

## **Pattern recognition of Kedah River water quality data by implementation of principal component analysis**

### **ABSTRACT**

This study examines Kedah River Basin, Kedah, Malaysia, to achieve the objective of identifying and recognizing pollutant sources contributing to the water quality using a large dataset extending over a period of eight years, from the year 1997 to 2006. Principal Component Analysis was applied to simplify and provide a better understanding for the complex relationships among water quality parameters such as DO, BOD, COD, SS, pH, NH<sub>3</sub>-NL, temperature, conductivity, turbidity, salinity, dissolved solids, total solids, NO<sub>3</sub>, Cl, Ca, PO<sub>4</sub>, As, Hg, Cd, Cr, Pb, Zn, Ca, Fe, K, Mg, Na, Oil and Grease, MBAS, E.coli and Coliform. Graphical presentation of the data also helps a better view of the overall analysis to appoint sources of pollutant in accordance to their effect. Similar pattern of water quality data reveals nine Principal Components responsible for the data structure and explained 73% of the total variance of the data set. PC score model provided apportionment of various sources contributing to the water quality. Consequently the nine causes of pollutants involved are natural causes in terms of strong river current and geological location of this river, industrial and factories effluent discharge, construction, coal and metal mining, agricultural and sewage plant, human waste and illegal oil dumping.

**Keyword:** Principal component analysis; Water quality; Sources apportionment and source of variations