A Multi-Scale Soil Moisture Monitoring Strategy for California: Design and Validation

Jennifer A. Curtis, Lorraine E. Flint, and Michelle A. Stern

Research Impact Statement: A multi-scale soil moisture monitoring strategy identifies where monitoring will best capture soil moisture response units that exhibit different hydrologic responses to changes in climate.

ABSTRACT: A multi-scale soil moisture monitoring strategy for California was designed to inform water resource management. The proposed workflow classifies soil moisture response units (SMRUs) using publicly available datasets that represent soil, vegetation, climate, and hydrology variables, which control soil water storage. The SMRUs were classified, using principal component analysis and unsupervised K-means clustering within a geographic information system, and validated, using summary statistics derived from measured soil moisture time series. Validation stations, located in the Sierra Nevada, include transect of sites that cross the rain-to-snow transition and a cluster of sites located at similar elevations in a snow-dominated watershed. The SMRUs capture unique responses to varying climate conditions characterized by statistical measures of central tendency, dispersion, and extremes. A topographic position index and landform classification is the final step in the workflow to guide the optimal placement of soil moisture sensors at the local-scale. The proposed workflow is highly flexible and can be implemented over a range of spatial scales and input datasets can be customized. Our approach captures a range of soil moisture responses to climate across California and can be used to design and optimize soil moisture monitoring strategies to support runoff forecasts for water supply management or to assess landscape conditions for forest and rangeland management.

(KEYWORDS: soil moisture; runoff; monitoring; geospatial analysis; statistics; water resource management.)

INTRODUCTION

Expected increases in climatic variability (IPCC 2014) and more frequent episodes of water scarcity make strategic management of water resources increasingly important. As climate change results in novel hydrologic conditions, real-time soil moisture monitoring is now considered an integral component of operational water resource management at local to regional scales (Schaefer et al. 2007; Illston et al. 2008; Dobriyal et al. 2012; Trenberth and Asrar 2014; Harpold et al. 2017).

Soil moisture monitoring and early recognition of anomalous conditions can improve runoff forecasting and support time-sensitive implementation of effective water-management solutions. For example, when spring snowmelt occurs on dry soils, a portion of snowmelt must fill the existing soil water deficit and runoff generation is both delayed and diminished. Deficits in soil moisture can result in the earlier onset of low flow conditions, while saturated soils more readily produce runoff that can result in earlier and larger streamflow peaks.

Incorporating soil moisture conditions into runoff forecast models will enable more accurate predictions...
of the timing and volume of runoff (Crow and Ryu 2009; Tayfur et al. 2014; Harpold et al. 2017). This information will greatly improve water resource management and reservoir operations; however, the ability to manage for anomalous soil moisture conditions requires new water-management strategies supported by real-time soil moisture monitoring.

Despite the importance of soil moisture in determining water balance conditions and associated hydrologic responses (Flint et al. 2008), there are relatively few soil moisture monitoring stations in comparison to streamflow stations (Quiring et al. 2015). There are even fewer soil moisture stations that measure deep subsurface soil moisture (>0.5 m) because many of the existing monitoring networks were developed for the calibration of remote sensing products (Larson et al. 2008, 2010; Entekhabi et al. 2010). These shallow networks capture only the top few centimeters (~5 cm); (Ochsner et al. 2013) and do not characterize the deeper root zone soil moisture that is utilized by transpiring vegetation. These shallow monitoring networks cannot effectively characterize surplus or deficit soil moisture conditions that exert first order controls on hydrologic responses, and the spatial resolution of satellite observations of soil moisture is too coarse for many hydrologic applications (Peng et al. 2017). This highlights the need for real-time soil moisture monitoring to characterize soil water storage.

Hydrologic classifications, which group rivers into hydrologic response units with similar streamflow regimes, have been implemented successfully over a range of spatial scales (Wolock et al. 2004; Olden et al. 2012). Unlike streamflow, which represents a cumulative watershed response, soil moisture data are fine-scale point measurements with high spatial and temporal variability (Vereecken et al. 2014), which creates a unique challenge for classification.

Soil moisture patterns have been described using cluster analysis at the watershed (Lin et al. 2006; Schrot et al. 2015) and hillslope-scale (Lee and Kim 2017; Liao et al. 2017). Hydrologic-based soil classifications have been developed using conceptual models of soil response (Boorman et al. 1995). Building on these studies, we developed an analytical workflow to classify soil moisture response units (SMRUs) at the regional to local-scale across California. The proposed workflow can be used to optimize the number and location of soil moisture stations, to design monitoring strategies that capture the range of soil moisture responses to climate, and to inform water resource management.

This study represents the first attempt to define SMRUs at a fine-spatial scale (~270 m) throughout California (Figure 1). A SMRU is defined as a landscape region with similar soil, climatic, hydrologic, and vegetation characteristics that produce a similar soil moisture response to climate. The size of a SMRU can range from relatively small areas that are typically located in mountainous regions, to very large areas, which tend to be located in gently sloping valleys. We developed a deductive approach that uses unsupervised cluster analysis to classify SMRUs across California. Deductive methods, which are often referred to as “top down” methods, approach classification from the general to the specific. This type of deductive classification is commonly used when an unsupervised approach, which does not constrain the classification process, is required due to the absence or scarcity of measured data (Sayre et al. 2009; Olden et al. 2012; Auerbach et al. 2016). For these reasons, unsupervised cluster analysis is considered an exploratory first principle tool that uses pattern recognition to define statistical groupings, which maximize within group similarities and between group differences.

Because of the spatial variability of soil properties, a primary challenge in the implementation of soil moisture monitoring is optimizing the number and location of sensors to represent the range of potential conditions (Robinson et al. 2008). We used preexisting environmental datasets discretized at a 270-m spatial resolution (Flint and Flint 2014; Wieczorek 2014, 2015; http://phenology.cr.usgs.gov; Table 1) and then validated watershed-scale SMRUs using measured soil moisture time series. Our workflow generates a series of map products that identify where monitoring will best support water resource management. We further illustrate the workflow with an example application in the Hetch-Hetchy water supply basin and discuss other predictive applications.

METHODS

Study Area

California’s Mediterranean climate is naturally variable with large latitude and elevation gradients that produce strong climate gradients. For example, California’s north coast region receives over 2,500 mm of average annual precipitation compared to 130 mm in the southeastern deserts. About 75% of the state’s annual precipitation falls within a narrow window of time between November and March. Annual precipitation anomalies are routinely 50%–200% of long-term averages, which result in greater inter-annual variability than most other locations in the conterminous United States (U.S.) (Dettinger et al. 2011). Rain-dominated landscapes occur at
lower elevations and the rain-to-snow transition occurs at about 2,000–2,500 m elevation (Curtis et al. 2014). The state’s largest reservoirs are located in northern California where water is captured and stored during the runoff season. Stored water is then released and transferred to population and agricultural centers in southern California via a complex system of dams, canals, aqueducts, levees, bypasses, weirs, and pumping plants.

