Multivariate Statistical Analyses of Groundwater Surrounding Fortymile Wash

A. Woocay, J.C. Walton
Environmental Science and Engineering
University of Texas at El Paso, El Paso, TX 79968
USA

ABSTRACT

Groundwater chemistry data from 211 sampling locations in the vicinity of Yucca Mountain, Nevada are analyzed using multivariate statistical methods in order to better understand groundwater chemical evolution, ascertain potential flow paths and determine hydrochemical facies. Correspondence analysis of the major ion chemistry is used to define relationships between and among major ions and sampling locations. A *k*-means cluster analysis is used to determine hydrochemical facies based on correspondence analysis dimensions. The derived dimensions and hydrochemical facies are presented as biplots and overlaid on a digital elevation model of the region giving a visual picture of potential interactions and flow paths. A distinct signature of the groundwater chemistry along the extended flow path of Fortymile Wash can be observed along with some potential interaction at possible fault lines near Highway I-95. The signature from Fortymile Wash is believed to represent the relict of water that infiltrated during past pluvial periods when the amount of runoff in the wash was significantly larger than during the current drier period. This hypothesis appears to be supported by hydrogen-2 and oxygen-18 data which indicate that younger groundwater is found in the upper part of the wash near Yucca Mountain and older groundwater is found in the lower region of the wash near Amargosa River. The range of the hydrogen-2 data corresponds to precipitation in a period of relatively cold climate and has a similar spatial signature to the oxygen-18 data. If the hypothesis that current groundwater chemistry primarily reflects past focused infiltration of surface runoff rather than regional groundwater migration is correct, then saturated zone transport from Yucca Mountain may be much slower than is currently anticipated.

INTRODUCTION

The conceptual hydrogeologic model of the Yucca Mountain region has evolved as more data are gathered and understanding of the region increases [1]. Several authors [2, 3, 4, 5] have conducted mathematical modeling of the Yucca Mountain conceptual model at the site and/or regional scale, and base their confidence in modeling results by comparing calculated to observed hydraulic heads, estimated to measured permeabilities, and comparing their result to results obtained by other mathematical models. Authors in references [2, 3, 4, 5] present groundwater flow in the Amargosa Desert region, generally from areas of higher hydraulic head under the mountains to the north to the low hydraulic head regions in the south (i.e. towards the Funeral Mountains)

Correspondence Analysis (CA) is a statistical analysis method that may reveal grouping or trends in data that otherwise would be overlooked and which are caused by external, often unknown, factors. In CA an Eigenvalue problem is applied to a weighed contingency table of the data [6], [7, page 514] and the derived Eigenvectors, generally called dimensions, are mutually orthogonal and each successive dimension includes a decreasing amount of the system’s variation, with the first one being the more interpretable. In this method, the relationship between variables and cases is simultaneously examined, thus making it a Q-mode type of analysis [6, 8]. CA is better that a principal Component Analysis (PCA) at examining non-linear relationship between variables and cases, such as a unimodal response, since it does not assume a linear relationship between variables or cases [9].

A variety of modeling approaches, including CA and PCA statistical methods, have been applied to better understand groundwater flow paths in the vicinity of Yucca Mountain. For example, [8] applied PCA and CA to groundwater data downgradient from Yucca Mountain, where the data set consisted of concentrations of major and trace elements in groundwater samples collected from 14 Nye County Early Warning Drilling Program (EWDP) monitoring wells. The major ion data were correlated with the trace elements data to aid in the interpretation of the analyses. The authors [8] were able to separate the more concentrated, deeper groundwaters, and therefore believed older, from the more dilute or younger groundwaters from volcanic rock (high SO_4^{2-}-HCO_3^- correlation). These waters in turn were shown to differ from those from carbonate rock (high Ca^{2+}-Mg^{2+}-HCO_3^- correlation). Finally; they were able to some extent separate waters exhibiting reducing conditions from waters exhibiting oxidizing conditions (i.e. As^{3+} vs. As^{5+}).

Kwicklis et al. [10] estimated groundwater flow paths in the Yucca Mountain region using Cl^-, SO_4^{2-}, and oxygen-18 as nonreactive tracers and assuming that groundwater flow is two-dimensional and that mixing and recharge could be neglected as a first approximation. The estimated flow paths were then used, along with PHREEQC, to develop inverse models to explain their chemical and isotopic composition though groundwater mixing and water/rock interactions. The inverse mixing and reaction models were used to verify the feasibility of flow paths estimated from non-reactive tracers.

