EGU2020-1313
https://doi.org/10.5194/egusphere-egu2020-1313
EGU General Assembly 2020

A nonlinear dynamics approach to data-enabled science: Reconstructing soil-moisture dynamics from big data collected by wireless sensor networks

Ray Huffaker and Rafael Munoz-Carpena
Department of Agricultural Engineering, University of Florida, Gainesville, FL, USA (Rafael.Munoz@unavarra.es)

The complex soil biome is a center piece in providing essential ecosystem services that humans rely on (carbon sequestration, food security, one-health interactions). Agricultural engineers and soil scientists are developing wireless sensor networks (WSN) that collect large/big data on the soil key state variables (water content, temperature, chemistry) to better understand the soil biome primary environmental drivers. The profession extracts information from WSN records with methods including soil-process modeling and artificial-intelligence (AI) algorithms. However, these approaches carry their own limitations. A recent review article faulted current soil-process modeling for inadequately detecting and resolving model structural (abstraction) errors. AI experts themselves caution against indiscriminant use of AI methods because of: a) problems including replication of past results due to inconsistent experimental methods; b) difficulty in explaining how a particular method arrives at its conclusions (the black box problem) and thus in correcting algorithms that learn 'bad lessons'; and c) lack of rigorous criteria for selecting AI architectures. An alternative approach to address these limitations is to investigate new strategies for reducing large/big data problems into smaller, more interpretable causal abstractions of the soil system.

We develop an innovative data diagnostics framework—based on empirical nonlinear dynamics techniques from physics—that addresses the above concerns over soil-process modeling and AI algorithms. We diagnose whether WSN and other similar environmental large/big data are likely generated by dimension-reducing (i.e., dissipative) nonlinear dynamics. An n-dimensional nonlinear dynamic system is dissipative if long-term dynamics are bounded within m<<n dimensions, so that the problem of modeling long-term dynamics shrinks by the n-m inactive degrees of freedom. If so, long-term system dynamics can be investigated with relatively few degrees of freedom that capture the complexity of the overall system generating observed data. To make this diagnosis, we first apply signal processing to isolate structured variation (signal) from unstructured variation (noise) in large/big data time series records, and test signals for nonlinear stationarity. We resolve the structure of isolated signals by distinguishing between stochastic-forcing and deterministic nonlinear dynamics; reconstruct phase space dynamics most likely generating signals, and test the statistical significance of reconstructed dynamics with surrogate data. If the reconstructed phase space is dimension-reducing, we can formulate low-dimensional

(phenomenological) ODE models to investigate nonlinear causal interactions between key soil environmental driving factors. When we do not diagnose dimension-reducing nonlinear real-world dynamics, then underlying dynamics are most likely high dimensional and the information-extraction problem cannot be shrunk without losing essential dynamic information. In this case, other high-dimensional analysis techniques like AI offer a better modeling alternative for mapping out interactions. Our framework supplies a decision-support tool for data practitioners toward the most informative and parsimonious information-extraction method—a win-win result.

We will share preliminary results applying this empirical framework to three soil moisture sensor time series records analyzed with machine learning methods in Bean, Huffaker, and Migliaccio (2018).