithms: Exploration Geophysics, **42**, 275–285, doi: 10.1071/EG11014.

Paasche, H., J. Tronicke, and P. Dietrich, 2010, Automated integration of partially colocated models: Subsurface zonation using a modified fuzzy *c*-means cluster analysis algorithm: Geophysics, **75**, no. 3, P11–P22, doi: 10.1190/1.3374411.

Paasche, H., J. Tronicke, K. Holliger, A. G. Green, and H. Maurer, 2006, Integration of diverse physical-property models: Subsurface zonation and petrophysical parameter estimation based on fuzzy *c*-means cluster analyses: Geophysics, **71**, no. 3, H33–H44, doi: 10.1190/1.2192927.

Parsekian, A. D., N. Claes, K. Singha, B. J. Minsley, B. J. Carr, E. Voytek, R. Harmon, M. A. Kass, A. Carey, D. Thayer, and B. Flinchum, 2017, Comparing measurement response and inverted results of electrical resistivity tomography instruments: Journal of Environmental and Engineering Geophysics, **22**, 249–266, doi: 10.2113/JEEG22.3.249.

Parsekian, A. D., T. J. Kelleners, F. A. Neves, M. Pleasants, and D. Grana, 2020, Geophysical measurements of CZO hillslopes and expectation maximization classification, http://www.hydroshare.org/resource/775a3bd3d8674bc38ceb160b902e5056.
Parsekian, A. D., K. Singha, B. J. Minsley, W. S. Holbrook, and L. Slater, 2015, Multiscale geophysical imaging of the critical zone: Reviews of Geophysics, **53**, doi: 10.1002/2014RG000465.
Pelletier, J. D., G. A. Barron-Gafford, H. Gutiérrez-Jurado, E. L. S. Hinckley, E. Istanbulluoglu, L. A. McGuire, G. Y. Niu, M. J. Poulos, C. Rasmussen, P. Richardson, and T. L. Swetnam, 2018, Which way do you lean? Using slope aspect variations to understand Critical Zone processes and feedbacks: Earth Surface Processes and Landforms, **43**, 1133–1154, doi: 10.1002/esp.4306.
Pleasants, M. S., T. J. Kelleners, and N. Ohara, 2017, Analysis of snowpack dynamics during the spring melt season for a sub-alpine site using point measurements and numerical modeling: Hydrological Processes, **31**, 4568–4585, doi: 10.1002/hyp.11379.
Riebe, C. S., W. J. Hahm, and S. L. Brantley, 2017, Controls on deep critical zone architecture: A historical review and four testable hypotheses: Earth Surface Processes and Landforms, **42**, 128–156, doi: 10.1002/esp.4052.
Robinson, J. S., M. Sivapalan, and J. D. Snell, 1995, On the relative roles of hillslope processes, channel routing, and network geomorphology in the hydrologic response of natural catchments: Water Resources Research, **31**, 3089–3101.
Rücker, C., T. Günther, and F. M. Wagner, 2017, pyGIMLi: An open-source library for modelling and inversion in geophysics: Computers and Geosciences, **109**, 106–123, doi: 10.1016/j.cageo.2017.07.011.
Rudman, A. J., J. F. Whaley, R. F. Blakely, and M. E. Biggs, 1976, Transformation of resistivity to pseudovelocity logs: AAPG Bulletin, **60**, 879–882, doi: 10.1306/83D91F47-16C7-11D7-8645000102C1865D.
Salvucci, G. D., and D. Entekhabi, 1995, Hillslope and climatic controls on hydrologic fluxes: Water Resources Research, **31**, 1725–1739, doi: 10.1029/95WR00057.
Schaap, M. G., F. J. Leij, and M. T. Van Genuchten, 2001, Rosetta: A computer program for estimating soil hydraulic parameters with hierarchical pedotransfer functions: Journal of Hydrology, **251**, 163–176, doi: 10.1016/S0022-1694(01)00466-8.
Seyfried, M., K. Lohse, D. Marks, G. Flerchinger, F. Pierson, and W. S. Holbrook, 2018, Reynolds creek experimental watershed and critical zone observatory: Vadose Zone Journal, **17**, 180129, doi: 10.2136/vzj2018.07.0129.
Shannon, C. E., 1948, A mathematical theory of communication: Bell System Technical Journal, **27**, 379–423, doi: 10.1002/j.1538-7305.1948.tb01338.x.
Slater, L., A. M. Binley, W. Daily, and R. Johnson, 2000, Cross-hole electrical imaging of a controlled saline tracer injection: Journal of Applied Geophysics, **44**, 85–102, doi: 10.1016/S0926-9851(00)00002-1.
Slater, L. D., and D. Lesmes, 2002, IP interpretation in environmental investigations: Geophysics, **67**, 77–88, doi: 10.1190/1.1451353.
Steindlberger, E., 2004, Volcanic tuffs from Hesse (Germany) and their weathering behaviour: Environmental Geology, **46**, 378–390, doi: 10.1007/s00254-004-1039-7.
Sun, J., and Y. Li, 2016, Joint inversion of multiple geophysical and petrophysical data using generalized fuzzy clustering algorithms: Geophysical Supplements to the Monthly Notices of the Royal Astronomical Society, **208**, 1201–1216, doi: 10.1093/gji/ggw442.
Thayer, D., A. D. Parsekian, K. Hyde, H. Speckman, D. Beverly, B. Ewers, M. Covalt, N. Fantello, T. Kelleners, N. Ohara, and T. Rogers, 2018, Geophysical measurements to determine the hydrologic partitioning of snowmelt on a snow-dominated subalpine hillslope: Water Resources Research, **54**, 3788–3808, doi: 10.1029/2017WR021324.
United States Environmental Protection Agency, 2020, Version 1.10. WATERSKMZ, Retrieved: 23 February 2020, from URL https://www.epa.gov/sites/production/files/2020-01/waterskmz_v1.10.kmz.
Walsh, D., P. Turner, E. Grunewald, H. Zhang, J. J. Butler, Jr., E. Reboulet, S. Knobbe, T. Christy, J. W. Lane, Jr., C. D. Johnson, and T. Munday, 2013, A small-diameter NMR logging tool for groundwater investigations: Groundwater, **51**, 914–926, doi: 10.1111/gwat.12024.
Waxman, M. H., and L. J. M. Smits, 1968, Electrical conductivities in oil-bearing shaly sands: Society of Petroleum Engineers Journal, **8**, 107–122, doi: 10.2118/1863-A.
West, N., E. Kirby, A. A. Nyblade, and S. L. Brantley, 2019, Climate preconditions the Critical Zone: Elucidating the role of subsurface fractures in the evolution of asymmetric topography: Earth and Planetary Science Letters, **513**, 197–205, doi: 10.1016/j.epsl.2019.01.039.
White, A., B. Moravec, J. McIntosh, Y. Olshansky, B. Paras, R. A. Sanchez, T. P. A. Ferré, T. Meixner, and J. Chorover, 2019, Distinct stores and the routing of water in the deep critical zone of a snow-dominated volcanic catchment: Hydrology and Earth System Sciences, **23**, 4661–4683, doi: 10.5194/hess-23-4661-2019.
White, A. F., and S. L. Brantley, 2003, The effect of time on the weathering of silicate minerals: Why do weathering rates differ in the laboratory and field? Chemical Geology, **202**, 479–506, doi: 10.1016/j.chemgeo.2003.03.001.
Yang, Y. P., and D. W. Emerson, 1997, Electromagnetic conductivities of rock cores: Theory and analog results: Geophysics, **62**, 1779–1793, doi: 10.1190/1.1444278.

Biographies and photographs of the authors are not available.