We selected study regions at three nested-spatial scales to optimize our workflow for classifying SMRUs (Figure 1). The regional scale includes the entire state of California (CA) (419,550 km²), which can be subdivided into 10 unique Jepson ecological regions (Hickman 1993) on the basis of similar air temperature and precipitation patterns. At the intermediate scale, we selected the Sierra Nevada ecoregion (SN) (62,560 km²), where more than 60% of California’s developed water supply originates (http://www.sierannevada.ca.gov/our-region/ca-primary-watershed). At the watershed-scale, we selected the combined Merced-Tuolumne watersheds (MT) (7,010 km²), which include the rain-to-snow transition boundary.

We further illustrate our workflow with example applications in the Hetch-Hetchy (1,175 km²) and Upper Carson River (2,480 km²) watersheds (CR). We selected the Hetch-Hetchy watershed because it is an important water supply basin. Hetch-Hetchy reservoir, one of California’s larger water supply reservoirs, supplies water via an aqueduct to 2.4 million San Francisco-Bay Area residents. Because much of California’s surface water supply is derived from Sierra Nevada snowmelt we selected

| Variable abbreviation | Soil response variable | Soil response variable type | Description                                                                 | Source                                      |
|-----------------------|------------------------|----------------------------|------------------------------------------------------------------------------|---------------------------------------------|
| psand                 | Percent sand           | Soil                       | Weighted average of sand throughout soil profile, in percent                 | Wieczorek (2014, 2015)                      |
| pclay                 | Percent clay           | Soil                       | Weighted average clay throughout soil profile, in percent                    | Wieczorek (2014, 2015)                      |
| por                   | Porosity               | Soil                       | Weighted average of porosity throughout soil profile, in meters/meter        | Wieczorek (2014, 2015)                      |
| thk                   | Soil depth             | Soil                       | Soil depth, in meters                                                        | Wieczorek (2014, 2015)                      |
| awc                   | Available water capacity| Soil                      | Water content at field capacity minus water content at wilting point multiplied by soil depth, in meters/meter | Wieczorek (2014, 2015)                      |
| ppt                   | Average annual precipitation | Climate                | Average precipitation from 1981 to 2010, in mm                               | Flint and Flint (2014)                      |
| apr1pck               | Average April 1 snowpack | Climate                | Average April 1 snowpack from 1981 to 2010, in mm snow water equivalent     | Flint and Flint (2014)                      |
| snowstd               | Standard deviation of April 1 snowpack | Climate | Standard deviation of April 1 snowpack from 1981 to 2010, in mm snow water equivalent | Flint and Flint (2014)                      |
| pet                   | Average annual potential evapotranspiration (PET) | Climate | Average PET from 1981 to 2010, in mm                                         | Flint and Flint (2014)                      |
| cwd                   | Average annual climatic water deficit (CWD) | Hydrologic | Average CWD from 1981 to 2010, in mm, calculated as PET minus actual evapotranspiration and accumulated annually | Flint and Flint (2014)                      |
| rch                   | Average annual recharge | Hydrologic               | Average recharge from 1981 to 2010, in mm                                    | Flint and Flint (2014)                      |
| run                   | Average annual runoff  | Hydrologic               | Average runoff from 1981 to 2010, in mm                                      | Flint and Flint (2014)                      |
| NDVI                  | Daily interpolated integration of NDVI above the baseline for the entire duration of the growing season | Vegetation | Average time-integrated NDVI from 2001 to 2015                               | USGS EROS Center (http://phenology.cr.usgs.gov) |

Notes: NDVI, Normalized Difference Vegetation Index; USGS, United States Geological Survey; EROS, Earth Resources Observation and Science.
the Upper Carson River watershed, to further validate our SMRU classification in a snow-dominated basin.

**Multivariate Analysis Geographic Information System-Workflow**

A multivariate workflow (Figure 2) was developed to classify SMRUs and to support design and optimization of the number and location of soil moisture monitoring stations. Publicly available environmental raster datasets (Flint and Flint 2014; Wieczorek 2014, 2015, U.S. Geological Survey (USGS) Remote Sensing Phenology: http://phenology.cr.usgs.gov) were reviewed and a list of 13 datasets that govern soil moisture response to climate were selected as explanatory variables (Table 1). The explanatory variables, referred to here as soil moisture response variables, represent a range of soil, climate, hydrology, and vegetation properties. Variables that describe soil properties included: soil depth, percent sand, percent clay, porosity, and available water capacity. All the soil variables, except soil depth, were weighted averages computed for the entire soil profile. Available water capacity, defined as the amount of water a soil can store, was calculated as water content at field capacity (i.e., the amount of
water stored after excess water has drained) minus water content at wilting point (i.e., a minimal point of soil moisture before plants begin to wilt) multiplied by soil depth. Climate variables were average values computed for the 1981–2010 period and included: precipitation, April 1 snowpack, potential evapotranspiration (PET), and the standard deviation of April 1 snowpack, which is a measure of snowpack variability that highlights the dynamic rain-to-snow transition zone. Hydrology variables were average annual values computed for the 1981–2010 period and included: recharge, runoff, and climatic water deficit (CWD). CWD was calculated as the evaporative demand (PET) minus available water (actual ET) accumulated annually. Vegetation was characterized using a time-integrated Normalized Difference Vegetation Index (NDVI), derived for the period spanning 2001–2015. The integrated NDVI characterizes plant
"greenness" or photosynthetic activity across the landscape.

Prior to classification, the 13 soil moisture response raster datasets (Table 1) were masked and standardized. A mask was developed using a land-use map from the California Department of Forestry’s Fire and Resource Assessment Program (Department of Forestry and Fire Protection, FRAP 2015). The land-use mask eliminated locations where soil moisture monitoring was either unfeasible or not recommended. Masked land-use areas were classified as: water, urban, barren, irrigated crops, agricultural land, and marsh land. An additional mask was derived from existing recharge and runoff maps (Flint and Flint 2014) to eliminate areas that did not produce enough recharge and/or runoff to be considered useful in a water supply context, which applies primarily to desert regions. Recharge plus runoff maps were created and a minimum threshold was set at 10 mm/year. All regions below this threshold were eliminated from our analysis. The 13 masked rasters were then standardized to convert all variables to a similar scale. This method reduces the artificial dominance of any single variable but allows extreme conditions to be preserved. A z-score method (Equation 1) was used to standardize the masked soil response rasters,

\[ z = \frac{(n - \bar{x})}{s}, \]

where \( z \) is the standardized value, \( n \) is the original raster pixel value, \( \bar{x} \) is the mean of the raster layer, and \( s \) is the standard deviation of the raster layer.

**Principal Component Analysis.** Each of the 13 soil moisture response variables was correlated with at least one other variable at all three spatial scales (Tables 2–4). Because data redundancy complicates pattern recognition, a key decision is how many variables should be used for classification. Although methods for eliminating variables are generally heuristic, the goal is to minimize the number of variables that can be used to explain the pattern of variance across the study region.

We used principal component analysis (PCA), a common statistical method (Pearson 1901; Abdi and Williams 2010; Shlens 2014), to define an optimal number of uncorrelated variables for cluster analysis and SMRU classification. An orthogonal transformation was used to convert the soil moisture response variables into a set of linearly uncorrelated principal components (PCs). Orthogonal transformation moves the largest possible variance into the first PC and each subsequent PC has the highest variance possible under the constraint that it is orthogonal to the prior component.

