DRA-0a. USE OF STATISTICS IN THE CHARACTERIZATION PROGRAM OF THE QRPWASP

V.T. McLemore, December 31, 2008, revised March 30, 2009 (reviewed by L. Brandvold, Ariel K. Dickins, G. Robinson)

1. STATEMENT OF PROBLEM

What statistics were used in the characterization program of the QRPWASP? Statistics are an integral part of the characterization program, but the types and methods utilized are not always described in the reports and DRAs, although some of the SOPs do describe statistics required for specific analyses of data reduction by the laboratories. The purpose of this component DRA is to summarize the various statistical analyses used in the characterization program.

2. PREVIOUS WORK

Statistics are mathematical tools used to describe, summarize, analyze, interpret and present data and, if the data are robust enough, also can be used to make future predictions (Cheeney, 1983; Wellmer, 1998; Schreuder et al., 2004). Descriptive statistics are mathematical tools that summarize or reduce the data sets to as small a subset of numbers as possible using indicators such as the mean, median, mode, standard deviation, coefficient of variation, etc. The mean or arithmetic average is the sum of the values divided by the number of values. The median is the middle value of a list ranged from lowest to largest. The mode is the value of a variable that occurs most frequently. The frequency distribution or the frequency of occurrence is typically illustrated by a histogram (a bar or line graph of values where the length of the line shows the frequency of that value), cumulative frequency or density function plots. A normal distribution or bell curve is a density function plot where the curve is regularly shaped; other distributions of data are likely (Schreuder et al., 2004). The standard deviation is a measure of how much the data deviates from the mean.

Several additional statistical tools were utilized in the QRPWASP. Regression analysis is the process used to determine the best fit of a line or curve to a set of data consisting of dependent and independent variables; linear regression is used to describe how well the trend fits along a line. Curvilinear regression is used to describe how well the trend fits along a curve. Factor analysis is a statistical tool used to group a set of many observed variables into two or more related factors. A correlation is used to define the linear relationship between two variables. Pierson correlation is obtained by dividing the covariance of the two variables by the product of their standard deviations. A correlation coefficient of 0.8 is generally a good correlation, although in some cases lower correlation coefficients could be acceptable. The term trend is used when the data indicate a relationship, but not a strong linear or curvilinear correlation.

Multivariate statistics (Pierson correlations, factor analyses, linear regressions, Pearce element ratio analysis) is another powerful tool in analyzing and interpreting geochemical data. Factor analysis was applied to determine the distribution and sources of various elements in soils in Turkey (Kumru and Bakaç, 2003). Reimann et al. (2002) also used factor analysis on a regional data set and discusses some of the difficulties in using such techniques because the nature of the data set (i.e. neither normal or log-normal distributions, strongly skewed, and multimodal data distributions). Whitebread and Moore (2004) described the use of Pearce element ratio analysis and isocon analysis in understanding the alteration of a mineral deposit in Australia. Some of the weathering indices in the literature are based upon statistical analysis (McLemore et al., 2008d).

3. TECHNICAL APPROACH

Numerous statistical packages were utilized in the characterization program to summarize, analyze and interpret the QRPWASP data. Some of these programs include Excel, WinStat, SigmaStat, ProStat, Rockware, and PSPlot. All of these computer programs utilize the same fundamental statistical mathematics and include documentation. The geochemical and some of the geotechnical laboratory data utilized means and error analysis in order to provide values for the parameters. Duplicate and triplicate measurements were averaged. Relationships, trends, and correlations of the geological, geochemical, mineralogical, and geotechnical data were examined using these statistical methods. Outliers are the sample points that appear to be inconsistent with the remainder of the collected data (Iglewicz, 1993). Possible sources of outliers are recording and measurement errors, incorrect distribution assumption, unknown data structure, or novel phenomenon. Numerous techniques can be used to statistically identify outliers. Wellmer (1998) and Schreuder et al. (2004) summarize the mathematical theory behind most of the statistics used in the QRPWASP characterization program.

4. CONCEPTUAL MODEL

Standard statistical practices were used in the QRPWASP as summarized by Wellmer (1998) and Schreuder et al. (2004).