The work presented herein includes a CA of major ion data from 211 sampling locations in the vicinity of Yucca Mountain, Nevada. Most sampling locations are wells while the rest are springs; furthermore, springs with high ion concentrations were eliminated from the analysis to avoid distortion of the analyses due to outliers. Multiple sampling locations were developed at wells sites that had more that one screen depth at which separate groundwater sample could be taken. CA results are then analyzed by a *k*-means Cluster Analysis (KMCA) Results are presented in a manner where similarities in chemical composition can easily be observed indicating possible groundwater flow pathways and/or evolution. Supporting groundwater stable isotopes (hydrogen-2 and oxygen-18) data interpretations are also presented.

METHODS

Correspondence Analysis

Statistica™ 7 [11] was used to conduct CA of major ion groundwater chemistry from 211 sampling locations in the vicinity of Yucca Mountain The input to the analysis is a tabular data file with columns representing major ion concentrations and rows representing each sampling location. In the CA the data is first normalized, thereby removing the input concentration units in the data, and making the results dimensionless. Next a series of new orthogonal axes (referred to as dimensions) are derived from the data. Dimension coordinates for each of the observation are also generated as part of the CA, and are used as input for the KMCA presented in the following subsection. The CA method allows visualization of how the sampling locations relate to each other, how the ions relate to each other, and how the samples relate to the ions. Since CA is based on a Chi-Squared type analysis, both linear and non-linear trends in the data may be revealed. CA can be viewed as a qualitative tool, which may visually reveal underlying trends and controlling features in the data

k-Means Cluster Analysis

A *k*-means cluster analysis (KMCA) attempts to minimize the variability within each cluster while maximizing the variability between clusters. The mean of a cluster, or centoid, has its components specified by the average of each variable in the analysis. The algorithm uses one initial observation per cluster as the mean for that cluster, and then evaluates each of the remaining observations for inclusion into a particular cluster. Initial observations are selected to maximize initial Euclidean distances between clusters and the number of clusters (*k*) is predetermined. As each observation is included, or clustered, the mean of each cluster is recalculated and previously clustered observations are re-evaluated to check for correct clustering. Observations and *k* means are evaluated at each step until no further improvement can be achieved and all observations have been clustered.

Using the same statistical software, CA results were then evaluated with a KMCA to cluster sampling locations with similar characteristics into nine separate sample groups, or hydrochemical facies. The KMCA variables evaluated are the three dimension coordinates generated for the sampling locations, and the observations are the coordinates for each sampling location. Both empirically and from previous analysis, it was determined to use nine groups for the KMCA.

Oxygen-18 and Hydrogen-2 Isotope Data

The stable oxygen-18 (O-18) and hydrogen-2 (H-2) composition of soil water and groundwater reflects the isotope composition of precipitation, which is correlated with mean annual temperature and may thus provide paleoclimate information [12]. Due to the process of fractionation at the moment of condensation, lighter waters are associated with cold temperatures and heavier waters are associated with warmer temperatures.

RESULTS

Table I presents the separation of the data's variation into the first three dimensions and the major ions' new coordinates onto these dimensions. The first three CA dimensions explain respectively 50.3, 28.6 and 9.6 percent of the variance in the data.

Table I First Three Dimensions' Coordinates for Major Ion Data

	Dimension 1	Dimension 2	Dimension 3
pH	-0.328	0.057	-0.145
Ca^{2+}	0.319	0.293	-0.207
Mg^{2+}	0.576	0.594	0.210
Na^{+}	-0.169	-0.179	0.028
K^{+}	0.117	0.089	-0.199
Cl^{-}	0.233	-0.348	-0.281
SO_4^{2-}	0.360	-0.118	0.104
Alkalinity	-0.146	0.102	0.013
Variation Explained (%)	50.3	28.6	9.6

Table II presents the nine hydrochemical groups or facies determined from the KMCA of the CA results; the names are spatially or compositionally descriptive and are presented along with the mean concentration of major ions demonstrating the different average ions compositions between the groups.

