Application of statistical models to estimate total dissolved solids in acid mine drainage

Ana Raquel Barroso[1], Teresa Maria Valente[1], Amélia Paula Reis[1,2], Isabel Margarida Antunes[1]

[1]*ICT – Institute of Earth Sciences, Pole of University of Minho, University of Minho, Braga, Portugal*

[2]*GEOBIOTEC, Geosciences Department, University of Aveiro, Campus Universitário de Santiago, 3810–193 Aveiro, Portugal*

Abstract

The present study presents an unexplored statistical approach using linear and multiple regression to estimate total dissolved solids in acid mine waters. One hundred twenty-one samples were collected from three distinct abandoned mining areas in the Iberian Peninsula: Valdarcas, São Domingos, and Campanario. The parameters analyzed included pH, electrical conductivity, total dissolved solids, acidity, and sulfate concentration. Discriminant analysis and linear regression were used to identify the best predictor variables of total dissolved solids. The results showed a high variability between the mining areas due to the hydrochemical characteristics at each site. The most relevant variables differentiating the three mines are sulfate, acidity, and electrical conductivity. However, for these specific samples, sulfate and acidity proved to be the most relevant parameters for estimating total dissolved solids.

Keywords: Hydrochemistry, Valdarcas north Iberia, Iberian Pyrite Belt, linear regression analysis, multiple regression analysis.

Introduction

Acid mine drainage (AMD) systems are typically characterized by low pH, high concentrations of iron and sulfate, and a wide variety of potentially toxic elements (Valente et al. 2013). The hydrochemical properties of AMD depend on several factors (e.g. geological context, climate condition, mine waste reactivity, and others), which determine their complexity. Assessing the hydrochemical characteristics of AMD is essential to comprehend the biogeochemical processes and evaluate the degree of contamination of the system. Therefore, several physicochemical parameters are analyzed, such as pH, electrical conductivity (EC), total dissolved solids (TDS), sulfate concentration, and total acidity. The determination of some of these parameters is straightforward, while others require laborious laboratory analysis. TDS is a critical parameter for characterizing acid mine waters, but their gravimetric standard analysis is time-consuming and expensive. Therefore, TDS is usually estimated through other parameters, namely EC.

Estimating TDS through EC is usually used in water quality studies. However, this is not always accurate, as several authors suggest a range of conversion factors is needed (Thirumalini and Joseph 2009; Marandi et al. 2013; Rusydi 2018). This conversation factor increases with water mineralization and depends on factors such as the activity of specific dissolved ions and ionic strength (Rusydi 2018). In the case of acidic mine waters, Hubert and Wolkersdorfer (2015) showed that for 45 water samples, the conversion factor varied from 0.25 to 1.34. These authors also suggested that using a single conversion factor can result in wrong estimation since these waters might change with local, diel, or seasonal variations. Barroso et al. (2023) proposed a new approach to estimate TDS based on the acidity of AMD water samples. The study indicates that acidity has a high and significant correlation with TDS, which can be explained by the role of the metal's hydrolysis that contributes with H+ protons to the solution. Furthermore, compared to the conventional estimation through EC, the authors propose this approach as less time-consuming and accurate.

Hence, the present work aims to increase the knowledge about the interactions between the physicochemical parameters that influence TDS estimation. A less conventional methodology using discriminant analysis and linear regression models is explored to identify the best predictive parameter of TDS.

Methods

A total of 121 water samples were collected in three abandoned mining areas of the Iberian Peninsula: Valdarcas, São Domingos, and Campanario (fig. 1). These areas were selected to characterize the hydrochemical diversity typically observed in AMD-affected systems. Valdarcas (NW of Portugal) is an old tungsten (W) mine associated with a skarn deposit (Valente and Gomes 2009). Although it was later rehabilitated, the accumulated waste continues to result in AMD that affects the Coura River (Alves et al. 2011). São Domingos (SW of Portugal) and Campanario (SE of Spain) are old copper (Cu) mines of the Iberian Pyrite Belt (IPB). The IPB is one of the largest metallogenic provinces in the world, comprising several massive sulfide deposits, where the intensive exploitation of base metals resulted in severe environmental degradation (Gomes et al. 2018). While the São Domingos mining complex is undergoing environmental rehabilitation, no remediation project has been implemented in the Campanario mine.