5. STATUS OF COMPONENT INVESTIGATION

Statistics were used at various levels in the QRPWASP (Table 1), but were not always specifically described in the project reports and DRAs. Descriptive statistics were used throughout the project and are included in most reports and DRAs to describe and summarize the various data collected in the project. McLemore et al. (2008a, c) and other reports further document statistical analysis used in this project. Histograms were plotted for many parameters in order to determine their distributions. The characterization team attempted to use statistical procedures to define correlations, which could then be used to make future predictions. However, much of the project data were too variable to define distinct correlations; only trends could be established, which were not significant enough to be used to make reliable predictions (DRA-27, 51). Multivariate statistics (Pierson correlations, factor analyses, linear regressions, Pearce element ratio analysis) were used to determine if there were any correlations between the various geological, geochemical, and geotechnical parameters, but were unsuccessful. Much of the characterization data were used to define trends, because the data were not robust enough to define significant correlations. T-tests and ANOVA (analysis of variation) were used to compare if groups of data were statistically similar or different. If t-tests could not be applied because the normality test failed (i.e. distribution of data are not normal), then the Mann-Whitney Rank Sum Test was used. ANOVA (Analysis of variances) results were obtained using

Kyuskal-Wallis One Way Analysis of variance on ranks, which is used on data that do not have a normal distribution (SigmaStat@).

Graf (2008) used the isocon method in his analysis of the residual weathering profiles in the alteration scars. The isocon method (Grant, 1986) was developed for metasomatically-altered rocks, but Graf (2008) modified the method to document changes in these weathering profiles. The final results from the isocon calculations document changes in the gains or losses of elemental oxide compositions relative to the least altered sample.

TABLE 1. Summary of specific QRPWASP tasks that used statistical analysis. Access database refers to the project database, which is available upon request.

Task	Location of specific data	Type of statistics used	Purpose of statistics	Reference
Lithologic atlas	McLemore et al. (2008e), Access database (412minpet3_29_09 table)	Average, standard deviation, ANOVA, t-tests	To show the variations in composition between rock type and alteration	DRA-1; McLemore et al. (2008e)
Questa rock pile characterization	McLemore et al. (2008a, c), Access database (412minpet3_29_09 table)	Descriptive statistics, histograms, t-tests, ANOVA	To determine if the Questa rock piles are similar in various parameters	DRA-2, 6; McLemore et al. (2008a, c)
Petrography	McLemore et al. (2008a, b, c), Access database (412minpet3_29_09 table)	Descriptive statistics, histograms	Summarize the different rock types and hydrothermal alteration assemblages	DRA-4; McLemore et al. (2008a, b, c)
Heat flow	Reiter (2009)	Curve fitting techniques	To provide temperature profiles through the rock piles	DRA-9; Reiter (2009)
Pore water	Campbell and Hendrickx (2008), Access database (isotopes_solid table)	Best line fit	To define local meteoric line	DRA-12; Campbell and Hendrickx (2008)
Literature review	McLemore et al. (2008b)	Descriptive statistics, histograms	To compare Questa rock piles with rock piles throughout the world	DRA-39; McLemore et al. (2008b)
Geotechnical characterization of Questa rock piles	Geotech summary spreadsheets	Descriptive statistics, histograms, t-tests, ANOVA	To compare geotechnical parameters of Questa rock piles and other materials	DRA-2, 6, 42, 44 ; McLemore et al. (2008a, c)
Slake durability and point load tests	Viterbo (2007), Ayakwah et al. (2009), Access	Descriptive statistics, histograms, best fit	To compare point load and slake durability indices of	DRA-46; Viterbo (2007), Ayakwah et al. (2009)

Task	Location of specific data	Type of statistics used	Purpose of statistics	Reference
	database (412minpet3_29_09 table)	line, t-tests, ANOVA	Questa rock piles and other materials	
Characterization of GHN	McLemore et al. (2008a), Access database (412minpet3_29_09 table)	Descriptive statistics, histograms, best fit line, t-tests, ANOVA	Characterize GHN	DRA-6; McLemore et al. (2008a)
Geochemical, mineralogic, and geotechnical data	McLemore et al. (2008a, c,d), Access database (412minpet3_29_09 table)	Pierson correlations, factor analyses, curve fitting, ANOVA, t-tests	To define relationships and correlations between various parameters	DRA-2, 6, 46; McLemore et al. (2008a, c, d)
Alteration scar geochemical data (Robertson GeoConsultants, NMIMT)	Access database (412minpet3_29_09 table)	Descriptive statistics, histograms, Pierson correlations, factor analyses	To define correlations between various parameters	McLemore et al. (in preparation)

Outliers were generally determined by visual inspection of scatter plots and data tables, or determined by values greater than 1 standard deviation. Proper sample recording and validation procedures as specified in the individual SOPs were employed. However some outliers were identified as recording errors by checking the outlier values with the original data set and such errors were corrected in the final data set. Any remaining outliers were identified, and, if appropriate, described on a case by case basis in the project reports. In some cases, the tests or measurements were rerun to confirm the outlier value; appropriate notes were recorded in the database.