A separate PCA was completed for each study region using the 13 standardized soil moisture response rasters (Table 1) and the PRINCIPAL COMPONENTS function in ArcGIS (http://pro.arcgis.com/en/pro-app/tool-reference/spatial-analyst/principal-components.htm). Note that ArcGIS functions are denoted here by capitalizing the function name. Again, all raster datasets were discretized at a 270-m spatial resolution.

There are numerous methods for selecting the number of PCs to retain for further analysis (Henson and Roberts 2006). A common approach is to simply select the first two PCs for further analysis. For example, Thorne et al. (2017) used PCA in a study that determined the sensitivity of vegetation groups to future climate distributions across California, Thorne et al. (2017) selected the first two PCs, which

| TABLE 2. Correlation matrix of 13 variables for the CA-scale cluster analysis. |
|-----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
|      | rch   | cwd   | pet   | psand | awc   | por   | thk   | pclay | run   | apr1pck | ppt   | NDVI   | snowstd |
|------|-------|-------|-------|-------|-------|-------|-------|-------|-------|---------|-------|--------|---------|
| rch  | —     | -0.43 | -0.39 | -0.17 | 0.15  | 0.11  | 0.12  | 0.12  | 0.25  | 0.14    | 0.74  | 0.09   | 0.07    |
| cwd  | -0.43 | —     | 0.93  | 0.18  | -0.47 | -0.51 | -0.35 | -0.12 | -0.51 | -0.54   | -0.65 | -0.41  | -0.32   |
| pet  | -0.39 | 0.93  | —     | -0.32 | -0.45 | -0.21 | -0.18 | -0.46 | -0.47 | -0.55   | -0.34 | -0.27  | —       |
| psand| -0.17 | 0.18  | 0.22  | —     | -0.04 | -0.19 | -0.03 | -0.89 | -0.06 | 0.23    | -0.18 | -0.07  | 0.15    |
| awc  | 0.15  | -0.47 | -0.32 | -0.04 | —     | 0.49  | 0.88  | 0.22  | —     | 0.19    | 0.22  | 0.29   | 0.10    |
| por  | 0.11  | -0.51 | -0.45 | -0.19 | 0.49  | —     | 0.22  | 0.16  | 0.15  | 0.27    | 0.22  | 0.26   | 0.15    |
| thk  | 0.12  | -0.35 | -0.21 | -0.03 | 0.88  | 0.22  | —     | 0.05  | -0.02 | 0.10    | 0.15  | 0.23   | 0.05    |
| pclay| 0.12  | -0.12 | -0.18 | -0.89 | 0.00  | 0.16  | 0.05  | —     | 0.05  | -0.27   | 0.10  | 0.02   | 0.17    |
| run  | 0.25  | -0.51 | -0.46 | -0.06 | 0.04  | 0.15  | -0.02 | 0.00  | —     | 0.47    | 0.81  | 0.24   | 0.25    |
| apr1pck| 0.14 | -0.54 | -0.47 | 0.23  | 0.19  | 0.27  | 0.10  | -0.27 | 0.47  | —       | 0.39  | 0.54   | 0.54    |
| ppt  | 0.74  | -0.65 | -0.55 | -0.18 | 0.22  | 0.22  | 0.15  | 0.10  | 0.81  | 0.39    | 0.25  | 0.20   | 0.25    |
| NDVI | 0.09  | -0.41 | -0.34 | 0.07  | 0.29  | 0.26  | 0.23  | 0.02  | 0.24  | 0.54    | 0.25  | 0.26   | —       |
| snowstd| 0.07 | -0.33 | -0.27 | 0.15  | 0.10  | 0.15  | 0.05  | -0.17 | 0.25  | 0.54    | 0.20  | 0.26   | —       |

Note: Variables exceeding a correlation threshold of ±0.50 are highlighted in bold. A correlation coefficient of ±1.0 indicates a perfect positive correlation, -1.0 indicates a perfect inverse correlation, and zero suggests that there is no correlation. Variables are described in Table 1.
explained approximately 80% of the variance, for their climate exposure analysis.

We selected PCs to retain for the cluster analysis and SMRU classification using a break in slope scree-test criterion (Cattell 1966), which is a common empirical method that defines a suite of PCs with the largest explanatory power. Larger eigenvalues represent greater explanatory power and the eigenvalues decrease sequentially for each additional PC, while asymptotically approaching zero. When the eigenvalues are plotted, a user-defined break in slope scree-test criterion defines the optimal number of PCs for further analysis. The scree-test criterion method allows the user to objectively identify the optimal number of PCs with adequate explanatory power for further analysis.

**Cluster Analysis.** The PCs for each study region with the largest explanatory power were used to classify SMRUs using an unsupervised $K$-means cluster analysis (Jain 2010). Separate cluster analyses were performed for each study region using the ISOCLUSTER function in ArcGIS (http://pro.arcgis.com/en/pro-app/tool-reference/spatial-analyst/iso-cluster.htm). ISOCLUSTER uses $K$-means clustering, one of the most widely used classification algorithms (Pang-Ning et al. 2006), and an optimization clustering algorithm, known as the migrating means technique, which iteratively minimizes the sum of the squared error. Error is defined as the distance between the raster cell’s multi-dimensional value and a cluster centroid. The cluster centroid is defined as the group mean using multi-dimensional squared error. Error is defined as the distance between a point and the nearest cluster centroid.

**TABLE 3. Correlation matrix of 13 variables for the SN-scale cluster analysis.**

|     | rch | cwd | pet | psand | awc | por | thk | pplay | run | aprlpc | ppt | NDVI | snowstd |
|-----|-----|-----|-----|-------|-----|-----|-----|-------|-----|--------|-----|-------|----------|
| rch | —   | —   | —   | —     | —   | —   | —   | —     | —   | —      | —   | —     | —        |
| cwd | 0.35| —   | —   | 0.89  | —   | —   | —   | —     | —   | —      | —   | —     | 0.37     |
| pet | —   | —   | 0.19| —     | 0.33| —   | —   | —     | —   | —      | —   | —     | 0.08     |
| psand| —   | —   | —   | —     | —   | —   | —   | —     | —   | —      | —   | —     | 0.08     |
| awc | 0.28| —   | —   | —     | 0.31| —   | —   | —     | —   | —      | —   | —     | 0.09     |
| por | —   | —   | 0.34| —     | —   | —   | —   | —     | —   | —      | —   | —     | 0.09     |
| thk | 0.12| —   | —   | —     | —   | 0.90| —   | —     | —   | —      | —   | —     | 0.17     |
| pplay| 0.03| 0.25| —   | —     | 0.13| 0.10| —   | —     | —   | —      | —   | —     | 0.14     |
| run | —   | —   | 0.52| —     | 0.11| —   | —   | —     | —   | —      | —   | —     | 0.14     |
| aprlpc| 0.30| —   | —   | 0.37  | 0.17| —   | —   | —     | —   | —      | —   | —     | 0.15     |
| ppt | 0.57| —   | —   | —     | 0.42| 0.14| —   | —     | —   | —      | —   | —     | 0.13     |
| NDVI| 0.12| —   | —   | 0.21  | 0.03| 0.08| —   | —     | —   | —      | —   | —     | 0.08     |
| snowstd| 0.08| —   | —   | —     | 0.21| 0.01| —   | —     | —   | —      | —   | —     | —        |

Note: Variables exceeding a correlation threshold of ±0.50 are highlighted in bold. A correlation coefficient of +1.0 indicates a perfect positive correlation, −1.0 indicates a perfect inverse correlation, and zero suggests that there is no correlation. Variables are described in Table 1.

