Prediction and Mapping of Pb Content in Overbank Sediments Affected by Coal-Mining Using Airborne Hyperspectral Imaging

Jamie-Leigh R. Abrahams[1], Emmanuel John M. Carranza[1]

[1]*Department of Geology, Faculty of Natural and Agricultural Sciences, University of the Free State, Bloemfontein, South Africa*

Abstract

Spectral absorption feature parameters (SAFPs) derived from airborne hyperspectral imaging (HSI) (396.0–2453.5 nm) were used to predict Pb contents in coalfield overbank sediments. The derived SAFPs were associated with goethite (~500 nm) and kaolinite (~1448 nm) in sediments. Sediment Pb contents correlated strongest with goethite-related absorption-depth (r = 0.6). The calibration model had a R^2 = 0.69 and standard error of estimation (SEE) = 3.97, after outlier removal. The validation model had a R^2 = 0.65 and SEE = 3.90. Overall, the results suggest that airborne HSI can complement conventional geochemical methods of detecting Pb contents in overbank sediments.

Keywords: Witbank Coalfield; reflectance spectroscopy; regression modeling

Introduction

Trace metal (TM) contamination of the environment is one of the biggest challenges related to acid mine drainage (AMD). TM analysis of sediments via conventional geochemical methods is often onerous, particularly when undertaken at large scales. Compared to conventional methods, hyperspectral imaging (HSI) sensors offer greater efficiency by rapidly measuring hundreds of contiguous spectral bands (Asadzadeh & de Souza Filho 2016), some of which can be used to predict TM contents in soils and sediments. The prediction of TM contents is based on their adsorption to minerals, such as Fe-oxides and clays in soils (Wu *et al.* 2005). The adsorption of TMs to these minerals can cause variations in the number of O^- and OH sites on the mineral's surface (Zachara & Westall 1999). These variations may cause changes in their spectral absorption feature parameters (SAFPs), namely absorption-peak depth, area, width and asymmetry, which can be linked to TM contents (Choe *et al.* 2008).

The city of Emalahleni is located in the Witbank Coalfield in South Africa, which has been mined for over a century (Hancox & Götz 2014). Thus, many rivers in the area have been affected by AMD-related TM contamination. There are limited studies on the use of airborne HSI in the vicinity of coal mines, none of which have examined overbank sediments. Among the TMs associated with coal mining, Pb is considered one of the TMs that are most hazardous to ecological and human health. Pb accumulation in the study area is linked largely to AMD related to coal mining and industrial waste (Abrahams & Carranza 2025). Thus, there is a need to efficiently monitor Pb content in the environment. This study therefore endeavoured to (i) derive SAFPs from airborne HSI for use as predictors of Pb content in overbank sediments along the Blesbokspruit River in Emalahleni, (ii) model Pb content in the study site by applying regression analysis to the airborne HSI data using GIS software and (iii) evaluate the model's predictive performance using the goodness of fit (R2) and standard error of estimation (SEE).

Methods

Overbank sediment samples

Twelve overbank sediment samples were collected at six different locations (i.e., two samples spaced approx. 5 m apart at each location), covering approx. 6 km of the Blesbokspruit River (Fig. 1). Samples were collected in this way because the spatial trends of overbank deposition can be highly variable (Simm & Walling 1998). Sample size was limited largely due to water-saturated sediments in a wetland along the stream (i.e., because high soil moisture contents typically cause interferences in spectral data) (Wu *et al.* 2005). For mineralogical analysis, air-dried samples were crushed and milled to < 75 μm and subjected to X-ray diffraction (XRD). For chemical analysis, air-dried samples were sieved (< 63 μm), treated with reverse aqua regia (3 HNO_3 : 1 HCl), which does not digest the silicate fraction (Petrović *et al.* 2022), and microwave digestion before analysis via inductively coupled plasma atomic emission spectroscopy/mass spectrometry (ICP-AES/MS).

