Title: Data associated with “Numerical modeling and data-worth analysis for characterizing the architecture and dissolution rates of a multicomponent DNAPL source”

Authors: the aqueous-phase concentration dataset was created by Kim Broholm, Stanley Feenstra, and John Cherry (Broholm et al. 1999: Solvent Release into a Sandy Aquifer. 1. Overview of Source Distribution and Dissolution Behavior). Permission to publish the dataset was granted by Stanley Feenstra (personal communication. [dnapl@sympatico.ca](mailto:dnapl@sympatico.ca)). All other files were produced by Andres Prieto-Estrada ([aestrad@vt.edu](mailto:aestrad@vt.edu)).

Corresponding Author Name: Andres E. Prieto-Estrada

Item type: datasets and processing/plotting scripts for article Prieto-Estrada et al., 2023: Numerical modeling and data-worth analysis for characterizing the architecture and dissolution rates of a multicomponent DNAPL source.

Keywords: PESTPP-iES, PEST\_HP, Python, SEAM3D, numerical modeling, groundwater, multilevel monitoring, mass discharge, multicomponent DNAPL dissolution, mass transfer

Description: A numerical solute transport model was calibrated to a high-resolution monitoring dataset to characterize a multicomponent source of nonaqueous phase liquids (NAPLs) and evaluate the uncertainty of estimated parameters. The dissolution of NAPL mass was simulated using SEAM3D with parameter zones including adjustable NAPL saturations and mass transfer coefficients, representing the heterogenous architecture of the source zone. Source zone parameters were simultaneously estimated using PEST from aqueous-phase concentrations measured in a multilevel monitoring transect and from mass recovery rates measured at extraction wells during a controlled field experiment. Data-worth analyses, facilitated by PEST ancillary software, linked maximum aqueous-phase concentrations of all compounds to reductions in the pre-calibration uncertainty of mass transfer coefficients. In turn, decreasing concentrations of the most soluble NAPL fraction constrained the source mass estimation. The accurate estimation of model parameters was possible by removing concentrations measured during early NAPL dissolution stages, identified as drivers of model bias using the iterative ensemble smoother PESTPP-iES. Although uncertainty analyses highlighted model limitations for representing sub-grid-scale heterogeneity of NAPL distribution and mass transfer rates, final stages of NAPL dissolution measured at multilevel ports eliminated parameter bias and produced long-term projections of multi-stage source zone depletion . Including mass discharge rates for model calibration further improved the accuracy of estimated residual source mass, complementing multilevel monitoring constraints on the saturation distribution and mass transfer coefficients.

**data.xlsx:** the excel (.xlsx) file named “data” contained measured and simulated multilevel and mass discharge (MD) data (e.g., pce\_measured and pce\_calculated). The file also includes prior-data conflicts (PDC) grouped by multilevel sampling (MLS) port (e.g., tcm\_pdc) and the MD PDC values (MD tab). The linear analysis (LA) results produced with the PEST software utilities GENLINPRED and PREDUNC are in tab Fig10. LA results of individual monitoring measurements of MLS port 606 are in tab Fig11-port606-zone6.1

**tcm\_mls-targets\_pdc.py:** python scripts with this name read the data.xlsx file to generate figures 5 through 7 in the manuscript (Prieto-Estrada et al., 2023).

**model:** folder containing SEAM3D input files and PESTPP-iES input and output files. Inside this folder:

* **Borden-Observations.csv:** contains all MLS and MD measured values, and initial DNAPL mass values in a list. Unique observation names in column A represent DNAPL component, sample port number, and monitoring day (e.g., TCM704\_11, TCM measurement, port 704, day 11).
* **plot\_pce\_MD\_realizations:** python scripts with this name read PESTPP-iES .csv output to plot MD realizations including PDC values.
* **SEAM3D Input Files:** these include the file types ADV, BIO, BTN, DSP, FIL, GCG, HFF, NPL, RCT. The SEAM3D model is available in the Groundwater Modeling System (GMS) licensed by Aquaveo (<https://www.aquaveo.com/products>). The reference Waddill & Widdowson (2000) in the manuscript provides a description of the SEAM3D model.
* **PESTPP-iES Input Files:** these include the file types INS, PST, and TPL. The file Borden-Observations.csv contains the observation data used with PESTPP-iES. The reference White et al. (2020) in the manuscript describes how to construct these files.
* **PESTPP-iES Output Files:** all other CSV files and the file types REC and RMR.

**PEST\_postprocessing:** folder contains results PEST results and a postprocessing script.

* **PEST\_results.xlsx:** contains a list of PEST\_HP results.
* **PEST\_processing.py:** reads PEST\_results.xlsx to produce tables suitable (e.g., pce\_measured tab in data.xlsx) for plotting MLS data and simulation results using the Seaborn library in Python.