



Machine Learning: The Secret Weapon in Unraveling Salinity Source Mysteries at Contaminated Sites to Support Regulatory Closure

Court Sandau, Chemistry Matters Inc and Paul Fuellbrandt, Statvis Analytics Inc.

Salinity impacts, predominantly chloride contamination, represent a significant and widespread environmental challenge across Alberta. Thousands of sites, particularly those associated with oilfield operations, road maintenance, agriculture, and industrial processes, have been adversely affected by chloride contaminants such as sodium chloride [NaCl] and calcium chloride [CaCl₂]. In road maintenance, chloride salts are extensively applied for ice and dust control, while agricultural sources primarily include fertilizers containing potassium chloride [KCl]. Oilfield operations contribute chloride contamination primarily through produced brines and drilling fluids. Many chloride impacted sites have multiple sources of chloride present due to these various sources.

Chloride-contaminated soils present substantial remediation challenges due to chloride's high solubility, mobility, and persistence in soil systems. Remediation of chloride contamination typically requires long-term management strategies, often spanning decades, making accurate source identification and delineation critically important.

Traditional statistical methods such as Principal Component Analysis (PCA) have been extensively used but often fall short in clearly differentiating multiple, co-occurring chloride sources. To overcome these limitations, we have preferentially used Uniform Manifold Approximation and Projection (UMAP), a powerful machine learning technique that provides significant advantages over PCA. UMAP excels in visualizing high-dimensional data by capturing and preserving intricate relationships among multiple contaminants, enhancing the clarity of distinct contamination patterns.

UMAP effectively maps chemical compositional data into a lower-dimensional visualization, allowing environmental professionals to intuitively recognize and interpret complex source relationships. Data points are presented as clearly delineated clusters representing unique

salinity signatures; while overlapping or transitional data nodes indicate potential mixing of multiple sources. The interpretability provided by UMAP significantly simplifies the identification and differentiation of contamination sources, particularly in sites impacted by multiple salinity sources.

Our integrated methodology combines UMAP visualizations with Hierarchical Cluster Analysis (HCA), radar plots, and geospatial visualization forming a comprehensive framework. Radar plots further enable visualization and comparison of chemical signatures, facilitating rapid identification of characteristic markers for each salinity source. HCA supports this by systematically categorizing samples based on their chemical similarities, thus streamlining the investigative process.

Together, UMAP, radar plots, and HCA form a robust analytical toolkit that enables environmental consultants and regulators to precisely and efficiently focus on contamination sources of concern. This approach minimizes unnecessary monitoring of unrelated sources, significantly reducing the complexity and cost associated with site investigations and long-term management.

Initially developed nearly two decades ago, this integrated method, now enhanced by machine learning with UMAP, has been successfully applied at numerous sites across Alberta, setting a new standard for forensic analysis and management of salinity impacts. The application of this workflow will be demonstrated on case studies to demonstrate different scenarios of sites where it has been successfully applied.

Court Sandau

Court D. Sandau, Ph.D., P.Chem., FRSC is the principal of Chemistry Matters Inc and co-founder of Statvis Analytics Inc. Dr. Sandau has 30 years of experience in the measurement, interpretation and statistical analysis of environmental contaminants. He has been applying chemical fingerprinting and environmental forensic analysis on contaminated sites for 20 years of his environmental consulting experience. The tools and techniques developed and used routinely by Dr. Sandau are now being incorporated into Statvis' web-based software so it can be used by all those environmental professionals interested.