Plotting of the results in a biplot allows both the normalized results for each observation (i.e. sample location) and each variable (i.e. ion) to be plotted relative to the new dimensions. Figure 1 shows a biplot of the CA results for the first two dimensions for the 211 sampling locations along with the relative composition of the major ions presented as green asterisk symbols whose location is derived from Table I. Using the biplot format, all the samples and all the ions can be plotted against the same new dimensions, thereby revealing the correspondence between samples and variables. Note that the positive and negative signs are not significant, only the relative locations along the new dimensions are important. The ions are plotted with larger symbols to help distinguish them visually from the sampling locations. Samples near a particular ion indicate a proportionately higher than average concentration of that ion in those samples. Dimension 1 can roughly be interpreted as the separation of samples into Ca^{2+} and $Mg^{2+\ -}$ on one end and pH, alkalinity on the other. In Figure 1, samples where there is a proportionately higher amount of alkalinity when compared to the other major ions are on the left side of Dimension 1 and those with a proportionately higher amount of Ca^{2+} and Mg^{2+} on the right side. Dimension 2 can be interpreted as roughly as the separation of samples into Ca^{2+} and Mg^{2+} on one end and Cl^- and SO_4^{2-} on the other end. Samples with proportionately higher Cl^- and SO_4^{2-}, and to a lesser extent Na^+, plot in the lower region of Dimension 2, while wells with proportionately higher Ca^{2+} and Mg^{2+} plot in the upper right quadrant on Figure 1. The third Dimension, explaining less than 10% of the variation in the data and is not graphically shown here. Dimension 3 is more difficult to interpret, but it is interesting to point out that Mg^{2+} and Cl^- oppose each other in this dimension.

Table II Summary of Major Ions and Isotopic Data by Hydrochemical Facies Derived by *k*-Means Cluster Analysis of Sample Coordinates

Group Name	pH	Ca^{2+} (mg/L)	Mg^{2+} (mg/L)	Na^+ (mg/L)	K^+ (mg/L)	Cl^- (mg/L)	SO_4^{2-} (mg/L)	Alkalinity as $CaCO_3$ (mg/L)	TDS (mg/L)	H-2 (permil)	O-18 (permil)
Yucca Mountain West Face	8.5	2.7	0.19	87.9	2.35	7.45	25.9	161.0	313.2	-101.4	-13.6
Yucca Mountain East Face	7.8	13.4	1.23	61.6	3.97	11.1	25.8	125.0	269.0	-102.4	-13.6
Mid Fortymile Wash	7.8	21.6	2.54	40.6	6.34	8.71	25.0	109.7	238.6	-100.8	-12.9
Oasis Valley	8.1	11.3	0.81	123.2	5.95	42.3	77.0	159.7	454.5	-112.0	-14.7
Amargosa Desert SW	7.8	46.4	8.91	128.2	10.4	56.4	156.8	195.4	645.4	-104.4	-13.4
High Alkalinity	7.9	26.0	9.22	90.8	9.84	16.2	69.3	214.4	482.7	-106.2	-13.6
East of Fortymile Wash	8.2	15.7	3.18	113.7	6.60	23.3	115.1	146.0	454.1	-108.3	-14.4
Amargosa Desert East	7.7	40.5	12.9	97.2	12.0	25.4	129.6	205.4	568.0	-104.9	-13.6
High Ca & Mg	7.7	56.2	28.3	69.8	7.75	18.4	110.3	253.2	599.5	-101.4	-13.2

When overlaid on a digital elevation model, the well groups and derived dimensions give a visual picture of potential interactions and flow paths in the region. More specifically, the first two CA dimensions were contoured in Surfer™ 8 software using natural neighbor gridding option and overlaid on the digital elevation model of the region. The first two contoured CA dimensions are shown, respectively, in Figures 2 and 3 along with red and blue arrows respectively demonstrating the approximate paths of the Amargosa River and Fortymile Wash. The first thing to notice is that although the hydrochemical facies were based entirely on the CA, they also show geographic trends. In Figure 2, samples found within the low (i.e. negative) contour area generally exhibit lower concentrations of Cl^- and SO_4^{2-}, in proportion to concentrations of alkalinity, than samples outside this area. Contours of Dimension 1 exhibit a strong trough following along the trend of Fortymile Wash, suggesting that the ground waters along the trough may share a common genesis.

