Integrated Application of RSIMCA, t-SNE, and PVA for Chemometric Mapping of PCDD/F Congeners

Mark J. Cejas (markcejas@gmail.com) and Robert C. Barrick (Infinity Solutions Group, Seattle, WA, USA)

Background/Objectives. Unique distributions of the 17 toxic polychlorinated dibenzodioxin and dibenzofuran (PCDD/F) congeners used in risk assessments are the basis for distinguishing major sources in sediment environments. Studies to date typically have approached data assessments using either sample classifications to assess relationships among sample PCDD/F distributions or linear mixing models to distinguish and apportion endmember fingerprints in samples. Both approaches have advantages but our objective was to improve interpretations by integrating these techniques in a tiered assessment of data.

Approach/Activities. Robust Independent Modeling of Class Analogy (RSIMCA) was applied to classify over 2,800 river sediment samples into seven “clusters” or as outliers to all clusters after an initial screening of a 3,255 sample dataset. This multivariate statistical output is compressed from seven latent dimensions into two interpretable dimensions using t-Distributed Stochastic Neighbor Embedding (tSNE) graphics. Polytopic Vector Analysis (PVA) is then used to identify distinct source end-members based on PCDD/F characteristics of the classified samples.

Results/Lessons Learned. Among several advantages, the integrated chemometrics approach a) applies emerging data visualization tools in this “Big Data” era to retain the fidelity of high-dimensional data attributes of a chemical dataset spanning over two decades of sample collection; b) employs a classification technique undisturbed by compositional outliers yet tracks those for subsequent investigations; c) provides an intuitive reduced-dimensional data visualization map for the PVA mixing polytope solution; d) fills a data gap in the contextual inventory of PCDD/F source dynamics in a complex river system; and e) serves as a backdrop for further forensics investigations of the finer structure of less dominant point sources and potential upland source end-members in sediments. This tiered chemometrics strategy provides a strong weight-of-evidence approach to the interpretation of sediment data.