Prediction of Water Quality Parameters Using Unmanned Aerial Vehicle Multispectral Imagery in Acidic Water Bodies in the Iberian Pyrite Belt (Tharsis, SW Spain)

Melisa Alejandra Isgró[1,2], María Dolores Basallote[1], Luis Barbero[2]

[1]*Department of Earth Sciences, Research Center on Natural Resources, Health and the Environment (RENSMA), University of Huelva, Campus El Carmen s/n, 21071 Huelva, Spain; meli.isgro@gmail.com, maria.basallote@dct.uhu.es*

[2]*Department of Earth Sciences, University of Cádiz, Campus Universitario de Puerto Real, 11510 Puerto Real, Cádiz, Spain, luis.barbero@uca.es*

Abstract

This study presents a novel approach of using high-resolution multispectral data acquired by an unmanned aerial system (UAS) combined with in situ chemical data to assess water quality parameters at 12 relatively small water bodies located in the Tharsis complex, an abandoned mining area highly affected by acid mine drainage (AMD) pollution. The spectral data jointly with water physicochemical data were used to estimate water quality parameters using regression analysis. Parameters including pH, ORP, EC, Al, Cu, Fe, Mn, S, Si, and Zn were estimated with high accuracy levels while Ba, Ca, and Mg showed low accuracy.

Keywords: Acid Mine Drainage, Abandoned Mine, Water Monitoring, Drone, Low Altitude Remote Sensing, Multispectral Sensor

Introduction

AMD is the main environmental pollution problem associated with coal and metal-bearing mineral mining. It is of international concern due to the difficulty in avoiding its formation and its long-lasting nature; it can occur indefinitely even after mining operations have ceased (Qian and Li 2019). Thus, there is a need to develop monitoring tools that can be used by the competent environmental agency and companies in charge of the mining concessions. Traditional procedures for water quality monitoring in reservoirs involve in-situ measurements, sampling, and laboratory analysis. Remote sensing provides a powerful alternative tool that is less time-consuming and provides spatial and temporal information to monitor water quality changes. However, this approach has been scarcely used to report the water quality status in mining areas. Water bodies associated with AMD have a complex composition that requires quantifying a wide range of parameters and few studies have addressed this issue by applying quantitative modeling of hydrochemical concentrations (Tesfamichael and Ndlovu 2018; Modiegi *et al*. 2020). Most recently, UAS-based hyperspectral data have been successfully used to monitor acidic water, generating high-resolution hydrogeochemical maps (Flores *et al*. 2021). Indeed, UAS is becoming increasingly popular in environmental monitoring due to its acquisition flexibility, high spatial and temporal resolution achieved, and the possibility of acquiring data not affected by cloud cover.

In this context, the calibration of empirical models is proposed through regression analysis to predict water quality parameters using in situ physicochemical parameters and spectral reflectance values obtained by a commercial sensor, the Micasense RedEdge-MX Dual. This sensor has already been tested for many purposes such as crop mapping, forestry, minerals mapping and land cover analysis. The Iberian Pyrite Belt (IPB), which hosts one of the largest concentrations of massive sulfides on Earth and is well-known for its mining tradition and extensive AMD environmental impacts, was selected as the study area. This work is

intended to implement an easily reproducible tool that can be used to monitor water bodies with different complex compositions in mine-affected zones.

Methods

Field campaigns were carried out during July and October 2020, consisting of simultaneous flight surveys with the Micasense RedEdge-MX Dual sensor onboard a UAS and in situ physicochemical data acquisition. The sampling sites involved 12 different water bodies containing acid and non-acid waters located in two abandoned mining sites in the Tharsis complex, in the IPB (fig. 1), the Tharsis Mine (fig. 1 A) and the Lagunazo Mine (fig. 1 B).

Field physicochemical parameters such as pH, electrical conductivity (EC), and oxidation-reduction potential (ORP) were measured at each sampling point (yellow circles, fig. 1) with a CrisonMM40 þ multimeter, previously calibrated with certified solutions. Water samples were collected in high-density polyethylene (HDPE) bottles previously washed with a solution of 10% HNO^3, filtered immediately after sampling through a 0.45 mm pore size cellulose nitrate membrane, and acidified to pH < 2 with HNO^3. The samples were analyzed by inductively coupled plasma-atomic emission spectroscopy (ICP-AES; Perkin-Elmer® Optima 3200 RL) for major elements determination (Al, Ba, Ca, Cu, Fe, K, Mg, Mn, Na, P, S, Si, Sr, and Zn) at the Institute of Environment Assessment and Water Research (IDAEA-CSIC, Barcelona). Sierra Bullones (SB, fig. 1A) was not water sampled due to its inaccessibility. However, as it is connected underground to Filón Norte, and since their chemical properties have been shown to be similar (González *et al.* 2018), they were considered the same for this study.