Airborne HSI data

Airborne HSI data were collected by the Council for Geoscience of South Africa over the study site via an airplane (Fig. 1b). The flight altitude was 2.35 km with solar zenith angle of 29.4° and solar azimuth angle of 80.9°. The spatial resolution was 1 m by 1 m, with 360 bands from 396.0 nm to 2453.5 nm wavelength range. The spectra were smoothed using Savitzky-Golay filtering to reduce noise and CO_2 mitigation was applied. The HSI data were corrected for atmospheric water by applying the widely used ATCOR4 model and transformed using continuum removal analysis (Choe *et al.* 2008). Variations in absorption-band position, depth(D), width

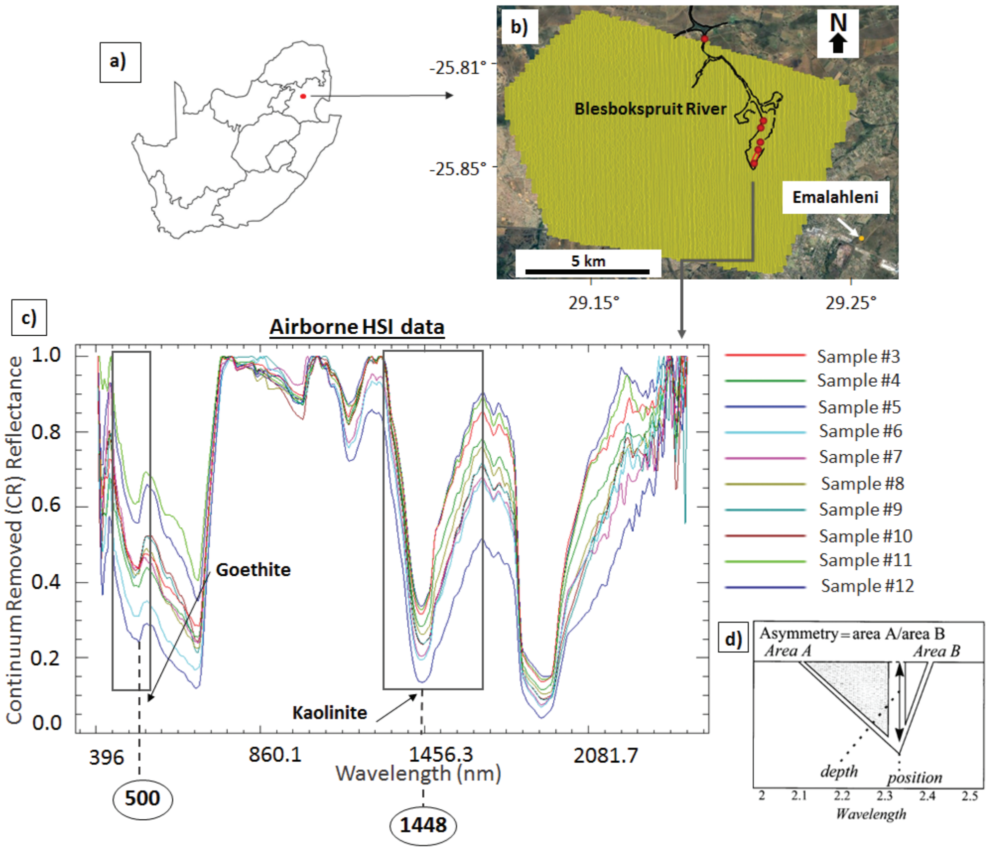

Figure 1 The study site location in the Mpumalanga Province, South Africa, b) airborne HSI coverage over the Blesbokspruit River area, c) continuum-removed reflectance data vs wavelength, and d) D and S parameters (after Van der Meer 1999). The W, which is not illustrated, is calculated as: Area A + Area B / 2D.

(W) and asymmetry (S), were derived from the most distinct SAFs at approx. 500 nm, which is associated with surface hydroxyl groups on goethite (Wu *et al*. 2005) and at approx. 1440 nm, which is associated with surface hydroxyl groups on kaolinite (Van der Meer 2004). The D, W and S were calculated according to Fig. 1d.

