

## **From pattern to process and process to pattern: insights on data-driven Critical Zone research from the Big Data collaborative network cluster**

JULIA N PERDRIAL<sup>1</sup>, KRISTEN UNDERWOOD<sup>1</sup>,  
SHAURYA SWAMI<sup>1</sup>, BYUNG SUK LEE<sup>1</sup>, IJAZ UL HAQ<sup>1</sup>,  
DUSTIN KINCAID<sup>1</sup>, BRYN STEWART<sup>2</sup>, ERIN SEYBOLD<sup>3</sup>,  
DONNA RIZZO<sup>1</sup>, LI LI<sup>2</sup>, SCOTT D HAMSHAW<sup>1</sup> AND  
JAMES B SHANLEY<sup>4</sup>

<sup>1</sup>University of Vermont

<sup>2</sup>Penn State University

<sup>3</sup>University of Kansas

<sup>4</sup>USGS

Presenting Author: [julia.perdrial@uvm.edu](mailto:julia.perdrial@uvm.edu)

Environmental data, including high resolution sensor-generated data, are abundant and can be used to understand Earth surface processes in order to respond to challenges of the Anthropocene. However, harnessing the potential of such big data—data that might be of high volume, veracity, variety, velocity and/or value—to understand Critical Zone (CZ) processes requires addressing challenges inherent to these data. The challenges include data cleaning and harmonization and applying appropriate analytical tools to gain insight from big data. Addressing these challenges is a core objective of our CZ collaborative network project, “Using Big Data to assess ecohydrological resilience across scales”. In this presentation, we will share our experiences with addressing some of these big data challenges. We will show examples such as data cleaning, where we apply machine-learning approaches to automate identification and correction of faulty data in sensor time series, freeing the domain expert to interpret the ecohydrological drivers. We will also demonstrate how we use an iterative “pattern-and-process” approach where we apply complex systems tools to extract patterns from big data and process-based models and observations to understand site-specific processes. Through this iterative process, which we are applying across scales, we will identify system characteristics that promote ecological resilience, and those that render systems vulnerable to disturbance.