



Multivariate modelling of a non-stationary process and its emergent properties using the Everglades, Florida, USA, as a case example

Ron Corstanje (1), Joanna Zawadzka (1), Delia Ivanoff (2), and Kathleen Pietro (2)

(1) Environmental Data Analytics Group Centre for Environmental and Agricultural Informatics (CEAI) Cranfield University, UK, (2) South Florida Water Management District, Florida, USA

Multivariate data analytics of large datasets are increasingly being used in the environmental studies to explore underlying factors, trends, and drivers which are not immediately obvious. Processes that influence observed temporal and spatial variability of terrestrial ecosystems often have stochastic properties and are non-linear. This has a significant effect on the resilience of a particular system, its capacity to withstand perturbations and the emerging properties which we observe as spatial or temporal patterns in a system (e.g. soil variability). The combination of remote sensing, large-scale sampling, and sensors has resulted in larger datasets becoming more prevalent in earth system studies, including soil. Here we apply these methods to the Everglades. This research focused on a large data set, collected over a period from 2002 to 2014 from the Everglades, using advanced data mining and analysis techniques including a) pattern seeking methods to test for trends and groupings in the data, b) confirmatory modelling approaches to test for predictive relationships between environmental variables, and c) spectral decomposition methods to determine the stability (or resilience) of the systems. The spatial analysis found common spatial patterns forming zonal patterns of phosphorus (P) distribution that may increasingly, over time align with the predominant flow path, micro-topography and vegetation communities. The results also suggest that coupled biogeochemical cycles may be a key component of these systems; i.e. the movement and transformation of P is coupled to that of nitrogen (N). The spectral decomposition of data was then used to determine the stability (and resilience) of these systems, illustrating how these methods can be used describe underlying data structures, and their stability over time.