Statistical treatment and geochemical mapping

Pb data were *ln*-transformed to ensure that the data approximate a normal distribution, prior to correlation and regression analysis. Spearman's rank correlation analysis was performed because of the small sample size ($n = 10$) and the uncertainty quantified based on statistical significance. The minimum number of observations necessary for a single explanatory variable is $n = 10$ (Van Voorhis & Morgan 2007). Thus, relationships between Pb contents and the HSI data were modeled using univariate regression analysis. Although predictive models typically perform better using a large number of samples, models based on smaller datasets can still provide guidance for more in-depth environmental analyses (Hernandez *et al*. 2006). To calculate 'predicted' Pb contents, the regression equation generated by the best calibration model was applied and the derived *ln*-contents were back-transformed to compare with the measured Pb contents. Regression models were evaluated in terms of the R^2, SEE, root mean squared error of prediction (RMSEP) calculated using leave-one-out cross-validation (LOO-CV) and the relative model stability assessed using the Chow statistic. For the predictive map, HSI data were subjected to NDVI (normalized difference vegetation index) analysis (Cherlinka 2019) and the regression equation from the best calibration model applied to bare soils using the raster calculator in QGIS (QGIS.org 2023). Predictions were validated using pXRF (portable x-ray fluorescence) measurements collected *in-situ*.

Results and Discussion

Mineralogy and geochemistry

Overbank sediments in the study site comprised mainly quartz (up to 95%) and kaolinite (up to 5%). Thus, the overbank sediment samples had a largely sandy texture, which is supported by the findings of Bell *et al*. (2002). Overbank sediments contained considerable Al (median = 4.47%) and Fe (median = 2.57%), with minor Si (median = 0.346%). Lead contents in the study site's overbank sediments (median = 25.2 mg/kg) appeared relatively consistent with those in stream sediments in the Witbank Coalfield (mean = 24.9 mg/kg) (Bell *et al*. 2002), the Baixo Jacuí coal mining region in Brazil (median = 27 mg/kg) (Teixeira *et al*. 2001) and the Jaintia Hills coal deposit in India (mean = 27.5 mg/kg) (Sahoo *et al*. 2017).

Correlations, regression analysis and geochemical mapping

Table 1 shows the correlations between *ln*-transformed Pb contents and the airborne-derived SAFPs. Iron oxide and clay minerals typically adsorb Pb in soils and sediments (Moreno *et al*. 2006), thus, providing support for the trends observed in subsequent correlation and regression analysis. Lead contents had the strongest correlation ($r = 0.6$) with Depth500 (the goethite-related D),

Table 1 *Correlations (r) between ln-transformed Pb contents and airborne-derived SAFPs. Also shown is the goodness-of-fit (R^2) and the standard error of estimation (SEE) of the calibration models.*

Predictor	*r*	R^2	SEE
Depth500	0.6	0.39	0.26
Asym500	-0.4	0.30	0.28
Width500	-0.5	0.20	0.30
Depth1448	0.5	0.25	0.29
Asym1448	0.5	0.27	0.29
Width1448	-0.5	0.29	0.28

although it was not statistically significant. Strong correlation with goethite is consistent with findings by Covelo *et al.* (2007) that Fe-oxide adsorbed and retained greater concentrations of Pb, compared to kaolinite. Results are also consistent with the findings of Kemper and Sommer (2002) that Pb contents in soils could be estimated using correlations with iron oxide contents and the findings of Zhao *et al.* (2022) that spectral bands at approx. 500 nm had strong correlation with Pb contents in soils. Stronger correlation with the goethite- related D, compared to the W and S, is likely because the D requires a simpler calculation compared to the W and S and, thus, calculations for W and S may be prone to greater error (Van der Meer 2004).