6. RELIABILITY ANALYSIS

Statistical analysis is an important tool in summarizing, describing, analyzing, interpreting and presenting data and can be used as a predictive tool, but must be used with caution and complete understanding of the data limitations and the problem to be addressed. It is important to realize that a correlation does not imply causation or that a correlation cannot always be used to infer a causal relationship between the variables. The characterization team struggled with this concept throughout the project, because correlations of various parameters were attempted before we had a complete understanding of the hydrothermal and weathering processes. Basically, before a correlation could be defined, there needed to be a geologic, geochemical, or geotechnical reason for the correlation to be valid. Furthermore, the variations in the data must be understood. In many cases the variation in sample heterogeneity and the error in the measurements were greater than the potential variation in the variable examined. For example, McLemore et al. (2008a) determined the variations in mineralogical and chemical composition of unweathered Questa rocks were larger than the variations due to weathering since the rock piles were constructed. This natural variation in rock lithology and hydrothermal alteration in most variables is larger than the variation one would expect due to weathering (DRA-27, 51).

Trends in geochemical and geotechnical diagrams are frequently used to describe certain data (von Eynatten, 2004), but are not always quantified or defined statistically (i.e. hand drawn or implied). Van Eynatten (2004) developed a mathematical approach to quantify some of these trends, especially as some of these trends apply to weathering processes (von Eynatten et al., 2003). This approach was examined briefly to describe compositional data in the QRPWASP, but was unsuccessful. Perhaps future studies could examine a similar approach, but in the terms of the sulfide weathering system instead of the silicate system as described by von Eynatten et al. (2003). Another approach that should be examined more rigorously is that the relative differences between observed ranges of values are more meaningful than the absolute values (Martín-Fernández et al., 2005).

7. CONCLUSION OF THE COMPONENT

Statistics were used at various levels in the QRPWASP, but were not always specifically described in the project reports and DRAs (Table 1). Descriptive statistics were used throughout the project and are included in most final reports and DRAs to describe the various data collected in the project. Histograms were plotted for most parameters in order to determine their distributions. The characterization team attempted to use statistical procedures to define correlations, which could then possibly be used to make future predictions. However, much of the project data were too variable to define distinct correlations; only trends could be established, which were not significant enough to be used to make reliable predictions based upon statistics. Multivariate statistics (Pierson correlations, factor analyses, linear regressions, Pearce element ratio analysis) were used to determine if there were any relationships between the various geological, geochemical, and geotechnical parameters, but were unsuccessful. Much of the characterization data were used to define trends not correlations, because the data were not robust enough to define correlations. Statistics were used to compare groups of data.

8. REFERENCES

Ayakwah, G.F., McLemore, V.T., Fakhimi, A. and Viterbo, V.C., 2009, Effects of weathering and alteration on point load and slake durability indices of the Questa rock piles, New Mexico: SME Annual Meeting, Preprint, Denver, Colorado.

Campbell, A. and Hendrickx, J., 2008, Final report on the stable isotopic composition of precipitation and soil samples at the Questa mine: revised unpublished report to Molycorp Tasks 1.12.3 and 1.12.4, 21 p.

Cheeney, R.F., 1983, Statistical methods in geology for field and lab decisions: George Allen and Unwin, London, 169 p.

Graf, G.J., 2008, Mineralogical and geochemical changes associated with sulfide and silicate weathering, natural alteration scars, Taos County, New Mexico: M. S. thesis, New Mexico Institute of Mining and Technology, Socorro, 193 p., http://geoinfo.nmt.edu/staff/mclemore/Molycorppapers.htm, accessed April 28, 2008.

Grant, J.A., 1986, The isocon diagram—A simple solution to Gresens' equation for metasomatic alteration: Economic Geology, v. 81, no. 8, p. 1976-1982

Iglewicz, B. and Hoaglin, D. C., 1993, How to Detect and Handle Outliers: American Society for Quality Control, Milwaukee, WI.