The flight surveys were performed with the multispectral sensor Micasense RedEdge-MX Dual Camera onboard a DJI Matrice 210 V2 RTK. For all the missions, the height was set at 120 m AGL (above ground level) altitude to ensure a ground sample distance (GSD) of 8 cm/pixel, the overlapping was set at 80% frontal and 75% side image overlap, the grid was simple, and the speed was set at 10 m/s. It is assumed that all the multispectral imagery is in the nadir position due to the location of the camera. The Downwelling light sensor (DLS 2) and MicaSense's calibrated reflectance panel (CRP) were used in all the flights. The multispectral images were processed using the Pix4D mapper Structure from Motion (SfM) software. To perform the extraction of the spectral signature of each of the water bodies from the multispectral imagery, the centroid of the water body shape was extracted and the mean reflectance value of all the pixel values included in a circular buffer of 3 m around each centroid point was estimated using the zonal statistics plugin of QGIS 3.10.7.

The collected data were divided into two subsets, the model calibration dataset (70%) and the validation dataset (30%). The waterbodies flights from July (EP) and from October (FC, FN, EG, EL, ML, SB, LLA, and LLB) were used for the model calibration, while the validation dataset consisted of the Th18, EG, and FS from July and LLC from

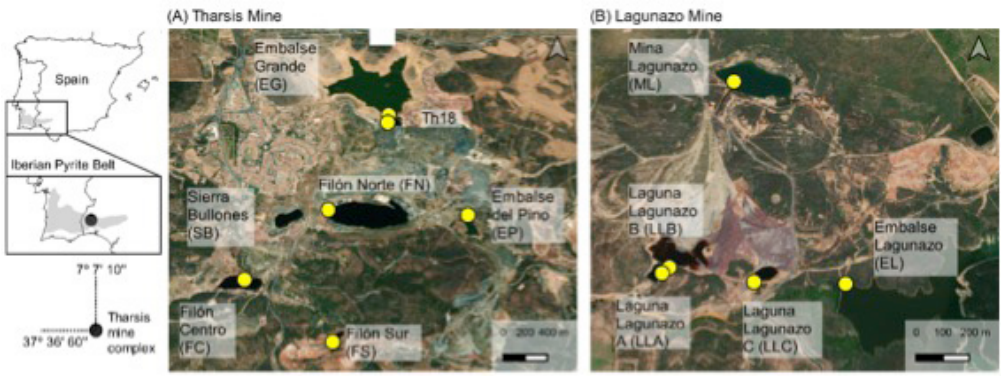

Figure 1 *Location map of the sampling sites. Yellow circles indicate the water samples collection.*

Table 1 List of Spectral Band Combinations (SBCs) tested in this study.

Algorithm	Band math	Reference
A1 (NDVI)[1]	(NIR-R650)/(NIR+R650)	(Rouse and Space 1978)
A2 (NDWI)[2]	(g560-NIR)/(g560+NIR)	(McFeeters 1996)
A3	r650/re705	Simple ratio
A4	re705/nir	Simple ratio
A5	(g531/g560)*r650	Three-band algorithm
A6	(re705/re717)*r668	Three-band algorithm
A7	(g560*g531)/b475	Three-band algorithm
A8	(g560-g531)/b475	Three-band algorithm
A9	(g560/g531)*b475	Three-band algorithm
A10	(re705/r650)*g531	Three-band algorithm

[1]Normalized difference vegetation index (NDVI); [2]Normalized difference water index (NDWI)

October (the only data for LLC presented). To determine which spectral bands and/or spectral band combinations (SBCs) were the best predictor for each chemical parameter, a correlation analysis was carried out between the raw and the ln-transformed chemical data versus the mean reflectance values of the spectral bands and the tested SBCs. Then, using the model calibration dataset, empirical models were constructed relating the mean reflectance values of the significant bands and/or the SBCs as the predictable variable with the chemical data and ln-transformed chemical data as the dependent variable. A total of 17 water quality parameters (pH, ORP, EC, Al, Ba, Ca, Cu, Fe, K, Mg, Mg, Na, P, S, Si, Sr, and Zn), the 10 sensor's spectral bands and 10 SBCs (table 1) were considered in this study. To ensure reliable models, simple linear regression (SLR) and multiple linear regression (MLR) using a stepwise selection method were tested. The prediction quality of the models was assessed on the validation dataset and the performance metric statistics calculated were: normalized root mean square error (RMSE%), mean absolute percentage error (MAPE), Mean absolute error (MAE), bias, and coefficient of determination (R^2). The best-fitted models were used to generate spatial distribution maps.