Water samples were collected from AMD in different points (e.g. watercourses, acidic lagoons, AMD streams, and pit lakes) under distinct climate conditions. In the field, pH and EC were measured using multi-parameter equipment (Thermo Scientific Model Orion Star A Series). Acidity, TDS, and sulfate were analyzed using standard methodologies (APHA, 2012). TDS concentration was determined by the standard method 2540 C, where the samples were filtrated onto a glass-fiber filter (Sartorius—glass microfiber discs of 47 mm diameter), dried at 180 °C, and following a cycle of drying-cooling-desiccating-weighing until a constant weight was obtained.

All variables were examined for outliers and normal distribution through histograms, box plots, quantile-quantile plots, and the Shapiro-Wilk test, because the techniques used assume multivariate normal distributions and are sensitive to the presence of extreme values. Variable transformations such as square, square root, and logarithmic transformations

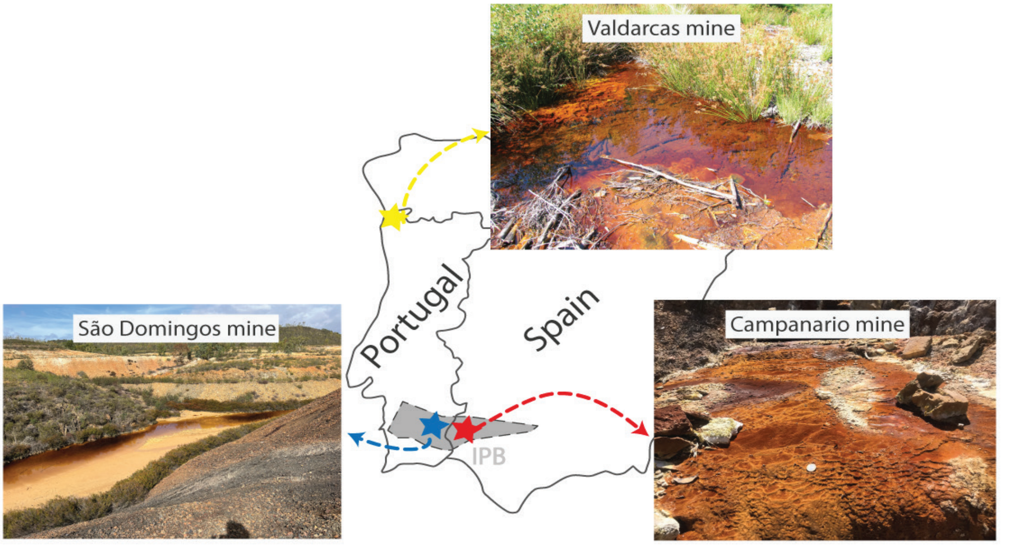

Figure 1 *Location of the three mining areas selected for the present study*

were attempted when normal distribution could not be accepted. The base-10 logarithm of TDS, EC, sulfate, acidity, and the square root of pH were found to help improve the distribution shape. Differences between groups were tested using the Kruskal–Wallis H test, with the results interpreted based on rank differences. A probability greater than 0.05 was considered significant in testing the null hypothesis of no differences across the three mining areas. Post-hoc pairwise multiple comparisons were performed using Bonferroni correction to detect significant differences between two specific mines (Cabral et al. 2019). Spearman's rank correlation coefficients (Spearman's rho) were calculated to identify relationships between the studied variables (Marinho-Reis et al. 2020). Linear discriminant analysis (LDA) is a supervised algorithm that computes the directions (the discriminant functions) that maximize the separation between multiple groups. In this study, LDA was performed to test differences between groups. The discriminant functions were then used to classify samples into groups (the three mines). The number of estimated discriminant functions is one less than the number of classified groups (Marinho-Reis et al., 2020). The Wilks' lambda test ($p < 0.005$) and leave-one-out cross-validation were the methods used to evaluate the ability of the LDA model. Finally, stepwise multiple linear regression (MLR) was performed to identify those hydrochemical parameters that are the best predictors of the TDS. The criteria for stepwise MLR were: probability of F to enter < 0.05 and probability of F to remove > 0.1. The Durbin–Watson test ensured the absence of first-order linear auto-correlation in our multiple linear regression data (Marinho Reis et al. 2018). All statistical techniques were performed using the IBM©SPSS (v. 28) software package.