Kumru, M.N. and Bakaç, M., 2003, R-mode factor analysis applied to the distribution of elements in soils from the Aydin basin, Turkey: Journal of Geochemical Exploration, v. 77, p. 81-91.

Martín-Fernández, J.A., Barceló-Vidal, C., Pawlowsky-Glahn, V., Kovács, L.Ó., and Kovács, G.P., 2005, Subcompositional patterns in Cenozoic volcanic rocks of Hungary: Mathematical Geology, v. 37, no. 7, p. 729-752.

McLemore, V.T., Ayakwah, G., Boakye, K., Campbell, A., Donahue, K., Dunbar, N., Gutierrez, L. Heizler, L., Lynn, R., Lueth, V., Osantowski, E., Phillips, E., Shannon, H., Smith, M., Tachie-Menson, S., van Dam, R., Viterbo, V.C., Walsh, P., and Wilson, G.W., 2008a, Characterization of Goathill North Rock Pile, New Mexico: unpublished report to Molycorp Tasks: 1.3.3, 1.3.4, 1.4.2, 1.4.3, 1.11.1.3, 1.11.1.4, 1.11.2.3, B1.1.1, B1.3.2, 191 p.

McLemore, V.T., Fakhimi, A., Ayakwah, G.F., Anim, K., Boakye, K., Ennin, F., Felli, P., Fredlund, D., Gutierrez, L.A.F., Nunoo, S., Tachie-Menson, S., and Viterbo, V.C., 2008b, Literature review of other rock piles: characterization, weathering, and stability: unpublished report to Molycorp Management, Task B4.2.5, 97 p.

McLemore, V.T., Anim, K., Fakhimi, A., Dickens, A.K., and Nunoo, S., 2008c, Characterization of Questa rock piles, New Mexico: Unpublished report to Chevron Mining, Inc and the QRPWASP, Task B1.1.

McLemore, V.T., Donahue, K., Dunbar, N. and Heizler, L., 2008d, Characterization of physical and chemical_weathering in the rock piles and evaluation of weathering indices for the Questa rock piles: unpublished report to Chevron, Task 1.3, B1.1.

McLemore, V.T., Sweeney, D., and Donahue, K., 2008e, Lithologic atlas: revised unpublished report to Molycorp (revised from appendix 2.3, May 2005 report).

Reimann, C., Filzmoser, P., and Garrett, R.G., 2002, Factor analysis applied to regional geochemical data: problems and possibilities: Applied Geochemistry, v. 17, p. 185-206.

Reiter, M., 2009, Fluid flow and the geothermal regime in the Roadside rock piles at the Chevron Mining Inc. molybdenum mine, Questa, New Mexico: publication in review.

Schreuder, H.T., Ernst, R., and Ramirez-Maldonado, H., 2004, Statistical techniques for sampling and monitoring natural resources: U.S. Department of Agriculture, Rocky Mountain Research Station, Fort Collins, CO, General Technical Report-GTR-126, 111 p.

Viterbo, V., 2007, Effect of premining hydrothermal alteration processes and postmining weathering on rock engineering properties of Goathill north rock pile at the Questa Mine, Taos County, New Mexico: M. S. thesis, New Mexico Institute of Mining and Technology, Socorro, NM, 209 p., http://geoinfo.nmt.edu/staff/mclemore/Molycorppapers.htm, accessed March 10, 2008.

von Eynatten, H., Barceló-Vidal, C. and Pawlowsky-Glahn, V., 2003, Modelling compositional change: the example of chemical weathering of granitoid rocks: Mathematical Geology, v. 35, no. 3, p. 231-251.

von Eynatten, H., 2004, Statistical modelling of compositional trend in sediments: Sedimentary Geology, v. 171, p. 79-89.

Whitebread, M.A. and Moore, C.L., 2004, Two lithogeochemical approaches to the identification of alteration patterns at the Elura Zn-Pb-Ag deposit, Cobar, New South Wales, Australia: use of Pearce element ratio analysis and Isocon analysis: Geochemistry, Exploration, Environment, Analysis, v. 4., p. 129-141.

Wellmer, 1998, Statistical evaluations in exploration for mineral deposits: Springer, New York, 379 p.

9. TECHNICAL APPENDICES

DRAs, SOPs, final reports