**Tools of The Trade: Application of the Isometric Logratio Transformation and Analysis of Similarities (ANOSIM) Statistic to VOC Emissions Composition Data from Oil and Gas Wells**

Bart Cubrich, Utah Department of Environmental Quality/Division of Air Quality

Lexie Wilson, Utah Department of Environmental Quality/Division of Air Quality

Oil and gas composition data were collected by sampling more than 70 well sites in the Uinta Basin as part of the Uinta Basin Composition Study. For photochemical modeling and other air quality analytical assessments in the Uinta Basin, it is necessary to apply the results from very few well samples in to the more than 1,100 producing wells in the Uinta Basin. To accomplish this, the composition data from are grouped into composite speciation profiles. Previous grouping efforts for Uinta Basin compositional data have been based on equipment type and company, or well type and geological formation. However, previous composite profiles made no effort to demonstrate the statistical validity or representativeness of their groupings. This novel analysis proposes a statistically robust methodology for grouping the compositional data collected in this study.

The focus of this research was answering two specific challenges of working with composition data:

1) in general, what statistics can be used to tell if a specific grouping schema adequately divides the data into distinct groups, and

2) statistical challenges in dealing with datasets in which each individual sample is represented by percentages of components rather than amounts.

In order to define a statistically robust analysis for these data, it is necessary to transform the data into a Euclidian space, which was accomplished using the isometric log-ratio transformation (ILR; Egozcue, 2003) We then used the Analysis of Group Similarities (ANOSIM) to statistically assess grouping schemes. Using this analysis we show the most statistically defensible grouping scheme is to dived the wells based on sample type and well type (4 profiles). Additionally, using well type, sample type, and formation produces very distinct profiles, but with too few samples per group (13 profiles). We propose these be used for future ozone modeling in the Uinta basin.