Results and discussion

The statistical summary shows high variability between mines due to the different hydrochemical conditions observed in the three study areas (tab. 1). In general, pH values were found to be in the acidic range (between 0.44 and 4.82), and the concentration of acidity varied from 96 to 429 250 mg/L $CaCO_3$. The concentrations of sulfate and TDS ranged between 153 to 410 601 mg/L and 296 to 640 086 mg/L, respectively. A statistical analysis of the datasets from the

Table 1 *Statistical summary by mining area for the entire dataset*

		pH	TDS mg/L	EC µS/cm	Sulfate mg/L	Acidity mg/L $CaCO_3$
Valdarcas (n=45)	Min	2.72	296	562	153	123
	Max	3.84	2 149	1 858	1 169	888
	Mean	3.13	1 007	1 300	602	431
	Median	3.11	1 024	1 411	589	381
	SD	0.21	529	436	314	245
Campanario (n=25)	Min	2.50	814	1 099	380	113
	Max	4.82	8 122	5 760	6 710	2 085
	Mean	3.33	4 845	3 799	2 947	1 465
	Median	3.05	5 133	3 790	2 700	1 500
	SD	0.74	1 512	970	1 196	455
São Domingos (n=51)	Min	0.44	510	826	247	96
	Max	4.27	640 086	42 560	410 601	429 250
	Mean	2.49	52 433	8 566	30 778	32 526
	Median	2.50	8 181	6 160	5 019	3 450
	SD	0.75	135 180	9 297	78 146	89 178
Total (n=121)	Min	0.44	296	562	153	96
	Max	4.82	640 086	42 560	410 601	429 250
	Mean	2.90	23 475	4 879	13 806	14 172
	Median	2.92	2 414	2 754	1 389	11 80
	SD	0.70	90 358	6 834	52 291	59 430

***Table 2** Spearman correlation matrix for measured parameters*

	pH	TDS	EC	Sulfate	Acidity
pH	1.00	-0.62	-0.67	-0.63	-0.64
TDS	-0.62	1.00	0.97	0.98	0.97
EC	-0.67	0.97	1.00	0.98	0.96
Sulfate	-0.63	0.98	0.98	1.00	0.98
Acidity	-0.64	0.97	0.96	0.98	1.00

Note: Significant level $\rho = 0.01$

different mines was also performed to better understand the primary factors governing AMD in each mine. São Domingos mine is characterized by significantly lower ($\rho < 0.05$) pH values than the other mines. At the same time, Valdarcas exhibits significantly lower values for the other parameters under study (TDS, EC, sulfate, and acidity), ($\rho < 0.05$).

Spearman correlation coefficients between pairs of variables are shown in Table 2. pH has a strong and significant ($p < 0.01$) negative correlation with all parameters. In contrast, TDS shows a robust significant ($p < 0.01$) correlation with EC ($r = 0.97$), sulfate ($r = 0.98$), and acidity ($r = 0.97$). These relationships are consistent with the typical characteristics of AMD, where the dissolution of sulfides generates acidic conditions and releases various ions into the water. However, the scatterplots suggested a non-linear relationship between TDS, EC, and pH (fig. 2).

Stepwise LDA was performed using the entire dataset to identify those water parameters, providing the best separation between the mining areas. While the Box's M test indicates the groups differ in their covariance matrices, violating an