**TABLE 4. Correlation matrix of 13 variables for the MT-scale cluster analysis.**

|     | rch | cwd | pet | psand | awc | por | thk | pplay | run | aprlpc | ppt | NDVI | snowstd |
|-----|-----|-----|-----|-------|-----|-----|-----|-------|-----|--------|-----|-------|----------|
| rch | —   | —   | —   | —     | —   | —   | —   | —     | —   | —      | —   | —     | —        |
| cwd | —   | —   | —   | —     | —   | —   | —   | —     | —   | —      | —   | —     | 0.44     |
| pet | —   | —   | —   | —     | —   | —   | —   | —     | —   | —      | —   | —     | 0.52     |
| psand| —   | —   | —   | —     | —   | —   | —   | —     | —   | —      | —   | —     | 0.51     |
| awc | 0.14| —   | —   | —     | —   | —   | —   | —     | —   | —      | —   | —     | 0.30     |
| por | 0.33| —   | —   | —     | —   | —   | —   | —     | —   | —      | —   | —     | 0.37     |
| thk | 0.27| 0.16| 0.29| —     | 0.24| —   | —   | —     | —   | —      | —   | —     | 0.35     |
| pplay| 0.27| —   | —   | 0.22  | 0.30| 0.01| —   | —     | —   | —      | —   | —     | 0.02     |
| run | 0.40| —   | —   | —     | —   | —   | —   | —     | —   | —      | —   | —     | 0.06     |
| aprlpc| 0.00| 0.00| 0.00| —     | 0.00| 0.00| —   | —     | —   | —      | —   | —     | 0.05     |
| ppt | 0.35| 0.35| 0.35| 0.35  | 0.35| 0.35| 0.35| —     | —   | —      | —   | —     | 0.35     |
| NDVI| 0.01| 0.01| 0.01| —     | —   | —   | —   | —     | —   | —      | —   | —     | 0.01     |
| snowstd| 0.01| 0.01| 0.01| —     | —   | —   | —   | —     | —   | —      | —   | —     | 0.01     |

Note: Variables exceeding a correlation threshold of ±0.50 are highlighted in bold. A correlation coefficient of +1.0 indicates a perfect positive correlation, −1.0 indicates a perfect inverse correlation, and zero suggests that there is no correlation. Variables are described in Table 1.
optimal number of classes is usually unknown, we started with a conservatively high number \((N = 30)\), analyzed the clusters, reduced the number of classes, and repeated the analyses until we achieved an optimal number of classes. Necessarily, the optimization of the number of classes will be constrained by project resources, objectives, or other limitations. We used the DENDROGRAM and MAXIMUM LIKELIHOOD CLASSIFICATION functions to interpret the cluster results and reduced the number of classes stepwise until we determined an optimal number of classes that captured variability across all the study regions.

In our worked example, a primary objective was to compare the classifications across the three nested spatial scales. For this reason, after defining an optimal number of classes, we specified the same number of classes \((N = 11)\) for all the study regions in the final ISOCLUSTER analyses to support comparison across spatial scales. However, the workflow allows a user to define the number of clusters on the basis of resources for implementation, ability to capture site heterogeneity, or other monitoring constraints or objectives.

The DENDROGRAM function (http://pro.arcgis.com/en/pro-app/tool-reference/spatial-analyst/dendrogram.htm) constructs a table and tree diagram that show the classification hierarchy, which were used to explore the potential for merging classes. The MAXIMUM LIKELIHOOD CLASSIFICATION function (http://pro.arcgis.com/en/pro-app/tool-reference/spatial-analyst/maximum-likelihood-classification.htm) calculates a percent certainty or probability for each raster cell, which represents the highest probability or “maximum likelihood” that the raster cell belongs to the assigned class. The probability values range from 0 to 1.0. Lower values indicate a higher likelihood of misclassification, whereas higher values indicate lower likelihood of misclassification. During the optimization, the confidence rasters were used to explore the potential for reducing the probability of misclassification.

**Landform Classification.** After the SMRUs are classified, landform classification is the final step in the workflow to determine the location for installation of monitoring equipment on the basis of local topographic conditions. Fine-scale topographic analysis was used to develop a landform map and explore potential locations for soil moisture monitoring in the Hetch-Hetchy watershed, an important water supply basin. This final step in the workflow is necessary to determine optimal sensor locations and to ensure that a wide range of landscape positions are represented when designing a monitoring network.

A landform map was derived in ArcGIS using a multi-scale topographic position index (TPI) and a slope raster (Weiss 2001; Jenness 2006; Deumlich et al. 2010) derived from a 10-m digital elevation model (DEM) (https://nationalmap.gov/elevation.html). The multi-scale TPI compares the elevation of individual focal cells to the mean elevation of the surrounding cells, where the surrounding cells are referred to as the “neighborhood” and the size and shape of the neighborhood are user-defined. We calculated a fine-scale TPI (circular neighborhood with a radius of 750 m) and large-scale TPI (circular neighborhood with a radius of 2,000 m) using the ArcGIS FOCAL MEAN function (https://pro.arcgis.com/en/pro-app/tool-reference/spatial-analyst/focal-statistics.htm) and a raster calculation:

\[
\text{TPI} = (\text{cell elevation} - \text{neighborhood focal mean}),
\]

where positive TPI values are assigned to cells with elevations higher than surrounding areas and negative TPI values are assigned to cells with elevations lower than surrounding areas. A slope raster was then calculated using the ArcGIS SLOPE function (http://pro.arcgis.com/en/pro-app/tool-reference/spatial-analyst/slope.htm). The final landform map was created using a nested conditional statement that applies classification thresholds to the 750-m TPI, 2,000-m TPI, and slope rasters (Weiss 2001; Jenness 2006). We used thresholds recommended by Deumlich et al. (2010) to define 10 landform classes (Table 5). The multi-scale TPI enables delineation of small depressions, large valleys, small hummocks, and large ridges. The resulting landform map provides a useful base for determining sensor location and facilitates consideration of landscape position in the design of soil moisture monitoring networks.

**Validation of SMRUs.** We used direct measurements of soil moisture as an independent dataset to validate the classification of SMRUs at the watershed-scale on the basis of summary statistics. We did not define relevant summary statistics a priori and the number of validation stations was constrained by available data. Eleven validation stations, located within the Sierra Nevada ecoregion (Figure 1), were selected on the basis of their geographical distribution, a continuous period of record spanning wet to dry conditions from 2009 to 2012, no more than 10% missing data, and multiple soil probe depths that span at least 50 cm to provide an integrated estimate of soil water storage (Table 6). A transect of six validation stations was selected within the Merced-Tuolumne watersheds along an elevation gradient. These
six stations span a range of climatic conditions: rain-dominated (Forty Mile, Merced Grove), rain-to-snow transition (Gin Flat, Smoky Jack), and snow-dominated (Olmsted Quarry, Dana Meadows).