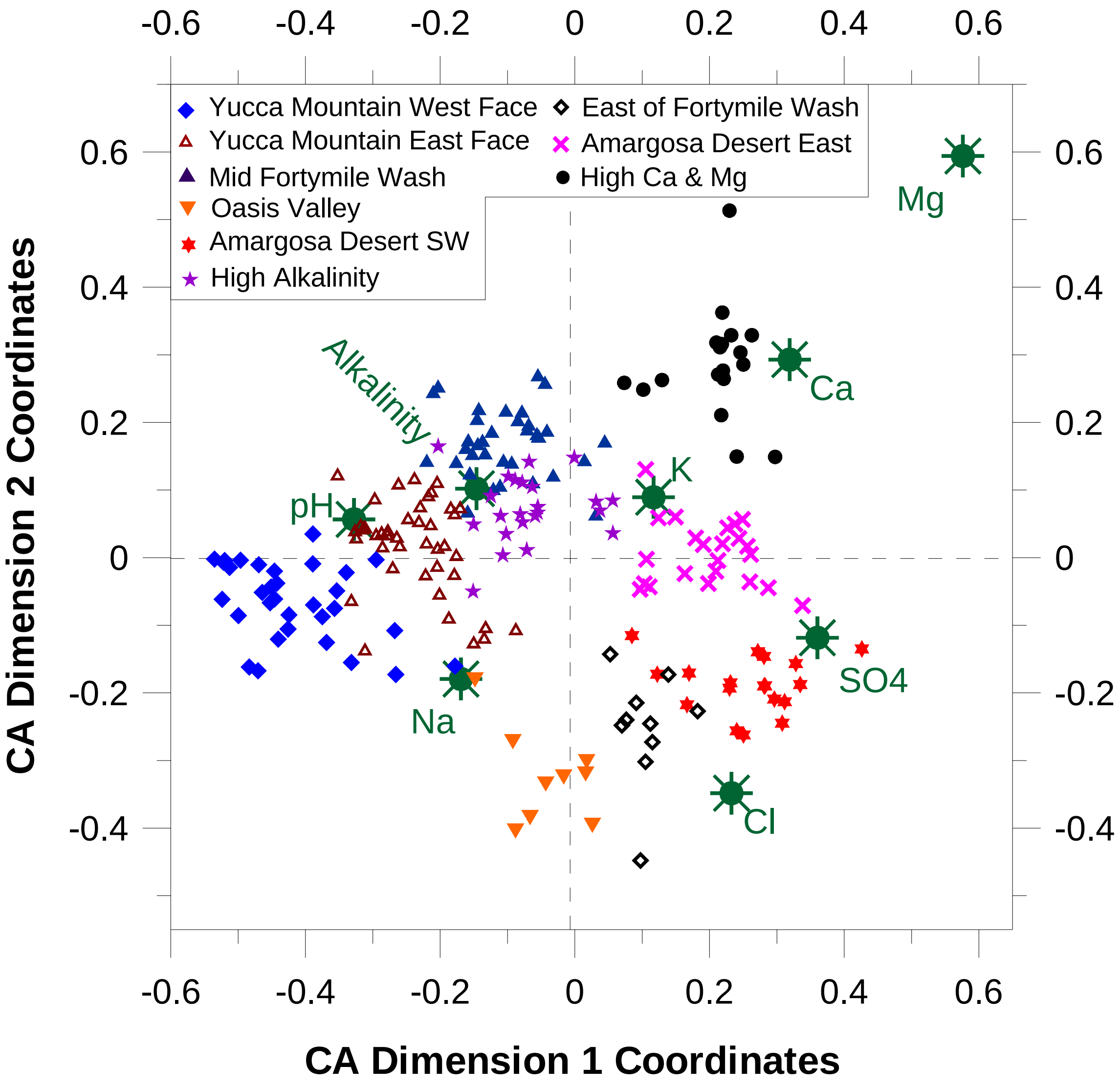

Figure 1 Correspondence analysis biplot with major ions and samples grouped into nine hydrochemical facies