Results

According to the in situ and laboratory measurements of the water quality parameters, which are representative of the water surface layer's composition at the time of sampling, the waterbodies selected showed a wide range of water compositions, from circumneutral to extremely acidic pH (0.01 - 7.76) and from low to high metal-enriched solutions (e.g., 0.04 - 4795.70 mg L^{-1} of Al, 0.04 - 318.67 mg L^{-1} of Mn, 0.06 - 2011.35 mg L^{-1} of Zn). Concerning the pH value, EP, FC, FN, FS, SB, ML, TH18, LLA, LLB, LLC correspond to acid waters (pH 0.01 - 3.81) while EG and EL are non-acid water (pH 6.62 - 7.76).

Apart from the physicochemical composition, the water bodies can be grouped based on their colors, which mainly depends on the contents of organic matter, algae, suspended particulates, and nutrients. Thus, while EG and ML have shades of brown or green, EP and ML present greenish-blue color, and the rest of the acid water bodies have a dark reddish-brown color. The Fe^{+2} ions give water a greenish color and when Fe^{+3} ions are more abundant, they are responsible for giving the water the intense red color (Schroeter and GläÄer 2011; Riaza *et al.* 2014; Davies and Calvin 2017; Flores *et al.* 2021). Although in this study, the Fe speciation was not quantified, the ORP values of about 500 mV found in all the dark reddish-brown acid waters indicate oxidized aqueous environments and the prevalence of ferric iron (Flores *et al.* 2021). In contrast, the redox potential in the greenish-blue acidic water bodies (i.e. EP and ML) was about 300 mV, suggesting that Fe^{+2} ions prevail over Fe^{+3}. The dissolved iron composition in the dark reddish-brown acid waters was between 352.37 (TH18) and 68940.00 (LLB) mg L^{-1}, while for the rest of the water bodies it ranged between 0.05 (EL) - 0.99 (ML) mg L^{-1}.

The elements K, P, Na, and Sr did not show a significant correlation with any spectral band or SBC, and they were discarded from the model calibration. The parameters correlated to a band or SBC with a Pearson coefficient exceeding 0.8 were used to construct different SLR and MLR models. Among the various regression representations with the validation dataset, the models having the best performance metric statistics were selected as the final model to predict the spatial distribution of water parameters (table 2). The lowest values of RMSE% registered by the selected models (RMSE% = 4 for ln(ORP) and RMSE% = 12 for ln(Ca) reflected the good prediction capability of the models. Ln(Al), ln(Ca), and ln(Mn) showed a tendency to underestimate the observed values (bias ranging from -0.11 to -0.73), while the rest of the models tended to overestimate the real values, giving a positive bias (from 0.0 to 0.75). Even though ln(Ca) showed good results in the model calibration and the validation metrics were better than other models, the R^2 showed poor correlation between the observed and modeled values (R^2 = 0.56). Due to the low accuracy in the prediction, ln(Ca), Ba, and ln(Mg) were dismissed ($0.42<R^2<0.70$) (table 2). The rest of the models presented good fitness between modeled and observed values with R^2 values between 0.81 and 0.99, showing the robust relationships found between the spectral and physicochemical data. Ln(Si) had the best correlation between modeled and observed values of all, confirming the high R^2 (0.99) and low RMSE% value (18%, table 2).

Few studies have quantitatively analyzed water bodies in mining environments applying remotely-sensed spectra to compare with the above findings. Recently Flores *et al.* (2021) applied UAS-hyperspectral imaging to map pH, redox, Al, and Fe concentration in the confluence between the Odiel River and the Tintillo River (Iberian Pyrite Belt, Huelva province). They applied a supervised random forest regression approach, obtaining R^2 values of 0.73, 0.82, 0.68, and 0.66, respectively, with the pixels only used as the validation set. Instead, our study found R^2 values of 0.98 (pH), 0.85 (ORP), 0.98 (Al), and 0.94 (Fe). Although our values are higher, the prediction of water quality parameters in a rapidly changing environment, such as a river, is a more challenging task since the spectral response of water can be affected by the water depth and dynamics. By contrast, the spectral response of a deep and steady water body is expected to be homogeneous, giving higher performance metrics. Schroeter and GläÄer (2011) characterized some water quality parameters in lignite mining lakes, among them pH and Fe. They used a bivariate correlation between Landsat TM satellite data and the chemical analysis, but no models were developed. The highest correlation was found between the red band and Fe (r=0.645), while pH and the red band were poorly correlated (r