The R^2 and SEE of the calibration regression models were consistent with the correlation analysis and showed that the best predictive model was obtained for Depth500. Based on the Chow statistic (0.095), the LOO-CV method yielded stable (i.e., $p > 0.05$) regression coefficients for airborne-predicted Pb contents. When the RMSEP (6.6 mg/kg) was evaluated in terms of the concentration range of Pb, it represented ≈20% of the concentration range. Fig. 2 shows the measured vs airborne-predicted Pb contents in overbank sediments in the study site. In general, the airborne-derived SAFPs showed a tendency to 'under'-predict higher Pb concentrations and slightly 'over'-predict lower Pb concentrations (Fig. 2a). This is consistent with the findings of Tan *et al.* (2020, 2021). The regression line and R^2 (red dotted line in Fig. 2a) were notably influenced by the high concentration outliers. Outliers (red dots) were detected using the median +- median absolute deviation (MAD) (Reimann *et al.* 2005) and were removed. Following outlier removal, the R^2 improved markedly from 0.28 to 0.69 (blue dotted line in Fig. 2a), which is consistent with Kemper & Sommer (2003), who noted a similar improvement to the R^2 when outliers were removed.

Fig. 3 shows airborne-predicted Pb contents in overbank sediments in the study site. Airborne-predicted Pb contents was calculated using the following regression equation: $y = 2.96x + 2.40$ (where x is Depth500). According to Fig. 3, the overbank sediments appeared largely dominated by Pb concentrations of 7–10 mg/kg and higher.

Predictions according to the validation model (Fig. 2b) were generally lower than measured pXRF contents. Kemper & Sommer (2003) suggested that the 'under'-prediction of higher metal concentrations is likely attributed to averaging across the area of one pixel, which can reduce higher concentrations to lower concentrations and incorporate it into the final measurement. Although the predictions were lower than the measured Pb contents and the goodness-of-fit of the validation model was only moderate ($R^2 = 0.65$), Choe *et al.* (2009) suggested that this technique is still useful because it provides a simple and rapid approximation of TM contents prior to performing more precise geochemical analysis.

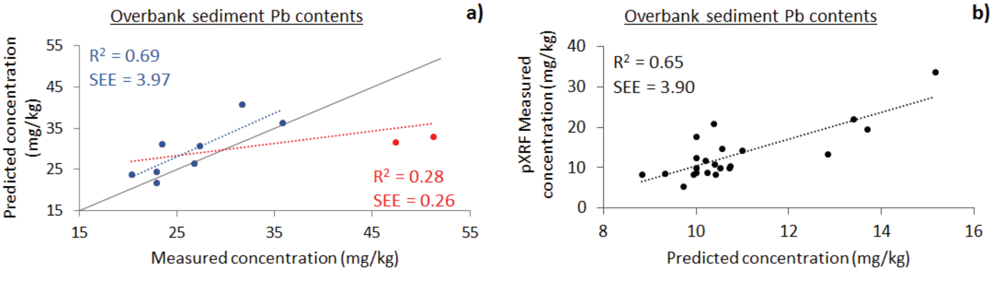

***Figure 2** a) Calibration models showing airborne HSI-predicted Pb content vs measured content in overbank sediments along the Blesbokspruit River, South Africa (the red dotted line represents the model including outliers, the blue dotted line represents the model excluding outliers and the black solid line represents the 1:1 trend line); and b) the validation model showing in-situ measured Pb contents vs. airborne HSI-predicted contents (the black dotted line represents the regression line).*