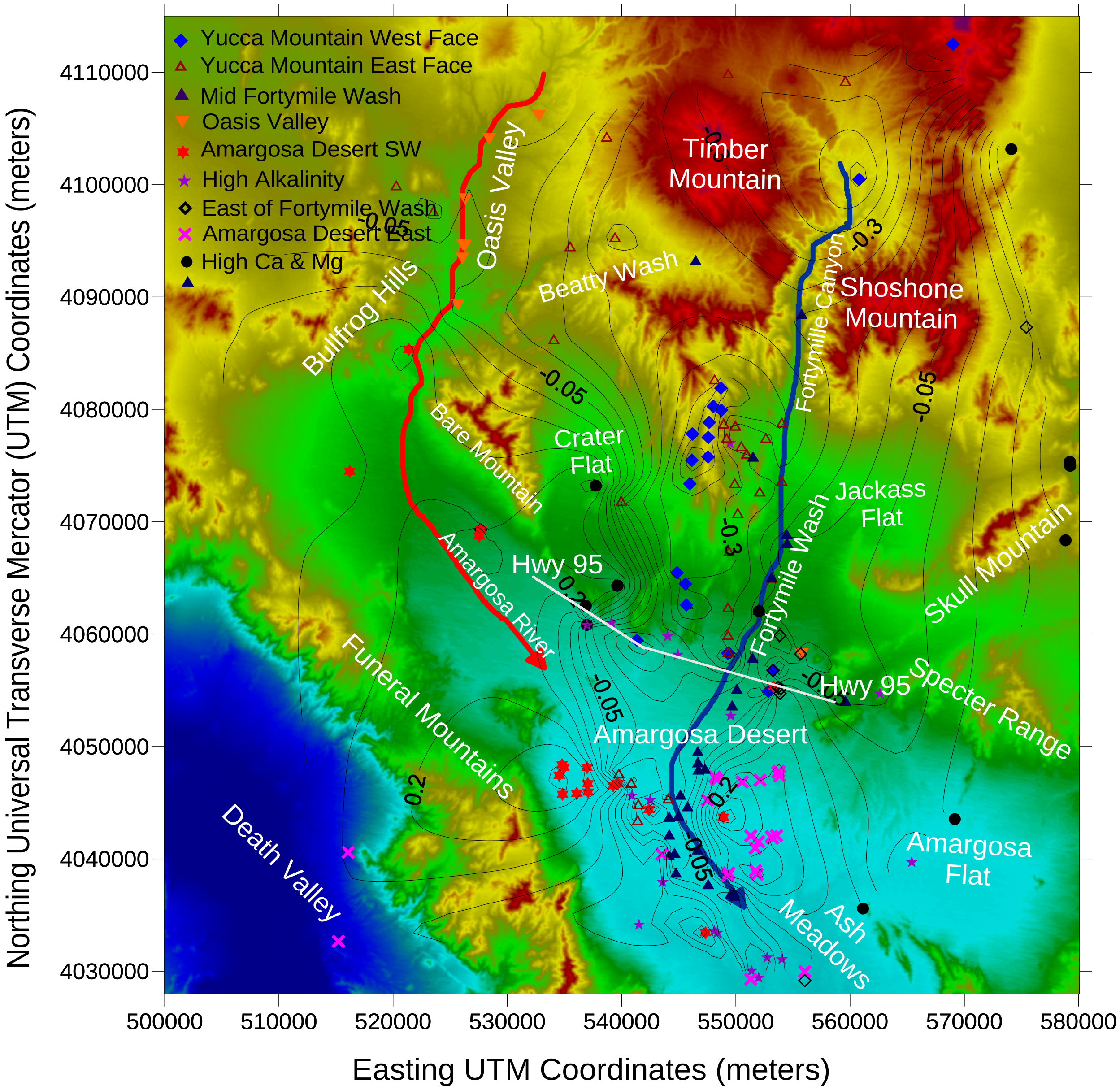

Figure 2 Correspondence analysis’ Dimension 1 overlain on a digital elevation model showing the Amargosa Rive and Fortymile Wash flowpaths based on results from the multivariate statistical methods, along with sampling locations grouped into nine hydrochemical facies

In Figure 3, high contour areas represent waters with higher relative amounts of Ca^{2+} and Mg^{2+} and are thought to reflect the influence of carbonate sediments from Bare Mountain. In addition, the evolving contours shown in Figure 3 following down the Amargosa River Valley suggest that this drainage may be an important factor in evolution of the ground water chemistry. This second potential flow path is uncertain due to an insufficient density of samples around the Amargosa River. Finally, both of the first two dimensions exhibit discontinuities along the trace of the inferred Highway I-95 fault, suggesting that the fault may be influencing the water chemistry.

