

Improved statistical assessment of a long-term groundwater-quality dataset with a non-parametric permutation method

Matthew A. Thomas¹ and Steve W. Wagner²

¹Sandia National Laboratories, 4100 National Parks Highway, Building A, MS 1395, Carlsbad, New Mexico, 88220-9006

²John Hart & Associates, 2815 Candelaria Road NW, Albuquerque, New Mexico, 87107-2914

Abstract

The Waste Isolation Pilot Plant (WIPP) is the only deep geological repository for transuranic waste in the United States. As the Science Advisor for the WIPP, Sandia National Laboratories annually evaluates site data against trigger values (TVs), metrics whose violation is indicative of conditions that may impact long-term repository performance. This study focuses on a groundwater-quality dataset used to redesign a TV for the Culebra Dolomite Member (Culebra) of the Permian-age Rustler Formation. Prior to this study, a TV violation occurred if the concentration of a major ion fell outside a range defined as the mean \pm two standard deviations. The ranges were thought to denote conditions that 95% of future values would fall within. Groundwater-quality data used in evaluating compliance, however, are rarely normally distributed. To create a more robust Culebra groundwater-quality TV, this study employed the randomization test, a non-parametric permutation method. Recent groundwater compositions considered TV violations under the original ion concentration ranges are now interpreted as false positives in light of the insignificant p-values calculated with the randomization test. This work highlights that the normality assumption can weaken as the size of a groundwater-quality dataset grows over time. Non-parametric permutation methods are an attractive option because no assumption about the statistical distribution is required and calculating all combinations of the data is an increasingly tractable problem with modern workstations.

Background

- The stability of Culebra groundwater composition on the timescale of the WIPP operational period is implicit in the steady-state, confined, and 2D flow/transport models used for Performance Assessment.
- Twenty-five rounds of water-quality sampling data (collected between 2000 and 2013) are available to extend the sample size with which to update the current ranges of acceptable ion (i.e., Na^+ , Ca^{2+} , Mg^{2+} , K^+ , Cl^- , SO_4^{2-} , and HCO_3^-) concentrations for the Culebra.
- Simply extending the sample size (i.e., from rounds 1-10 to 1-35), however, would still require the Beauheim et al. (2000) assumption that the data follow a normal distribution.
- The ion concentration data for seven chemical constituents, for six wells, including sampling rounds 1-35, were evaluated for normality and log-normality with the Shapiro-Wilk test (Shapiro and Wilk, 1965). More than 50% of the datasets fail the statistical test for normality and log-normality.

Methods

To provide a more robust assessment of Culebra groundwater composition, we employ a non-parametric method known as the randomization test (Ernst, 2004). Non-parametric methods, including the randomization test, generate reference distributions by calculating a chosen statistic for all combinations of the data. The primary advantage of the randomization test is that no assumption about the distribution of groundwater composition is needed to make comparisons between the sampling round-of-interest and the baseline group of samples.

We follow five steps to employ the randomization test:

1. Calculate the test statistic with the control group and treatment group.
2. Find all combinations of the control and treatment group.
3. Calculate the test statistic for all combinations.
4. Count the number of combinations whose test statistic meet or exceed the threshold criteria.
5. Calculate the p-value to determine if the treatment group is statistically significant compared to the control group.

$$p = \frac{\sum_{i=1}^N I(t_i \leq t^*)}{\binom{N}{n}}$$

where p is the p-value, $\frac{N}{n}$ is the number of combinations, I is the indicator function, t_i is the calculated test statistic, and t^* is the observed test statistic

Findings

Under the revised Culebra Groundwater Composition TV, three consecutive sampling rounds exhibiting a p-value less than or equal to 0.05 constitutes a violation. The ongoing TV violations reported in Wagner and Thomas (2016) are no longer violations when re-interpreted with the non-parametric permutation method. When the sampling round 36 data is re-evaluated with the new TV definition, two of the 42 tests (i.e., six wells multiplied by seven chemical constituents) indicate the treatment group is statistically significant. Concentration histories and test statistic histograms for the two statistically significant cases are shown in Figure 1. When the sampling round 37 data is re-evaluated with the new TV definition, no consecutive offense is observed and, therefore, no possibility of violation exists. Figure 1 shows that the statistically significant sampling round 36 cases are statistically insignificant for sampling round 37.

