Geophysical Research Abstracts
Vol. 21, EGU2019-7701, 2019
EGU General Assembly 2019

Probability distribution of concentration time series data

Artur Kohler

Eötvös Loránd University of Sciences, Faculty of Natural Sciences, Department of Applied Geology, Hungary (kohler@ftr2000.hu)

Groundwater contamination often proves to be a persistent feature of the affected groundwater regime. The groundwater plume is a complex and dynamic entity with a life cycle consisting of expansive, stable, diminishing and exhausting phases. Persistent plumes regularly monitored, concentration data gained by repeated sampling of monitoring wells and laboratory analyses of the samples are used to assess the actual state of the plume. Concentration data are ordered into time series.
Concentration time series are evaluated for trends. Trend evaluation often turns out to be difficult. One of the most common trend evaluation methods is regressions using least squares and using linear regression is widespread.
The significance of the regression line slope is often determined by F-test. F-test calculation requires the calculation of SSE (Sum of Squares Error), which in turn requires the use of data mean. Whether data mean is a representative (expected) value for the data set is depend on data distribution. Data mean is representative for the data set if data are normally distributed, that is one of the reasons F-test is sensitive to data distribution.
Concentration time series data can never be normally distributed from a theoretical point of view, however their factual probability distribution is discussed on rare occasions. Additional problem is that concentration time series datasets are aggregating in time, trends should be evaluated for each increment, and data distribution also should be evaluated for each increment. The resulting process is a data distribution evolution in time.
From a strictly theoretical point of view a time series with three data is a minimum requirement to facilitate trend evaluation. Therefore the proposed distribution evolution assessment starts with n= 3 data and consist n-3 construction of histograms. Probability distributions are identified using these histograms. Identified probability distributions may provide reliable mean or expected values to use in significance testing.