Table 2 *Performance metric statistics for the best-fitted water quality parameters.*

WQP	Regression	Candidate model	RMSE%	MAPE	bias	R^2
pH	MLR	2.046-5.674*A1+56.413*A7	15%	0.19	0.31	0.98
ln(ORP)	SLR	6.397-25.657*A7	4%	0.03	0.16	0.85
ln(EC)	SLR	3.633-138.510*g531	32%	0.04	0.09	0.87
ln(Al)	MLR	6.230-142.199*A7+6.751*A1	14%	-0.04	-0.17	0.98
Ba	SLR	0.003+0.818*g560	57%	0.97	0.00	0.42
ln(Ca)	SLR	6.223-81.167*g531	12%	0.10	-0.11	0.56
ln(Cu)	SLR	4.957-170.154*A7	57%	0.17	1.09	0.92
ln(Fe)	SLR	9.489-275.345*A7	25%	0.01	0.75	0.94
ln(Mg)	SLR	6.606-114.772*g531	14%	0.10	0.13	0.7
ln(Mn)	SLR	4.865-135.763*A7	75%	0.056	-0.73	0.95
ln(S)	SLR	9.874-181.692*g560	14%	0.09	0.33	0.81
ln(Si)	SLR	4.336-99.318*A7	18%	-0.19	0.19	0.99
ln(Zn)	SLR	6.259-184.166*A7	40%	0.11	0.22	0.83

SLR: Simple linear regression; MLR: Multiple linear regression

= −0.378). However, in the present study, the correlations found were stronger for both parameters. The pH was correlated with r650 (r = 0.765) but had the highest correlation with A1 (r = −0.92, table 1). The Fe concentration was also correlated with r650 (r = −0.66) having the best correlation with A7 (r = −0.93, tab.). Tesfamichael and Ndlovu (2018) estimated physicochemical parameters from ASTER and Landsat imagery in a gold mining area. Both satellites performed similarly in estimating Eh and Ca ($0.25 < R^2 < 0.36$ and $19 < RMSE\% < 56$). Redox potential was estimated using the blue band (Landsat) and SWIR 6 (ASTER), while Ca was estimated using short-wave infrared band (Landsat) and NIR (ASTER). In this study, ORP was best estimated using A7 ($R^2 = 0.85$ and RMSE% = 4 and ln(Ca) using g531 (R^2 0.56 and RMSE% = 12). In light of this, the models generated with UAS-multispectral imagery were different in the selected bands than in the previous studies but the accuracy of the obtained models was higher, enhancing the value of low altitude remote sensing at a local scale.

Finally, to show the real potential of UAS for monitoring acidic water and predicting water quality parameters, the empirical relationships obtained were extended to the LLC validation dataset, producing spatial distribution maps for each studied parameter (fig. 2). LLC (fig. 1B) is a small waterbody highly affected by AMD, originated by surficial waters flowing through the waste dump and pyritic waste. The water is stagnant and its composition is expected to be homogeneous throughout the water surface. Nevertheless, fig. 2 shows the surface water intake at the right edge, which has a different composition from the rest of the water body. This water intake was analyzed and presented a pH value of 3.12, EC of 0.415 mS/cm (ln EC=0.415 mS/cm), and ORP of 415 mV (ln ORP=6.028 Mv). These values are coincident with the ones observed in fig. 2.

Conclusion

The current study demonstrates that the application of empirical models to generate spatial distribution maps can be an effective and easily applicable monitoring tool in AMD-affected sites. Moreover, the Micasense RedEdge-MX Dual commercial sensor performed well, predicting several water quality parameters, which is especially valuable for small water bodies that cannot be monitored by satellites due to their low spatial resolution. The results obtained here are intended to contribute to the water