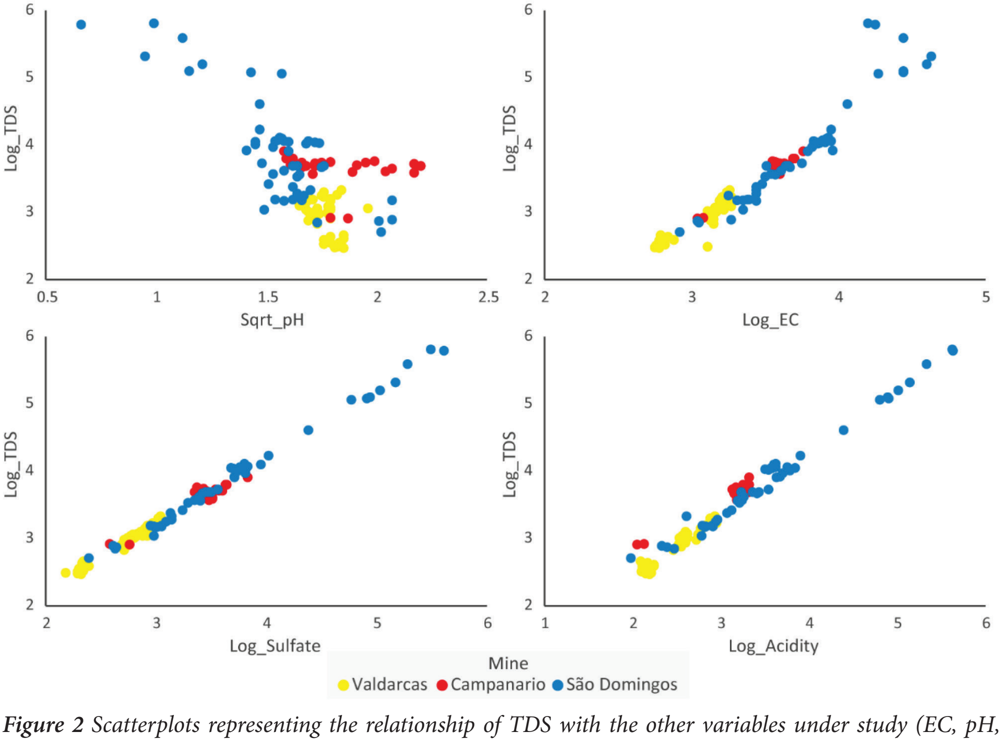

***Figure 2** Scatterplots representing the relationship of TDS with the other variables under study (EC, pH, sulfate, and acidity)*

assumption of LDA, the log determinants are relatively equal and LDA is robust even when the homogeneity of variances assumption is not met. Table 3 presents the sample classification matrix for the stepwise MDA model, where the lines show the actual classification of the samples, and the columns show the classification given by the LDA. EC, acidity, and sulfate are the most significant variables in groups' separation. Based on Table 3, all Valdarcas samples were correctly classified, and two Campanario samples and thirteen from São Domingos were misclassified, according to the leave-one-out classification matrix. Overall, 87.6% of the samples are correctly classified. However, for São Domingos, only 74.5% of the samples were well classified, indicating that the model is poorer for discriminating between São Domingos samples and the other groups. This poorer discriminating ability is well demonstrated by the partial overlapping between groups observed in Fig. 3. This lower discriminating ability can be explained by the diversity of microenvironments within the area, with highly variable physicochemical and ecological properties, where the water samples were collected.

Considering that TDS was not an important factor in the discrimination between mining areas, stepwise MLR was used to investigate which hydrochemical variable better predicts the TDS concentration for AMD samples. The linear models obtained, using the transformed variables, are shown in Table 4. R2 indicates the proportion of the variance accounted for by each regression model, which is not improved by the second model. Nevertheless, both regression models are statistically significant ($\rho < 0.005$). Although acidity and sulfate seem to be good predictors of TDS, sulfate alone explains 99% of the total variance of the variable. This result indicates that, for this dataset, sulfate is the best predictor of TDS.

Conclusions

To evaluate the mine contamination level, the characterization of water and effluents requires the determination of several physicochemical parameters, such as TDS, pH, EC, acidity, and sulfate content. The TDS is often estimated from EC due to the time-consuming laboratory effort for gravimetric analysis. However, AMD has no standard conversion factor, as these waters can have highly different hydrochemical compositions. Therefore, the present work investigated the correlation between typical physicochemical parameters to understand which variable best predicts TDS. The chemical analysis of 121 samples revealed high variability due to the heterogeneity of the three mining areas. Stepwise LDA showed that sulfate, acidity, and EC were the parameters that better classified the water samples into their corresponding mining area. On the other hand, both linear regression models suggested that sulfate is a relevant predictive parameter of TDS. However, the role of the acidity in the estimation of this parameter should not be disregarded, and further investigation is required to increase current knowledge of the interactions between the

Table 3 *Sample classification matrix for the stepwise multiple discriminant analysis (MDA)*

	Valdarcas	Campanario	São Domingos	Total
Valdarcas	45	0	0	45
Campanario	0	23	2	25
São Domingos	6	7	38	51

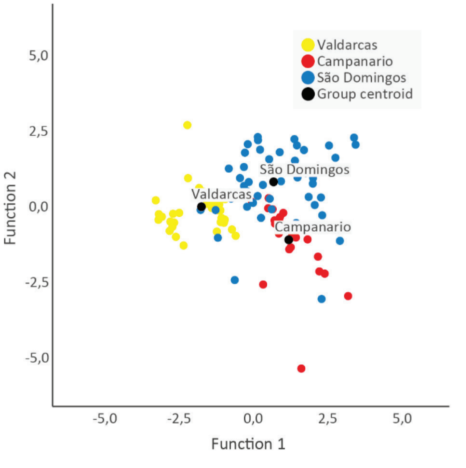

Figure 3 *Combined-group plot resulting from the canonical discriminant functions*

***Table 4** Linear regression models to predict TDS concentrations from hydrochemical parameters.*

Models	Equation	R2
Model 1	{Log_TDS} = 0.237 + 0.996[Log_Sulfate]	99.1%
Model 2	{Log_TDS} = 0.278 + 0.874[Log_Sulfate] + 0.116[Log_Acidity]	99.1%

different physicochemical parameters that influence TDS estimation in AMD systems.