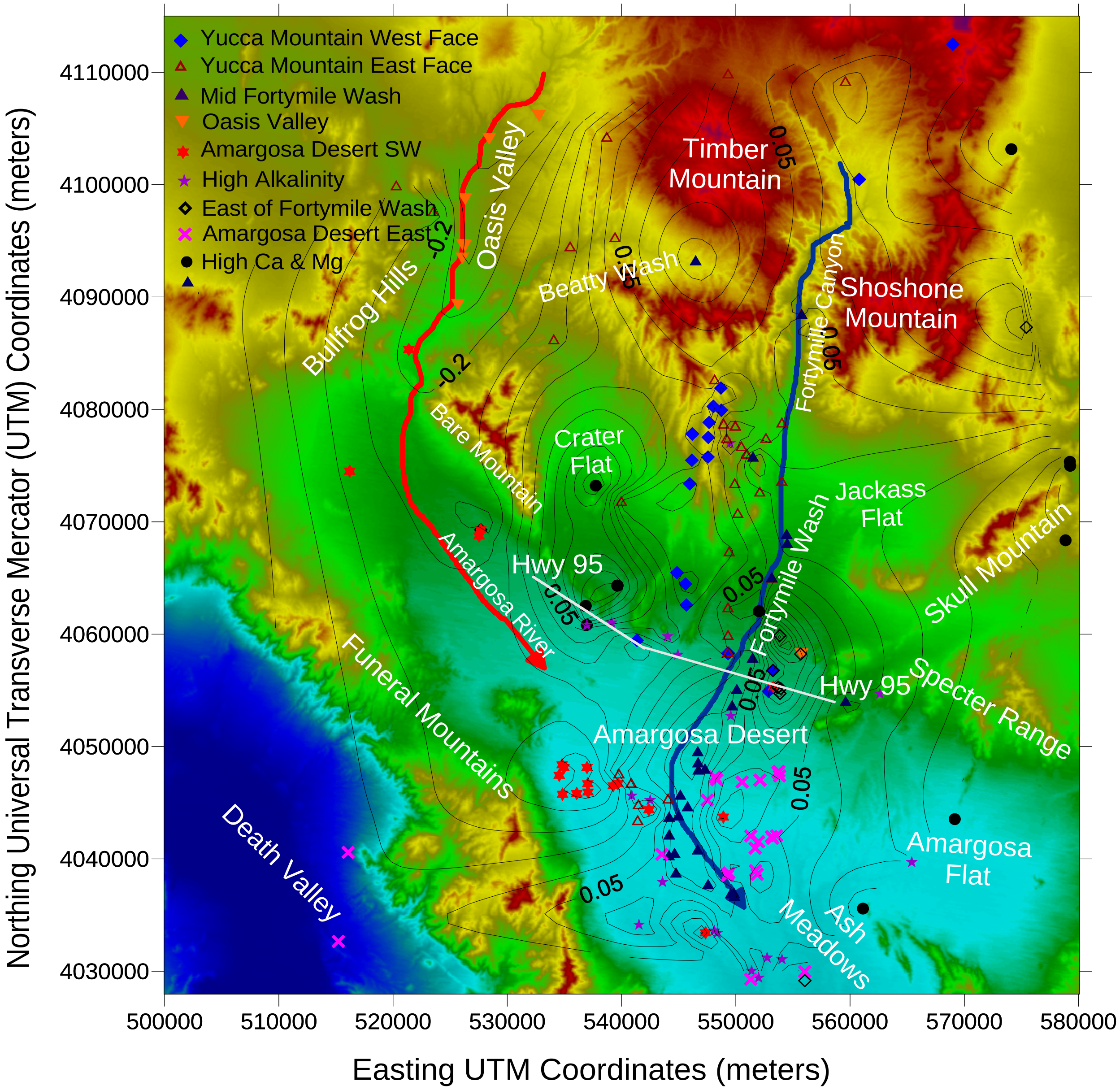

Figure 3 Correspondence analysis' Dimension 2 overlain on a digital elevation model showing the Amargosa Rive and Fortymile Wash flowpaths based on results from the multivariate statistical methods, along with sampling locations grouped into nine hydrochemical facies

In summary, the CA and associated figures indicate that the statistical grouping of the data corresponds to hydrochemical facies and topographical and geologic features in the region. Two spatial trends in the CA dimensions suggest that Fortymile Wash and the Amargosa River have influenced the groundwater chemistry.

The range of H-2 data from 115 locations of the 211 sampling locations for which information was available are shown in Figure 4 and correspond to a relatively cold climate precipitation. Less depleted H–2 values in the upper part of Fortymile Wash correspond to the warmer part of a period of cold climate [13], but still cold relative to the present; whereas more depleted values found in the lower portion of the wash correspond to colder climates. Plots of H−2 values versus O−18 compared with the Global Meteoric Water Line (GMWL) (not shown) demonstrate a humid climate type of precipitation with little surface evaporation before infiltration [13], which along with a cold climate, suggest focused infiltration.