In a separate analysis, we applied the SMRU workflow in the Upper Carson River watershed, which is located just above the rain-to-snow transition, and then validated using a cluster of five validation stations (Poison Flat, Monitor Pass, Forestdale Creek, Blue Lakes, Burnside Lake). The cluster of Carson River validation stations were selected to show an example study region with low climate variability among stations located at similar elevations across a snow-dominated watershed.

Datasets were downloaded from the Scripps Institution of Oceanography (Scripps and University of California, San Diego; http://meteora.ucsd.edu/), the North American Soil Moisture Database (Quiring et al. 2015), the Department of Water Resources California Data Exchange Center (http://cdec.water.ca.gov/), the Natural Resources Conservation Service (NRCS) Snowpack Telemetry (SNOTEL) (http://www.wcc.nrcs.usda.gov/snow/), and USGS datasets (Stern et al. 2018). These continuous soil moisture time series were checked for missing data. Cumulative soil moisture, which is a measure of soil water storage, was calculated on a daily basis using the equation:

\[
\text{Cumulative Soil Moisture} = \int_{t_0}^{t} \text{Soil Moisture} \, dt
\]

### TABLE 5. Definition of landform classes on the basis of a multi-scale TPI and slope rasters (Weiss 2001; Deumlich et al. 2010), derived from a 10-m digital elevation model.

| Landform classification number | Landform classification                                      | Neighborhood TPI             |
|-------------------------------|-------------------------------------------------------------|------------------------------|
| 1                             | Canyons, deeply incised streams                             | TPI ≤ -1                     |
| 2                             | Mid-slope drainages, shallow valleys                       | TPI ≤ -1                     |
| 3                             | Upland drainages, headwaters                               | TPI ≤ -1                     |
| 4                             | U-shaped valleys                                            | -1 < TPI < 1                 |
| 5                             | Plains                                                      | -1 < TPI < 1                 |
| 6                             | Open slopes                                                 | -1 < TPI < 1                 |
| 7                             | Upper slopes, mesas                                        | -1 < TPI < 1                 |
| 8                             | Local ridges, hills in valleys                              | TPI ≥ -1                     |
| 9                             | Mid-slope ridges, small hills in plains                     | TPI ≥ -1                     |
| 10                            | Mountain tops, high ridges                                 | TPI ≥ -1                     |

### TABLE 6. Soil moisture station description including name, latitude and longitude, elevation, period of record, depths, and source.

| Station name         | Latitude/longitude | Elevation (m) | Period of record | Depths (cm) | Source                                                      |
|----------------------|-------------------|---------------|------------------|-------------|-------------------------------------------------------------|
| Forty Mile           | 37.77/–119.86     | 1,720         | 2008–present     | 30, 60, 90  | Scripps Institution of Oceanography — Climate Research Division (2008) |
| Merced Grove         | 37.76/–119.84     | 1,810         | 2008–present     | 30, 60, 90  | Scripps Institution of Oceanography — Climate Research Division (2008) |
| Gin Flat             | 37.77/–119.77     | 2,149         | 2005–present     | 10, 36, 71  | Stern et al. (2018)                                         |
| Smoky Jack           | 37.82/–119.71     | 2,180         | 2008–present     | 30, 60, 90  | Scripps Institution of Oceanography — Climate Research Division (2008) |
| Poison Flat          | 38.51/–119.63     | 2,358         | 2003–present     | 5, 20, 51   | NRCS SNOTEL                                                 |
| Forestdale Creek     | 38.68/–119.96     | 2,444         | 2003–present     | 5, 20, 51   | NRCS SNOTEL                                                 |
| Blue Lakes           | 38.61/–119.92     | 2,456         | 2005–present     | 5, 20, 51   | NRCS SNOTEL                                                 |
| Burnside Lake        | 38.72/–119.89     | 2,477         | 2003–present     | 5, 20, 51   | NRCS SNOTEL                                                 |
| Monitor Pass         | 38.67/–119.61     | 2,533         | 2005–present     | 5, 20, 51   | NRCS SNOTEL                                                 |
| Olmsted Quarry       | 37.81/–119.51     | 2,601         | 2008–present     | 30, 60, 90  | Scripps Institution of Oceanography — Climate Research Division (2008) |
| Dana Meadows         | 37.90/–119.26     | 2,965         | 2005–present     | 10, 36, 53, 79 | Stern et al. (2018)                                        |

Notes: NRCS, Natural Resources Conservation Service; SNOTEL, Snowpack Telemetry.
where $SM$ is the cumulative soil moisture ($\%$), depth is summed across all soil layers in cm, profile is the total soil profile thickness in cm, and $X$ is the soil moisture value for each soil layer ($\%$). We selected 2010, an average water year (Jian et al. 2011), for validation using statistical measures of central tendency, dispersion, and extremes.

RESULTS

The explanatory power of the PCs differed across the range of spatial scales represented by the study regions (Table 7). A break in slope scree-test criterion (Figure 3) retained eight PCs for cluster analysis at the California-scale and seven PCs for the Sierra Nevada ecotone-scale and the Merced-Tuolumne watershed-scale. The first PC for the California-scale analysis explained 35% of the variance compared to 42% and 52% for the Sierra Nevada and Merced-Tuolumne regions, respectively. The PCs used for SMRU classification at the state (Figure 4), ecotone (Figure 5), and watershed (Figure 6) scales explained 96%, 94%, and 97% of the variance, respectively.

At the California-scale, the SMRU classification (Figure 4a) is strongly influenced by elevation. At this scale, the Merced-Tuolumne watersheds are dominated by six SMRUs (Figure 4b), with the highest likelihood of misclassification occurring near the rain-to-snow transition (Figure 4c), which coincides with the transition from SMRU #3 to #10. The lowest elevations, which are rain-dominated, are primarily classified as SMRU #5 and #7 and the highest elevations that are snow-dominated are primarily classified as SMRUs #10 and #11.

TABLE 7. Principal component analysis.

