

Self-organizing maps for the detection and classification of natural nanoparticles and nanoparticle systems using single-particle ICP-time-of-flight-MS

CHAD W CUSS¹, MARC F. BENEDETTI², CARLA COSTAMANGA³, LUCAS MESNARD⁴ AND MICKAEL THARAUD²

¹Memorial University of Newfoundland

²Université Paris Cité, Institut de Physique du Globe de Paris, CNRS

³Instituto de Química Física de Materiales, Ambiente y Energía (INQUIMAE)

⁴Université PSL (Paris Sciences & Lettres)

Single-particle ICP-ToF-MS has revolutionized our ability to measure the properties of nanoparticles (NPs). Data processing for these analyses largely evolved from measurements of monodisperse and mono-elemental engineered nanoparticles (ENPs) using single-particle ICP-MS; however, spICP-ToF-MS facilitates multi-elemental measurements on the *millions* of polydisperse nanoparticles typically detected per liter of natural waters. This leads to data with very different properties, requiring new approaches to answer fundamental analytical questions such as:

- What is the best way to distinguish particle pulses from the natural background of metals caused by simple complexes and small molecules in e.g. a six-minute analysis, which produces more than four million data points for each element?
- What is the best way to characterize and compare these natural nanoparticle systems (NNPS) for relating their properties to impacts in environmental roles?

These questions are at the core of emerging research in *environmental nanobiogeochemistry*, but current approaches relying on clustering or distributions struggle with such large, complex data sets typified by many dimensions, unusual distributions, sparse data (i.e. many zeroes), and frequent outliers.

Kohonen networks are unsupervised single-layer artificial neural networks with excellent visualization capabilities that are broadly used for pattern recognition and classification, more commonly known as self-organizing maps (SOM). The SOM algorithm performs an unsupervised topology-preserving projection using adaptive cluster centers, which can be formally delineated by clustering on the converged map. It can thus be considered a prototype-based clustering approach like fuzzy clustering and mixture models; however, the SOM is uniquely suited for complex data due to the iterative adaptation of cluster centers to data properties. The SOM algorithm is also nonparametric, robust to noise and outliers, and useful for sparse

and high-dimensional data with unusual distributions like nonlinear and compositional data. It therefore performs well in multiple pattern recognition and classification applications that are otherwise problematic.

After a brief introduction to environmental nanobiogeochemistry and key features of SOM, this talk will discuss advances, challenges and opportunities in applying SOM to detect and characterize the multi-elemental composition of NPs, distinguishing NNPS from different locations, and detecting multi-elemental ENPs in NNPS.