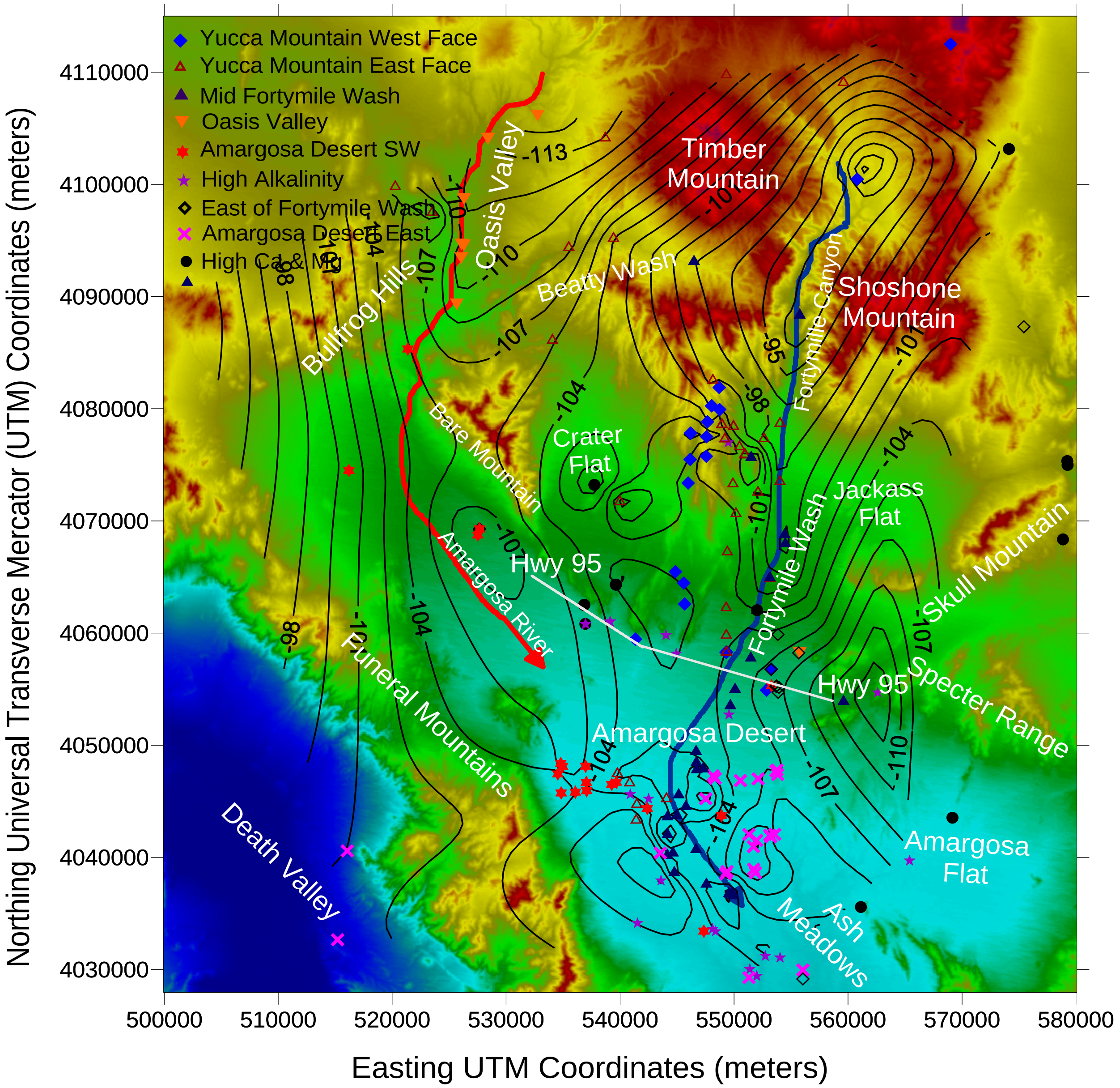

Figure 4 Ground water δH-2 concentrations in permil

INTERPRETATION

The first CA dimension (Figure 2) shows a clear chemical signature for the groundwaters below the surface of Fortymile Wash, which extends down to and then follows the Amargosa River. The second CA dimension (Figure 3) shows the possible influence both of carbonate rocks in Crater Flat and of the Amargosa River on ground water chemistry. Both dimensions show a small discontinuity along the trace of the Highway I-95 fault.