| PC   | Eigenvalue | Percent of eigenvalues (%) | Cumulative percent of eigenvalues (%) | PC   | Eigenvalue | Percent of eigenvalues (%) | Cumulative percent of eigenvalues (%) | PC   | Eigenvalue | Percent of eigenvalues (%) | Cumulative percent of eigenvalues (%) |
|------|------------|----------------------------|---------------------------------------|------|------------|----------------------------|---------------------------------------|------|------------|----------------------------|---------------------------------------|
| pc1  | 1.26       | 35                         | 35                                    | pc1  | 0.29       | 42                         | 42                                    | pc1  | 3.80       | 52                         | 52                                    |
| pc2  | 0.61       | 17                         | 52                                    | pc2  | 0.11       | 15                         | 57                                    | pc2  | 1.38       | 19                         | 71                                    |
| pc3  | 0.51       | 14                         | 66                                    | pc3  | 0.09       | 9                          | 70                                    | pc3  | 0.64       | 9                          | 80                                    |
| pc4  | 0.34       | 9                          | 76                                    | pc4  | 0.06       | 9                          | 78                                    | pc4  | 0.50       | 7                          | 87                                    |
| pc5  | 0.22       | 6                          | 82                                    | pc5  | 0.05       | 7                          | 85                                    | pc5  | 0.37       | 5                          | 92                                    |
| pc6  | 0.19       | 5                          | 87                                    | pc6  | 0.04       | 5                          | 90                                    | pc6  | 0.23       | 3                          | 95                                    |
| pc7  | 0.18       | 5                          | 92                                    | pc7  | 0.03       | 4                          | 94                                    | pc7  | 0.16       | 2                          | 97                                    |
| pc8  | 0.15       | 4                          | 96                                    | pc8  | 0.02       | 3                          | 97                                    | pc8  | 0.08       | 1                          | 99                                    |
| pc9  | 0.08       | 2                          | 98                                    | pc9  | 0.01       | 2                          | 99                                    | pc9  | 0.04       | 1                          | 99                                    |
| pc10 | 0.03       | 1                          | 99                                    | pc10 | 0.01       | 1                          | 100                                   | pc10 | 0.02       | 0                          | 100                                   |
| pc11 | 0.01       | 0                          | 100                                   | pc11 | 0.00       | 0                          | 100                                   | pc11 | 0.02       | 0                          | 100                                   |
| pc12 | 0.01       | 0                          | 100                                   | pc12 | 0.00       | 0                          | 100                                   | pc12 | 0.01       | 0                          | 100                                   |
| pc13 | 0.02       | 0                          | 100                                   | pc13 | 0.00       | 0                          | 100                                   | pc13 | 0.00       | 0                          | 100                                   |

Note: Variables highlighted in bold were used for cluster analysis and classification of SMRUs. Variable names are described in Table 1.
At the ecoregion-scale (Figure 5a), topography exerts a larger influence on the SMRU classification and the Merced-Tuolumne SMRUs (Figure 5b) illustrate greater variability across the study region. Previous climate modeling at the ecoregion-scale indicates ridge and valley topography

FIGURE 4. CA-scale analysis showing validation stations and SMRUs for (a) the state and (b) close-up of MT, as well as (c) the maximum likelihood classification (MLC) for the MT; red colors indicate a higher likelihood of misclassification and green colors indicate a lower likelihood of misclassification.

FIGURE 5. Ecoregion-scale analysis showing validation stations and the SMRUs for (a) SN and (b) close-up of the MT, as well as (c) the MLC for the MT; red colors indicate a higher likelihood of misclassification and green colors indicate a lower likelihood of misclassification.
strongly influence the variability of snow processes (Curtis et al. 2014) and this assertion is supported by the SMRU classification. At this intermediate-scale, the Merced-Tuolumne watersheds are dominated by three SMRUs (#3, #4, and #7) below the rain-to-snow transition that remain strongly influenced by the elevation gradient. Above the rain-to-snow transition, the spatial extent of four SMRUs (#7, #8, #10, and #11) is more variable and shows the influence of ridge and valley topography. At this spatial scale, the highest likelihood of misclassification remains near the rain-to-snow transition (Figure 5c).

At the watershed-scale (Figure 6a), the Merced-Tuolumne SMRUs below the rain-to-snow transition remain strongly influenced by the elevation gradient and are dominated by four SMRUs (#6, #7, #9, and #10) but the spatial extent of the five SMRUs above the rain-to-snow transition (#1, #2, #3, #4, and #5) is highly variable and influenced by finer-scale topography and associated hydrologic processes. At this scale, the highest likelihood of misclassification

---

**TABLE 8.** MLC, the percent likelihood of cells correctly assigned to a class, for CA, SN, MT, and Upper CR describing the number of cells (# of cells) in each confidence class, percent of each study region within each confidence class (conf.), and the confidence cumulative percent (cum %).

| MLC (%) | CA (# cells) | CA conf. % | CA cum % | SN (# cells) | SN conf. % | SN cum % | MT (# cells) | MT conf. % | MT cum % | CR (# cells) | CR conf. % | CR cum % |
|---------|--------------|------------|----------|--------------|------------|----------|--------------|------------|----------|--------------|------------|----------|
| 100     | 72,898       | 2          | 2        | 6,459        | 1          | 1        | 542          | 1          | 1        | 28           | 0          | 0        |
| 99.5    | 85,368       | 2          | 3        | 7,331        | 1          | 2        | 803          | 1          | 2        | 51           | 0          | 1        |
| 97.5    | 218,156      | 5          | 7        | 20,887       | 3          | 5        | 2,964        | 4          | 5        | 105          | 1          | 1        |
| 95      | 278,614      | 6          | 13       | 33,321       | 4          | 9        | 4,102        | 5          | 10       | 344          | 2          | 4        |
| 90      | 401,207      | 9          | 24       | 65,489       | 9          | 17       | 6,647        | 8          | 18       | 587          | 4          | 7        |
| 75      | 812,543      | 19         | 43       | 150,353      | 20         | 37       | 16,337       | 19         | 37       | 1,452        | 10         | 17       |
| 50      | **922,128**  | **21**     | **67**   | **190,311**  | **25**     | **62**   | **18,330**   | **22**     | **59**   | **2,428**    | **16**     | **34**   |
| 25      | 724,430      | 17         | 84       | 150,162      | 20         | 81       | 13,743       | 16         | 75       | 2,520        | 17         | 50       |
| 10      | 432,995      | 10         | 93       | 79,834       | 10         | 91       | 8,685        | 10         | 86       | 1,300        | 9          | 59       |
| 5       | 164,713      | 4          | 96       | 25,908       | 3          | 95       | 3,918        | 5          | 90       | 525          | 4          | 63       |
| 2.5     | 92,247       | 2          | 98       | 12,882       | 2          | 97       | 2,037        | 2          | 93       | 354          | 2          | 65       |
| 1       | 60,845       | 1          | 99       | 9,821        | 1          | 98       | 1,717        | 2          | 95       | 390          | 3          | 68       |
| 0.5     | 22,685       | 1          | 98       | 4,077        | 1          | 98       | 770          | 1          | 96       | 199          | 1          | 59       |
| 0       | 61,634       | 1          | 100      | 12,735       | 2          | 100      | 3,589        | 4          | 100      | 4,603        | 31         | 100      |

Note: The percent of each study region with a 50% or greater likelihood of cells assigned to the correct class is highlighted in bold.
moved away from the rain-to-snow transition to topographic highlands and lower elevation ridges (Figure 6b).

A comparison of the maximum likelihood classification rasters across spatial scales illustrates a lower confidence and higher likelihood for misclassification at finer-scales (Table 8). The percent of each study region with a 50% or greater likelihood of cells assigned to the correct class is 67%, 62%, and 59%, respectively, for the state, ecoregion, and watershedscales. The decrease in confidence and associated increase in the probability of misclassification at finer-scales is further exemplified in the SMRU classification for the Upper Carson River watershed.

To expand the number of validation sites, we applied the SMRU workflow to the Upper Carson River watershed, which is a snow-dominated basin located along the Sierra Nevada crest (Figure 1). We selected five PCs for the cluster analysis and defined the optimal number of classes (N = 11) for classification but the ISOCCLUSTER function only classified four SMRUs (Figure 7a). The maximum likelihood of misclassification (Figure 7b) was much higher in comparison to the Merced-Tuolumne watersheds and the percent of the Carson River watershed with a 50% or greater likelihood of cells assigned to the correct class was only 34% (Table 8).