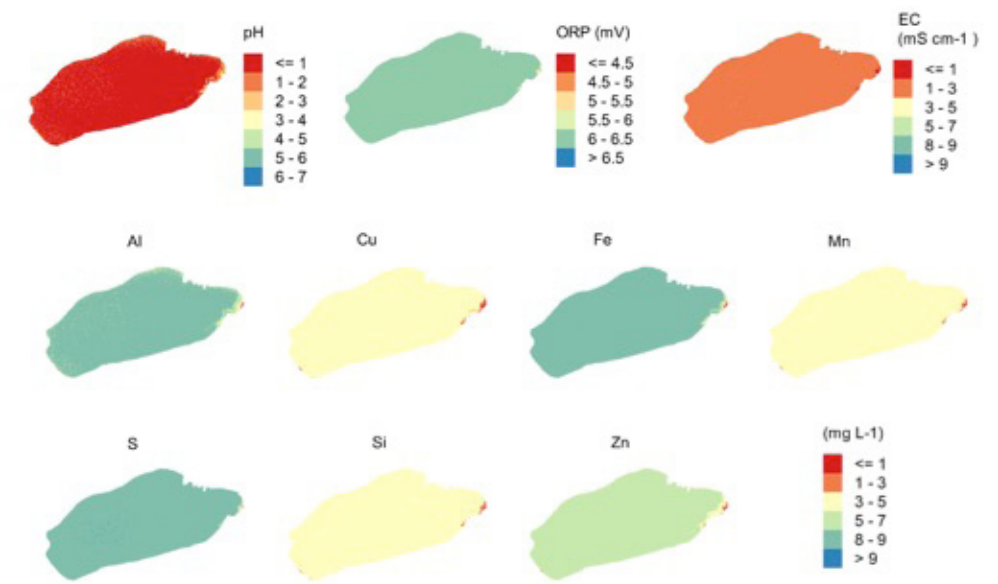

Figure 2 *Spatial distribution maps for the estimated water quality parameters in Laguna Lagunazo C (LLC). All the models were performed with ln-transformed data, except pH.*

resources management and decision-making process during the exploitation and closure phase of mining sites. It is noteworthy that the empirical models are data-driven and are based on in situ water quality measurements. Thus, the models found in this study should be calibrated for application in other locations, water type, and/or season. For instance, if in a mining area is neutral and alkaline metalliferous drainage rather than just AMD, this must be considered during the calibration phase. Further studies should be done to investigate the influence of seasonal variability on the reflectance values of the water bodies due to changes in the dissolved concentrations of pollutants.

Acknowledgements

This study was supported in part by the Erasmus Mundus Joint Master Degree (EMJMD) in Water and Coastal Management (WACOMA) with the contribution of the Erasmus+ Programme of the European Union. This work is also supported by Plan Andaluz de Investigación RNM 166 Environmental radioactivity research group (LB). UAS equipment from University of Cádiz Drone Service supported by MINECO infrastructure projects (EQC2018-00446-P and UNCA-2013-1969). M.D. Basallote thanks the Spanish Ministry of Science and Innovation for the Postdoctoral Fellowships granted under application reference IJC2018-035056-I.

References

Davies GE, Calvin WM (2017) Quantifying Iron Concentration in Local and Synthetic Acid Mine Drainage: A New Technique Using Handheld Field Spectrometers. Mine Water Environ 36:299–309. https://doi.org/10.1007/s10230-016-0399-z

Flores H, Lorenz S, Jackisch R, *et al* (2021) UAS-Based Hyperspectral Environmental Monitoring of Acid Mine Drainage Affected Waters. 1–25

González RM, Olías M, Macías F, *et al* (2018) Hydrological characterization and prediction of flood levels of acidic pit lakes in the Tharsis mines, Iberian Pyrite Belt. J Hydrol 566:807–817. https://doi.org/10.1016/j.jhydrol.2018.09.046

McFeeters SK (1996) The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. Remote Sens Environ 25:687–711

Modiegi M, Rampedi IT, Tesfamichael SG (2020) Comparison of multi-source satellite data for quantifying water quality parameters in a mining environment. J Hydrol 591:125322. https://doi.org/10.1016/j.jhydrol.2020.125322

Qian G, Li Y (2019) Acid and Metalliferous Drainage – A Global Environmental Issue. J Min Mech Eng 1–4

Riaza A, Buzzi J, García-Meléndez E, *et al* (2014) Monitoring acidic water in a polluted river with hyperspectral remote sensing (HyMap). Hydrol Sci J 60:1064–1077. https://doi.org/10.1080/02626667.2014.899704

Rouse JW, Space G (1978) Monitoring the vernal advancement of retrogradation of natural vegetation. Third ERTS Symp 1:48–62

Schroeter L, GläÄer C (2011) Analyses and monitoring of lignite mining lakes in Eastern Germany with spectral signatures of Landsat TM satellite data. Int J Coal Geol 86:27–39. https://doi.org/10.1016/j.coal.2011.01.005

Tesfamichael SGU of A and L for quantifying hydrochemical concentrations in abandoned gold mining, Ndlovu A (2018) Utility of ASTER and Landsat for quantifying hydrochemical concentrations in abandoned gold mining. Sci Total Environ 618:1560–1571. https://doi.org/10.1016/j.scitotenv.2017.09.335