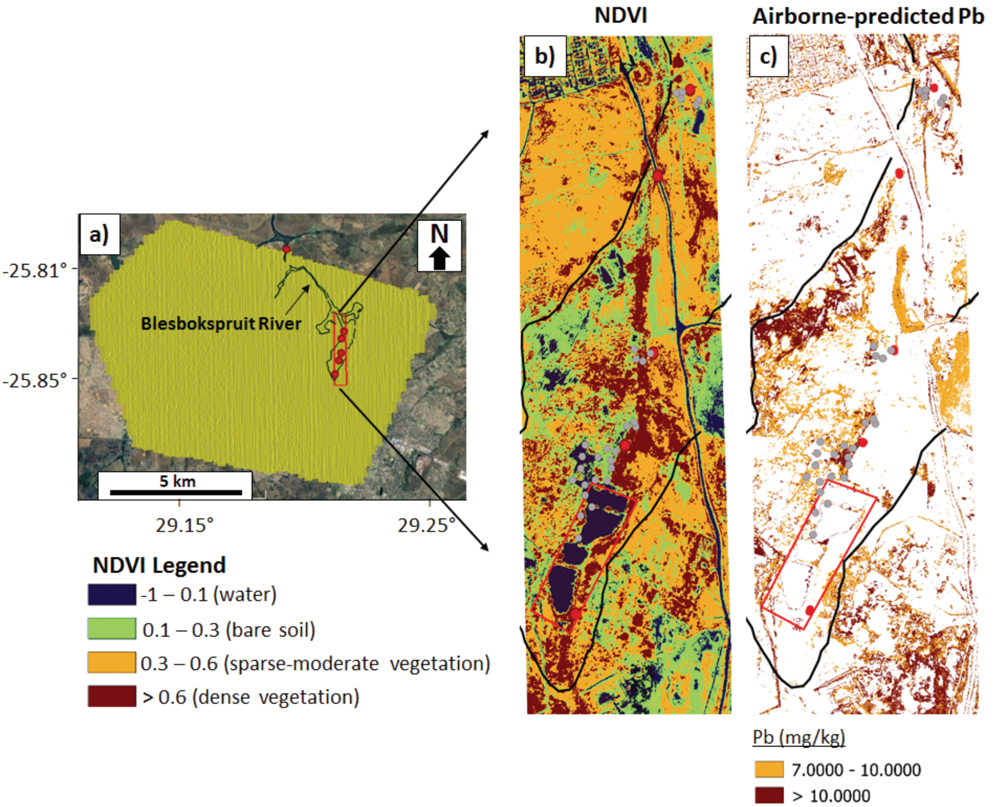

Figure 3 a) Airborne HSI data coverage over the study area, b) NDVI and c) airborne-predicted Pb contents in exposed overbank sediments in the study site. Red dots represent ICP–MS sample sites (used in model calibration) and grey dots represent pXRF sample sites (used in model validation). The red rectangle in b) and c) highlights acid ponds in the study site.

Conclusions

This study found that the most prominent TM-predicting SAFs were associated with goethite (at approx. 500 nm) and kaolinite (at approx. 1448 nm) in overbank sediments. Pb data correlated strongest with Depth500 (goethite-related D). The calibration model had a R^2 of 0.28 and SEE of 0.26. Outlier removal markedly improved the R^2 of the calibration model from 0.28 to 0.69. The corresponding validation model had a R^2 of 0.65 and SEE of 3.90. The results suggest that Depth500 may serve as a proxy for Pb contents in overbank sediments. Due to the small dataset (n = 10) and strong influence of outliers on the regression analysis, further research using a larger dataset is necessary to improve the calibration models. When more data are added and the models are improved, airborne HSI may be useful as a rapid screening method for detecting TM contents related to coal mining in overbank sediments prior to more intensive data collection and geochemical analysis.

Acknowledgements

The authors thank the South African National Space Agency (SANSA) for sponsoring this research project.

References

Abrahams, J-LR, Carranza EJM (2025) Assessment of trace metal contamination in overbank sediments of the Witbank Coalfield, South Africa. Environ Earth Sci 84:153

Asadzadeh S, de Souza Filho CR (2016) A review on spectral processing methods for geological remote sensing. Int J App Earth Obs and Geoinf 47:69–90

Bell FG, Hälbich TFJ, Bullock SET (2002) The effects of acid mine drainage from an old mine in the Witbank Coalfield, South Africa. Q J Eng Geol Hydrogeol 35:265–278

Cherlinka V (2019) NDVI FAQ: All You Need To Know About Index, EOS Data Analytics. Retrieved July 2023 from https://eos.com/blog/ndvi-faq-all-you-need-to-know-about-ndvi/

Choe E, van der Meer F, van Ruitenbeek F, van der Werff H, de Smeth B, Kim KW (2008) Mapping of heavy metal pollution in stream sediments using combined geochemistry, field spectroscopy, and hyperspectral remote sensing: A case study of the Rodalquilar mining area, SE Spain. Remote Sens Environ 112(7):3222–3233

Choe E, Kim KW, Bang S, Yoon IH, Lee KY (2009) Qualitative analysis and mapping of heavy metals in an abandoned Au-Ag mine area using NIR spectroscopy. Environ Geol 58(3):477–482