The watershed-scale SMRUs were validated using measured data from six stations along an elevation gradient within the Merced-Tuolumne watersheds (Figure 8a) and a cluster of five stations within the Upper Carson River watershed (Figure 8b). The transect of stations across the Merced-Tuolumne watersheds illustrates the range of soil moisture responses along an elevation gradient that spans the rain-to-snow transition. In comparison, the cluster of stations in the Upper Carson River watershed illustrates a range of soil moisture responses within a narrow elevation range across a snow-dominated watershed. We used the 2010 water year for validation because it represented an average year but we also show a series of wet to dry years, for the period spanning 2009–2012.

We used summary statistics for water year 2010 to describe the shape of the annual soil moisture time series and to further illustrate that the SMRUs capture a range of soil moisture responses to climate across the study regions. The summary statistics demonstrate that SMRU classes are poorly explained by measures of central tendency and better explained by statistical metrics of dispersion and extremes (Figure 9). In the Merced-Tuolumne and Carson River watersheds, measures of dispersion (standard deviation, coefficient of variation, range, variance, and the interquartile range) and extremes (maximum, minimum, skewness, and kurtosis) vary systematically and highlight the primary differences among the SMRUs.

The response of soil moisture to climatic variations is further illustrated by comparing summary statistics for three stations (Figure 9) that span the rain-to-snow transition within the Merced-Tuolumne watersheds. Merced Grove (MT10) is rainfall-dominated, with a low variability in soil moisture over an annual scale but a rapid response to rainfall. This station had a low standard deviation, coefficient of
FIGURE 8. (a) Soil moisture time series data for transect of stations (Table 7) located in the MT combined watershed (Figure 1). (b) Soil moisture time series data for cluster of stations (Table 7) located in the Upper CR watershed (Figure 1). Histograms and boxplots display only data from water year 2010, whereas exceedance curves and annual time series are shown for water years 2009 (black), 2010 (blue), 2011 (green), and 2012 (red).
variation, and skewness but the highest minimum value. The lack of snowpack at Merced Grove results in a narrower range of soil water content because the soil properties enable the soil to continually drain. Gin Flat (MT6) is located within the rain-to-snow transition, where the soil responds to early season rainfall, but soil moisture rapidly declines and stabilizes during the winter as snowpack accumulates. There is a second gradual rise followed by a rapid decline in soil moisture during the spring snowmelt runoff season. Although Gin Flat had mean and median values that were similar to Merced Grove, values for the standard deviation, coefficient of variation, range, variance, maximum, skewness, and minimum were higher.

FIGURE 9. Histogram of statistical metrics derived using measured soil moisture time series during water year 2010, which describe the central tendency, dispersion, and extremes for each validation station. See Figure 1 for station locations. SMRUs for the Merced-Toulumne watershed are labeled MT2, MT6, MT10, and MT5 and SMRUs for Carson watershed are labeled CR2 and CR3. CV, Coefficient of Variation; IQR, Interquartile Range.
FIGURE 10. (a) SMRUs for the Hetch-Hetchy watershed showing percent of SMRU coverage and (b) the MLC; red colors indicate a higher likelihood of misclassification and green colors indicate a lower likelihood of misclassification.

FIGURE 11. Landform classification map for (a) the Hetch-Hetchy watershed and (b) a close-up of the Dana Meadows soil moisture station (Table 6). Note that there are only nine landform classes; plains (Class 5, Table 5) do not exist in this high-relief mountain watershed.
kurtosis were larger and the minimum and interquartile range were smaller. Dana Meadows (MT2) is snow-dominated with a small response to early season precipitation and a strong snowmelt signal from May to August. Colder air temperatures maintain snowpack at Dana Meadows for a longer period resulting in stable soil moisture contents over the winter and throughout much of the summer. Dana Meadows had the lowest mean, median, and minimum values but highest standard deviation, coefficient of variation, range, variance, interquartile range, maximum, skewness, and kurtosis. Dana Meadows also had the largest range of soil moisture conditions, but rapidly loses soil moisture, indicating better soil drainage.

The final step in the proposed workflow is a landform classification that can be used to design and optimize the number and location of soil moisture monitoring stations. We show an example application for the Hetch-Hetchy watershed (1,175 km², Figure 10), an important water supply basin in the Sierra Nevada and located upstream of the Hetch-Hetchy reservoir within the MT watershed boundary (Figure 1). A comprehensive soil moisture network would have stations located in several combinations of landforms and SMRUs but costs and accessibility for operation and maintenance of monitoring equipment must be considered. A more realistic application is to select a transect across a gradient of SMRUs and landforms similar to the Merced-Tuolumne validation transect. The recommended workflow uses a landform classification (Figure 11) along with the maximum likelihood classification (Figure 10) to select soil moisture monitoring locations at the local-scale.

DISCUSSION

The primary objective of this study was to design and validate a statewide soil moisture monitoring strategy that can be applied over a broad range of spatial scales. A workflow within a geographic information system was developed to classify SMRUs using a deductive or “top down” approach. Our workflow combines several user-friendly ArcGIS statistical tools and is intended to be easily repeatable and our nested-scale analysis illustrates how spatial scale influences the SMRU classification. At the California-scale, SMRUs were strongly influenced by elevation. At the ecoregion-scale, the SMRUs captured more heterogeneity above the rain-to-snow transition. At the watershed-scale, the SMRUs capture the influence of finer-scale topography and the integrated effects of soil, vegetation, climate, and hydrology.

Validation was performed using available soil moisture time series data and our results indicate the SMRUs capture soil moisture variability across the study regions. In heavily instrumented watersheds, where the relationship between soil water storage and runoff is well understood, a supervised classification could be developed with classes defined on the basis of user criteria.

Maps of the maximum likelihood classification can be used to explore the potential for reducing statistically misclassification and to optimize site selection and sensor location. This is particularly relevant at the watershed-scale because the likelihood of misclassification increased at finer-spatial scales.

The final step in our workflow utilizes a landform classification to ensure that a wide range of topographic conditions are represented within a soil moisture monitoring strategy. Site selection and sensor location at the local-scale will benefit from the finescale landform classification that facilitates selection of a range of hillslope positions (e.g. soil catena) for soil moisture monitoring.

Our proposed workflow will be used to develop a statewide soil moisture monitoring strategy to inform water resource management throughout California. The relation between soil moisture conditions and spring snowmelt is of particular interest. An optimized soil moisture monitoring network can help inform reservoir operations managers by providing an assessment of soil water storage that might be filled with snowmelt, thus reducing snowmelt discharge available for reservoir storage. Other potential applications include: quantifying landscape stress and wildfire risk assessment (Krueger et al. 2016), estimating irrigation demand and crop productivity (Rivers et al. 2015), and risk assessment and early warning of floods (Ionita et al. 2015) and landslides (Mirus et al. 2016).

Many soil moisture sensors are installed at shallow depths, which are intended to calibrate satellite data but these shallow sensors fail to characterize the complete soil profile. Although imperative for determining soil response to climate, there are relatively few stations with sensors installed below the top 5–10 cm of soil. We recommend a standardized approach for monitoring that includes installation of deeper subsurface sensors. Ideally, sensors would be installed in at least three different soil horizons that include the rooting zone and the soil and bedrock interface (Flint et al. 2008).