Given the age of the water (several thousand years), and how closely the CA signature follows the surface water drainage, one explanation is that current ground water chemistry reflects infiltration of surface water which flowed along Fortymile Wash and the Amargosa River during a past, pluvial climate. Based upon contours of potentiometric head, groundwater would be unlikely to so precisely follow the surface contours in this arid region. The surface water drainage/infiltration hypothesis also appears to be supported by O-18 and H-2 data which indicate that younger warmer groundwater is found in the center of Fortymile Wash relative to the water on either side of the wash. The water along the wash also has lower Total Dissolved Solids (TDS), suggesting it is not derived from migration of adjacent groundwaters, but instead by focused infiltration.

If the hypothesis that current groundwater chemistry along the major washes primarily represents past focused infiltration of surface runoff rather than groundwater migration is correct, then it also follows that groundwater movement since the end of the last ice age has been too slow to erase the old signature. Thus, saturated zone transport from Yucca Mountain may be much slower than currently estimated in regional groundwater flow models [2, 3, 4, 5].

ACKNOWLEDGEMENTS

The Nye County NWRPO funded this work through the U.S. Department of Energy Office of Civilian Radioactive Waste Management cooperative research grant DE-FC28-02RW12163.

REFERENCES

[1]. A. L. Flint, L. E. Flint, G. S. Bodvarsson, E. M. Kwicklis and J. T. Fabryka-Martin, “Evolution of the conceptual model of unsaturated zone hydrology at Yucca Mountain, Nevada”, Journal of Hydrology, 247, 1-30, (2001).

[2]. A. A. Eddebbarh, G. A. Zyvoloski, B. A. Robinson, E. M. Kwicklis, P. W. Reimus, B. W. Arnold, T. Corbet, S. P. Kuzio, and C. Faunt, “The saturated zone at Yucca Mountain: an overview of the characterization and assessment of the saturated zone as a barrier to potential radionuclide migration”, Journal of Contaminant Hydrology, 62–63, 477– 493 (2003).

[3]. J. R. Winterle, A. Claisse, and H. D. Arlt, “An Independent Site-Scale Groundwater Flow Model for Yucca Mountain”, IHLRWM 2003, Las Vegas, NV, March 30-April 2, 151-158, (2003).

[4]. S. Kelkar, P. Tseng, T. Miller, R. Pawar, A. Meijer, B. Robinson, G. Zyvoloski, E. Kwicklis, A. A. Addebbarh, and B. Arnold,” Site/Subsite Scale Saturated-Zone Flow – Transport Models for Yucca Mountain”, IHLRWM 2003, Las Vegas, NV, March 30-April 2, 114-121, (2003).

[5]. J. Liu, E. L. Sonnenthal, G. S. Bodvarsson, “Calibration of Yucca Mountain unsaturated zone flow and transport model using porewater chloride data”, Journal of Contaminant Hydrology, 62–63, 213– 235, (2003).

[6]. M. Mellinger, “Multivariate Data Analysis: Its Methods”, Chemometrics and Intelligent Laboratory Systems, 2, 29-36, (1987).

[7]. A. C. Rencher, ”Methods of Multivariate Analysis”, Second edition, Wiley-Interscience, USA, (2002).

[8]. I. M. Farnham, K. H. Johannesson, A. K. Singh, V. F. Hodge, and K. J. Stetzenbach, “Factor analytical approaches for evaluating ground trace element chemistry data”, Analytica Chimica Acta, 490, pp. 123-138, (2003).

[9]. C. J. F. ter Braak, P. Smilauer, “CANOCO Reference Manual and User’s Guide to CANOCO for Windows: Software for Canonical Community Ordination (version 4)”, Microcomputer Power, Ithaca, NY, 352 pp, (1998).

[10]. E. Kwicklis, A. Meijer, and J. T. Fabryka-Martin, “Geochemical inverse model of groundwater mixing and chemical evolution in the Yucca Mountain area”, HLRWM 2003, Las Vegas, NV, March 30-April 2, (2003).

[11]. StatSoft, Inc. 1984-2006. “SATISTICA for Windows [Computer program manual]”, Tulsa, OK. Kernel release 5.5, www.statsoft.com, (1984-2006).

[12]. B. Liu, F.Phillips, S. Hoines, A. R. Campbell and P. Sharma, “Water movement in desert soil traced by hydrogen and oxygen isotopes, chloride, and chlorine-36, southern Arizona”, Journal of Hydrology, 168, 91–110, (1995).

[13]. I. Clark, and P. Fritz, “Environmental Isotopes in Hydrogeology”, Lewis Publishers, Boca Raton, (1997).