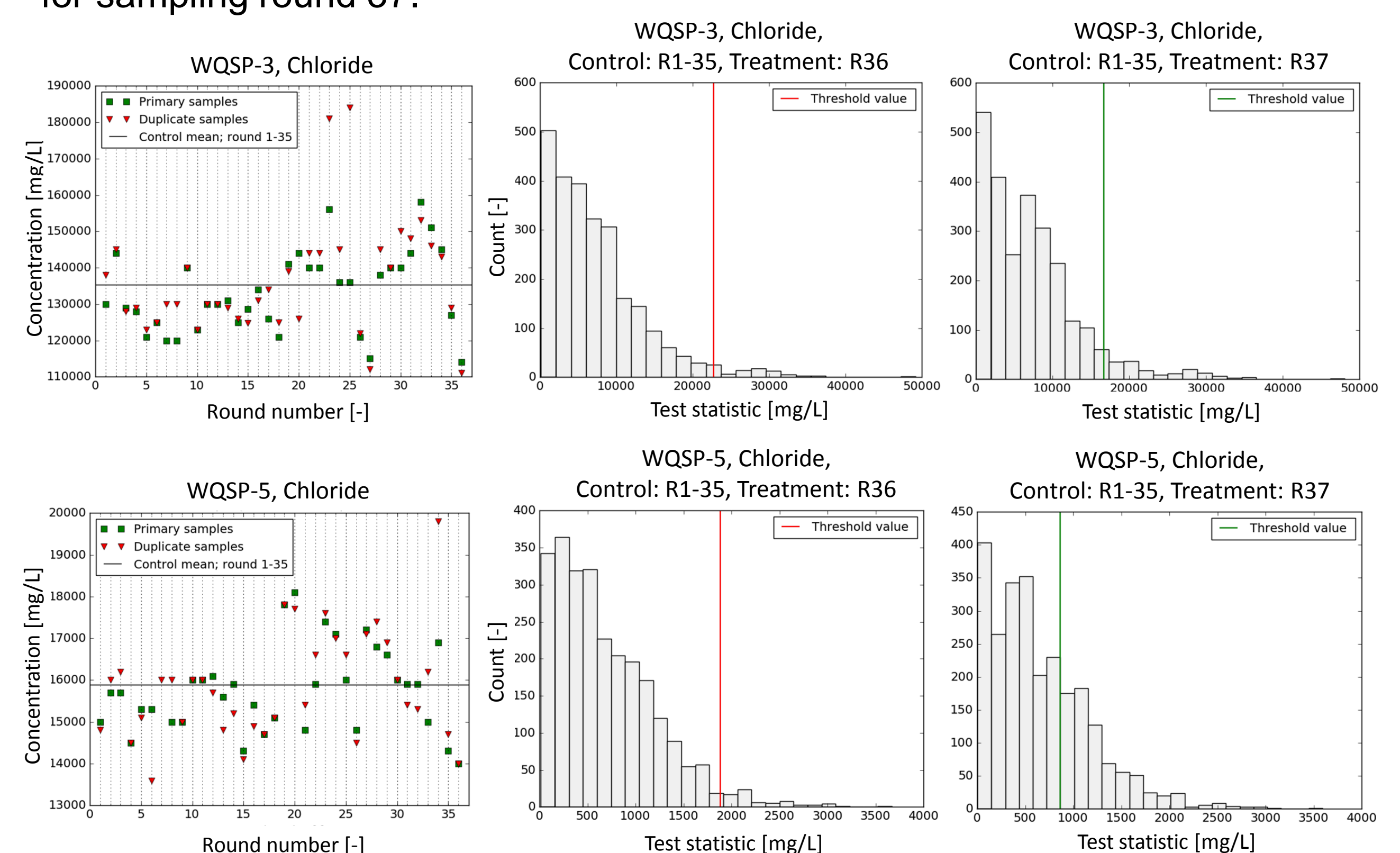


Figure 1 (row1) Concentration history (sampling rounds 1-37) and histograms of test statistic values for sampling rounds 36 and 37 for chloride in well WQSP-3. (row2) Concentration history (sampling rounds 1-37) and histograms of test statistic values for sampling rounds 36 and 37 for chloride in well WQSP-5. Vertical line (green if statistically insignificant, red if statistically significant) indicates the absolute difference between the mean of sampling rounds 1-35 and sampling rounds 36 and 37.

References

- Beauheim, R., F. Hansen, S. Wagner, and D. Guerin. 2000. *Trigger Value Derivation Report*, Revision 0, Sandia National Laboratories, Carlsbad, NM.
- Ernst, M.D. 2004. *Permutation Methods: A Basis for Exact Inference*, Statistical Science, Vol. 19, No. 4, pp. 676-685.
- Shapiro, S.S. and M.B. Wilk. 1965. *An analysis of variance test for normality (complete samples)*, Biometrika, Vol. 52, No. 3/4, pp. 591-611.
- Wagner, S. and M. Thomas, 2016. *Sandia National Laboratories Compliance Monitoring Parameter Assessment for 2015*, WBS 4.3.1, Sandia National Laboratories, Carlsbad, NM.