CONCLUSIONS

Hydrologic conditions under future climates are expected to be more variable and more extreme,
which will make quantifying the role of soil moisture within the water balance increasingly important for runoff forecasting and water resource management. Soil moisture monitoring is essential to understanding soil moisture response to variable climate conditions and to forecast extreme or anomalous conditions related to spring snowmelt discharge and summer baseflows. Our workflow for designing and validating statewide soil moisture monitoring strategies will be an integral component of water resource management. Our example applications were able to capture a range of soil moisture responses to climate across regional to watershed-scales and can be used to optimize the number and location of soil moisture stations necessary to improve hydrologic characterization. The workflow that has been developed here is unique and can be implemented over a range of spatial scales using existing and publicly available input datasets and software, or it can be customized by the user for their specific needs.

A better understanding of the linkages between soil water storage and runoff, particularly during snowmelt discharge and summer baseflow periods, could greatly advance the understanding of soil-runoff processes, could improve runoff forecasts, and inform the timing of surface water reservoir operations. Improvements in data collection platforms are needed, particularly real-time data collection over a range of soil depths within the soil profile and the network of soil moisture monitoring locations should be expanded. Implementing a real-time soil moisture monitoring strategy for California could usher in a new paradigm of water supply management that incorporates a broader spectrum of measured data to monitor the hydrologic cycle as completely as possible.

LITERATURE CITED

Abdi, H., and L.J. Williams. 2010. “Principal Component Analysis.” WIREs Computational Statistics 2: 433–59. https://doi.org/10.1002/wics.101.

Auerbach, D.A., B.P. Buchanan, A.V. Alexiades, E.P. Anderson, A.C. Encalada, E.L. Larsson, R.A. McManamay et al. 2016. “Towards Catchment Classification in Data-Scarce Regions.” Ecohydrology 9: 1235–47. https://doi.org/10.1002/eco.1721.

Boorman, D.B., J.M. Hollis, and A. Lilly. 1995. “Hydrology of Soil Types: A Hydrologically-Based Classification of the Soils of the United Kingdom.” Institute of Hydrology, Report No. 126, 146 pp. http://nora.nerc.ac.uk/id/eprint/7369.

Cattell, R.B. 1966. “The Scree Test for the Number of Factors.” Multivariate Behavioral Research 1 (2): 245–76. https://doi.org/10.1207/s15327906mbr0102_10.

Crow, W.T., and D. Ryu. 2009. “A New Data Assimilation Approach for Improving Runoff Prediction Using Remotely-Sensed Soil Moisture Retrievals.” Hydrology and Earth System Sciences 13: 1–16. https://doi.org/10.5194/hess-13-1-2009.

Curtis, J.A., L.E. Flint, A.L. Flint, J.D. Landquist, B. HUDGENS, E.E. BOYDSTON, and J.K. Young. 2014. “Incorporating Cold-Air Pooling Into Downscaled Climate Models Increases Potential Refugia for Snow-Dependent Species within the Sierra Nevada Ecoregion,” CA.” PLoS ONE 9 (9): e106984. https://doi.org/10.1371/journal.pone.0106984.

Department of Forestry and Fire Protection, Fire and Resource Assessment Program (FRAP). 2015. http://frap.fire.ca.gov/data/frapidata-subset.

Dettinger, M.D., F.M. Ralph, T. Das, P. Neiman, and D. Cayan. 2011. “Atmospheric Rivers, Floods, and the Water Resources of California.” Water 3: 455–78. https://doi.org/10.3390/w3020445.

Deumlich, D., R. Schmidt, and M. Sommer. 2010. “A Multiscale Soil–Landform Relationship in the Glacial-Drift Area Based on Digital Terrain Analysis and Soil Attributes.” Journal of Plant Nutrition and Soil Science 173 (6): 843–51. https://doi.org/10.1002/jpln.200900094.

Dobriyal, P., A. Qureshi, R. Badola, and S.A. Hussain. 2012. “A Review of the Methods Available for Estimating Soil Moisture and Its Implications for Water Resource Management.” Journal of Hydrology 458–459: 110–17. https://doi.org/10.1016/j.jhydrol.2012.06.021.

Entekhabi, D., E.G. Njoku, P.E. O’Neill, K.H. Kellogg, W.T. Crow, W.N. Edelstein, J.K. Entin et al. 2010. “The Soil Moisture Active Passive (SMAP) Mission.” Proceedings of the IEEE 98 (5): 704–16. https://doi.org/10.1109/JPROC.2010.2043918.

Flint, A.L., L.E. Flint, and M.D. Dettinger. 2008. “Modeling Soil Moisture Processes and Recharge Under a Melting Snowpack.” Vadose Zone Journal 7 (1): 350–57. https://doi.org/10.2136/vzj2006.0135.

Flint, L.E., and A.L. Flint. 2014. “California Basin Characterization Model: A Dataset of Historical and Future Hydrologic Response to Climate Change.” U.S. Geological Survey Data Release. https://doi.org/10.5066/F76T0JPB.

Harpold, A.A., K. Sutcliffe, J. Clayton, A. Goodbody, and S. Vazquez. 2017. “Does Including Soil Moisture Observations Improve Operational Streamflow Forecasts in Snow-Dominated Watersheds?” Journal of the American Water Resources Association 53 (1): 179–96. https://doi.org/10.1111/1752-1688.12490.

Henson, R.K., and J.K. Roberts. 2006. Use of Exploratory Factor Analysis in Published Research: Common Errors and Some Comment on Improved Practice. Educational and Psychological Measurement 66 (3): 393–416. https://doi.org/10.1177/0016495205282485.

Hickman, J.C. 1993. The Jepson Manual: Higher Plants of California. Berkeley: University of California Press.

Illston, B.G., J.B. Basara, D.K. Fisher, R. Elliott, C.A. Fiebrich, K.C. Crawford, K. Humes, and E. Hunt. 2008. “Mesoscale Monitoring of Soil Moisture Across a Statewide Network.” Journal of Atmospheric and Oceanic Technology 25: 167–82. https://doi.org/10.1175/2007JTECHA1993.1.

Ionita, M., M. Dima, G. Lohmann, P. Scholz, and N. Rimbu. 2015. “Active Passive (SMAP) Mission.” Proceedings of the IEEE 57. https://doi.org/10.1109/ICIP.2015.7351378.

IPCC (International Panel on Climate Change). 2014. “IPCC Fifth Assessment Report: Climate Change 2014; Synthesis Report.” Core Writing Team, edited by R.K. Pachauri and L. Meyer. Geneva, Switzerland: International Panel on Climate Change. https://www.ipcc.ch/report/ar5/

Jain, A.K. 2010. “Data Clustering: 50 Years Beyond K-Means.” Pattern Recognition Letters 31 (8): 651–66. https://doi.org/10.1016/j.patrec.2009.09.011.

Jenness, J. 2006. “Topographic Position Index (tpi_jen.avx) Extension for ArcView 3.x, v. 1.3a.” Jenness Enterprises. http://www.jennessent.com/arcview/tpi.htm.