Covelo EF, Vega FA, Andrade ML (2007) Competitive sorption and desorption of heavy metals by individual soil components. J Hazard Mater 140:308–315

Hancox PJ, Götz AE (2014) South Africa's coalfields - A 2014 perspective. Int J Coal Geol 132:170–254

Hernandez PA, Graham CH, Master LL, Albert DL (2006) The effect of sample size and species characteristics on performance of different species distribution modeling methods. Ecography 29(5):773–785

Kemper T, Sommer S (2002) Estimate of heavy metal contamination in soils after a mining accident using reflectance spectroscopy. Environ Sci Technol 36:2742–2747

Kemper T, Sommer S (2003) Mapping and monitoring of residual heavy metal contamination and acidification risk after the Aznalcóllar mining accident (Andalusia, Spain) using field and airborne hyperspectral data. In: Habermeyer, M., Müller, A., Holzwarth, S. (Eds.) Proceedings of 3rd EARSEL Imaging Spectroscopy, Herrsching, Germany, CD-ROM ISBN 2-908885-56-5, 333–343

Moreno AM, Quintana JR, Pérez L, Parra JG (2006) Factors influencing lead sorption–desorption at variable added metal concentrations in Rhodoxeralfs. Chemosphere 64(5):758–763

Petrović S, Mrmošanin J, Pavlović A, Alagić S, Tošić S, Stojanović G (2022) The Influence of Agricultural Soil Preparation Methods on the Pseudo-Total Element Content Determined By ICP-OES. Studia Universitatis Babes-Bolyai Chemia 67(1):43–60

QGIS.org (2023) QGIS Geographic Information System. QGIS Association. Retrieved August 2023 from http://www.qgis.org

Reimann C, Filzmoser P, Garrett RG (2005) Background and threshold: Critical comparison of methods of determination. Sci Total Environ 346(1–3):1–16

Sahoo PK, Tripathy S, Panigrahi MK, Equeenuddin SM (2017) Anthropogenic contamination and risk assessment of heavy metals in stream sediments influenced by acid mine drainage from a northeast coalfield, India. Bull Eng Geol Environ, 76(2):537–552

Simm DJ, Walling DE (1998) Lateral variability of overbank sedimentation on a Devon flood plain. Hydrol Sci J 43:715–732

Tan K, Ma W, Chen L, Wang H, Du Q, Du P, Yan B, Liu R, Li H (2021) Estimating the distribution trend of soil heavy metals in mining area from HyMap airborne hyperspectral imagery based on ensemble learning. J Hazard Mater 401:123288

Tan K, Wang H, Chen L, Du Q, Du P, Pan C (2020) Estimation of the spatial distribution of heavy metal in agricultural soils using airborne hyperspectral imaging and random forest. J Hazard Mater 382: 120987

Teixeira EC, Ortiz LS, Alves MFCC, Sanchez JCD (2001) Distribution of selected heavy metals in fluvial sediments of the coal mining region of Baixo Jacuí, RS, Brazil. Environ Geol 41(1–2):145–154

Van der Meer F (1999) Can we map swelling clay with remote sensing? Int J App Earth Obs and Geoinf 1:27–35

Van der Meer F (2004) Analysis of spectral absorption features in hyperspectral imagery. Int J App Earth Obs and Geoinf 5(1):55–68

Van Voorhis CW, Morgan BL (2007) Understanding power and rules of thumb for determining sample sizes. Tutorials in Quantitative Methods for Psychology 3(2):43–50

Wu YZ, Chen J, Ji JF, Tian QJ, Wu XM (2005) Feasibility of reflectance spectroscopy for the assessment of soil mercury contamination. Environ Sci Technol 39(3):873–878

Zachara JM, Westall JC (1999) Chemical modeling of ion adsorption in soils. In Sparks DL (Ed), Soil physical chemistry, Boca Raton: CRC Press, p 47–95

Zhao D, Xie D, Yin F, Liu L, Feng J, Ashraf T (2022) Estimation of Pb Content Using Reflectance Spectroscopy in Farmland Soil near Metal Mines, Central China. Remote Sens 14(10):2420