Acknowledgments

This research was funded by FCT (Fundação para a Ciência e a Tecnologia, Portugal) by the research fellowship under the ICT-UMinho (Institute of Earth Sciences, Pole of University of Minho) supported by FCT, I.P. with reference UI/BD/151330/2021. This work was also co-funded by FCT through projects UIDB/04683/2020, UIDP/04683/2020, and Nano-MINENV 029259 (PTDC/CTA-AMB/29259/2017).

References

Alves, R. C., Valente, T. M. F., Braga, M. A., & Gomes, C. L. (2011). Mineralogical composition and metals retention in the fine-fraction streambed precipitates of on AMD affected system. In Proceedings of the International Mine Water Association Congress-IMWA, Aachen, Germany, 4–11 September; ISBN 9781618393050.

APHA Standard Methods for the Examination of Water and Wastewater, 20th ed.; (2012) American Public Health Association, American Water Works Association, Water Environment Federation: Washington, DC, USA.

Barroso, A., Valente, T., Marinho Reis, A. P., & Antunes, I. M. H. (2023). A new acidity-based approach for estimating total dissolved solids in acidic mining influenced water. Water, 15(16), 2995.

Cabral Pinto, Marina M. S., Paula Marinho-Reis, Agostinho Almeida, Edgar Pinto, Orquídia Neves, Manuela Inácio, Bianca Gerardo, Sandra Freitas, Mário R. Simões, Pedro A. Dinis, and et al. (2019). "Links between Cognitive Status and Trace Element Levels in Hair for an Environmentally Exposed Population: A Case Study in the Surroundings of the Estarreja Industrial Area" International Journal of Environmental Research and Public Health 16, no. 22: 4560. https://doi.org/10.3390/ijerph16224560

Gomes, P., Valente, T., & Pereira, P. (2018). Addressing quality and usability of surface water bodies in semi-arid regions with mining influences. Environmental Processes, 5, 707-725.

Hubert, E., & Wolkersdorfer, C. (2015). Establishing a conversion factor between electrical conductivity and total dissolved solids in South African mine waters. Water SA, 41(4), 490-500.

Marandi, A.; Polikarpus, M.; Jõeleht, A. (2013) A new approach for describing the relationship between electrical con-ductivity and major anion concentration in natural waters. Appl. Geochem., 38, 103–109.

Marinho-Reis, A. P., Costa, C., Rocha, F., Cave, M., Wragg, J., Valente, T., ... & Noack, Y. (2020). Biogeochemistry of household dust samples collected from private homes of a Portuguese industrial city. Geosciences, 10(10), 392.

Reis, A. P. M., Cave, M., Sousa, A. J., Wragg, J., Rangel, M. J., Oliveira, A. R., ... & Noack, Y. (2018). Lead and zinc concentrations in household dust and toenails of the residents (Estarreja, Portugal): a source-pathway-fate model. Environmental Science: Processes & Impacts, 20(9), 1210-1224.

Rusydi, A. F. (2018). Correlation between conductivity and total dissolved solid in various type of water: A review. In IOP conference series: earth and environmental science (Vol. 118, p. 012019). IOP Publishing.

Thirumalini, S., & Joseph, K. (2009). Correlation between electrical conductivity and total dissolved solids in natural waters. Malaysian Journal of Science, 28(1), 55-61.

Valente, T. M., & Gomes, C. L. (2009). Fuzzy modelling of acid mine drainage environments using geochemical, ecological and mineralogical indicators. Environmental geology, 57, 653-663.

Valente, T., Ferreira, M. J., Grande, J. A., de la Torre, M. L., & Borrego, J. (2013). pH, electric conductivity and sulfate as base parameters to estimate the concentration of metals in AMD using a fuzzy inference system. Journal of Geochemical Exploration, 124, 22-28.