impact International Journals and Publications

© IIJP 2026 | Volume 2, Issue 4, pp. 1266-1286 | ISSN: 2636-4484

SPATIAL DISPERSION PATTERNS OF CHEMICAL ELEMENTS AS PREDICTIVE INDICATORS OF GOLD IN GOLD MINE SPOILS: A CASE STUDY OF GOLD MINES IN NIGER, TARABA AND ZAMFARA STATES, NIGERIA

Bagudu Shehu Ezekiel[1,*], Samuel Shikaa[2], Myina Othniel Mdughga[3], Eunice Adebowale Medjor[4], Felix Egharevba[5] & Weltime Odeguonor Medjor[6]

[1,3,6] Department of Chemical Sciences, Faculty of Science, Taraba State University, Jalingo, Taraba State, Nigeria

[2]Department of Mathematics, Faculty of Science, Taraba State University, Jalingo, Taraba State, Nigeria.

[4]Department of Civil Engineering, Taraba State University, Jalingo, Nigeria.

[5]Department of Chemistry, Ambrose Alli University, Ekpoma, Edo State, Nigeria

** Corresponding author: bagudu.shehuezekiel@gmail.com*

ARTICLE INFO

Article No.: 0717

Accepted Date: 15/10/2026

Published Date: 27/11/2026

Type: Research

ABSTRACT

Gold mine spoils are widespread in Nigeria but have rarely been used as a source of geochemical information for gold exploration. We investigated the spatial dispersion patterns of chemical elements as predictive indicators of gold in the gold mine spoils of Niger, Taraba and Zamfara States. A total of 291 spoil samples (20 from each of 15 Local Government Areas, depth 15 cm) were digested and analysed for 55 elements by ICP-OES, with certified reference materials (OREAS-25 and SARM-36) used for quality control. Concentrations were converted to enrichment factors (EF) using iron as reference element and the upper continental crust as background. Random Forest regression, interpreted with SHapley Additive exPlanations (SHAP), ranked the elements that predict gold in each state; the top 20 predictors of each state were then subjected to stepwise multiple regression to develop the Taraba, Niger and Zamfara models, while the combined dataset (291 samples) was subjected directly to stepwise regression to develop a combined model, TANIZA. The predictors differed among states: Sm, Er, Gd and Dy in Taraba (R^2 = 98.77%; predicted R^2 = 88.35%); W, Si, Hg, Sm and Gd in Niger (94.48%; 81.83%); and Pt, Pd, Sm and Ce in Zamfara (62.45%; 55.72%). The TANIZA model, with Ba, W, B, Hg, La, Si, Tb and Sm, explained 98.55% of the variation in gold (predicted R^2 = 97.49%). The graphical performance counts of the Taraba, Zamfara, Niger and TANIZA models were 85.6%, 87.7%, 98.1% and 98.9% respectively, all above the 50% acceptance criterion. Rare earth elements (Sm, Gd, Er, Dy, Ce, La, Tb), the platinum-group elements (Pt, Pd), W, Si, B, Ba and Hg emerged as strong statistical predictors of gold in Nigerian basement-complex mine spoils and, subject to independent validation, may extend the conventional pathfinder suite.

Keywords: gold exploration; mine spoils; enrichment factor; Random Forest; SHAP; stepwise regression; rare earth elements; Nigeria

© IIJP 2026 | Volume 2, Issue 4, pp. 1266-1286 | ISSN: 2636-4484

Introduction

Geochemical exploration for gold relies on pathfinder elements, that is, elements that are easier to detect, more widely dispersed or less erratically distributed than gold itself and that are associated with it in the primary mineralisation (Rose et al., 1979). In orogenic gold systems the classical suite comprises As, Sb, W, B, Hg and, depending on the deposit, Bi, Te, Mo, Ag, Cu, Pb and Zn (Groves et al., 1998). In deeply weathered basement terrains such as those of northern Nigeria, however, the surface expression of gold depends on weathering, regolith evolution and climate, so that the pathfinder suite proven in one terrain cannot be assumed to apply in another (Webster and Mann, 1984; Butt et al., 2000). Pathfinder suites must therefore be established locally.

Gold mine spoils (the heaps and tailings left by decades of artisanal and small-scale mining) are a largely unused geochemical resource. They sample the ore and its alteration envelope, are abundant and are easy to collect. Their use is complicated, however, by the redistribution of particles by water and wind and by processing contaminants such as mercury, and the multi-element datasets they yield are high-dimensional and intercorrelated, often with few samples relative to the number of variables. Machine-learning methods are well suited to such data. Random Forest (Breiman, 2001), in particular, has been shown to perform as well as or better than neural networks, regression trees and support vector machines in mineral prospectivity modelling (Rodriguez-Galiano et al., 2015) and to work with few prospects and incomplete data (Carranza and Laborte, 2015); it has been used with other machine-learning methods to identify geochemical anomalies (Zuo and Xiong, 2018) and used to predict gold pathfinder elements from stream-sediment geochemistry in central Tanzania (Nunoo et al., 2024). Its main drawback, the opacity of the fitted ensemble, can be addressed with SHapley Additive exPlanations (SHAP), which attribute each prediction to the contributions of individual variables (Lundberg and Lee, 2017).

This paper reports a data-driven identification of the elements that predict gold in the gold mine spoils of Niger, Taraba and Zamfara States, north-central, north-eastern and north-western Nigeria, respectively. Specifically, we (i) determined 55 elements in 291 spoil samples by ICP-OES and converted them to enrichment factors; (ii) ranked the elements by their ability to predict gold using Random Forest with SHAP; (iii) developed state-specific and combined (TANIZA: Taraba–Niger–Zamfara) stepwise regression models; and (iv) assessed the models numerically (R^2, adjusted R^2, predicted R^2) and graphically (performance count).

Study area and sampling

The three states lie on the Precambrian basement complex of Nigeria, locally overlain by Cretaceous–Tertiary sedimentary or volcanic cover, and host gold in quartz veins, shear zones, stockworks and placers. Zamfara State (north-west) is in the Sudano-Sahelian zone and its gold is hosted by the Anka and Maru schist belts, which belong to the same north-western schist-belt system as the Bin Yauri belt of neighbouring Kebbi State; Niger State (north-central) has extensive basement terrain with documented gold in the Minna and Bida areas and in greenstone-belt remnants; Taraba State (north-east) has gold in quartz veins and shear zones of the schist-belt terrain of central and southern Taraba, together with Mambilla Plateau highlands and younger basalts (Obaje, 2009). Five Local Government Areas (LGAs) with a confirmed history of gold mining were selected in each state (Table 1). From each LGA, 20 composite samples were collected from spoil heaps and tailings at a depth of 15 cm: each comprised about five sub-samples taken within a 5-m radius with stainless-steel tools, mixed to give about 1 kg and sealed in acid-washed HDPE bags. The coordinates of each point were recorded with a handheld GPS

receiver, and samples were transported under chain-of-custody documentation. In all, 291 samples were collected; the enrichment-factor dataset used for modelling comprised 100 samples from Taraba, 100 from Zamfara and 91 from Niger State (291 in total).

Figure 1-3: **Location area Maps**

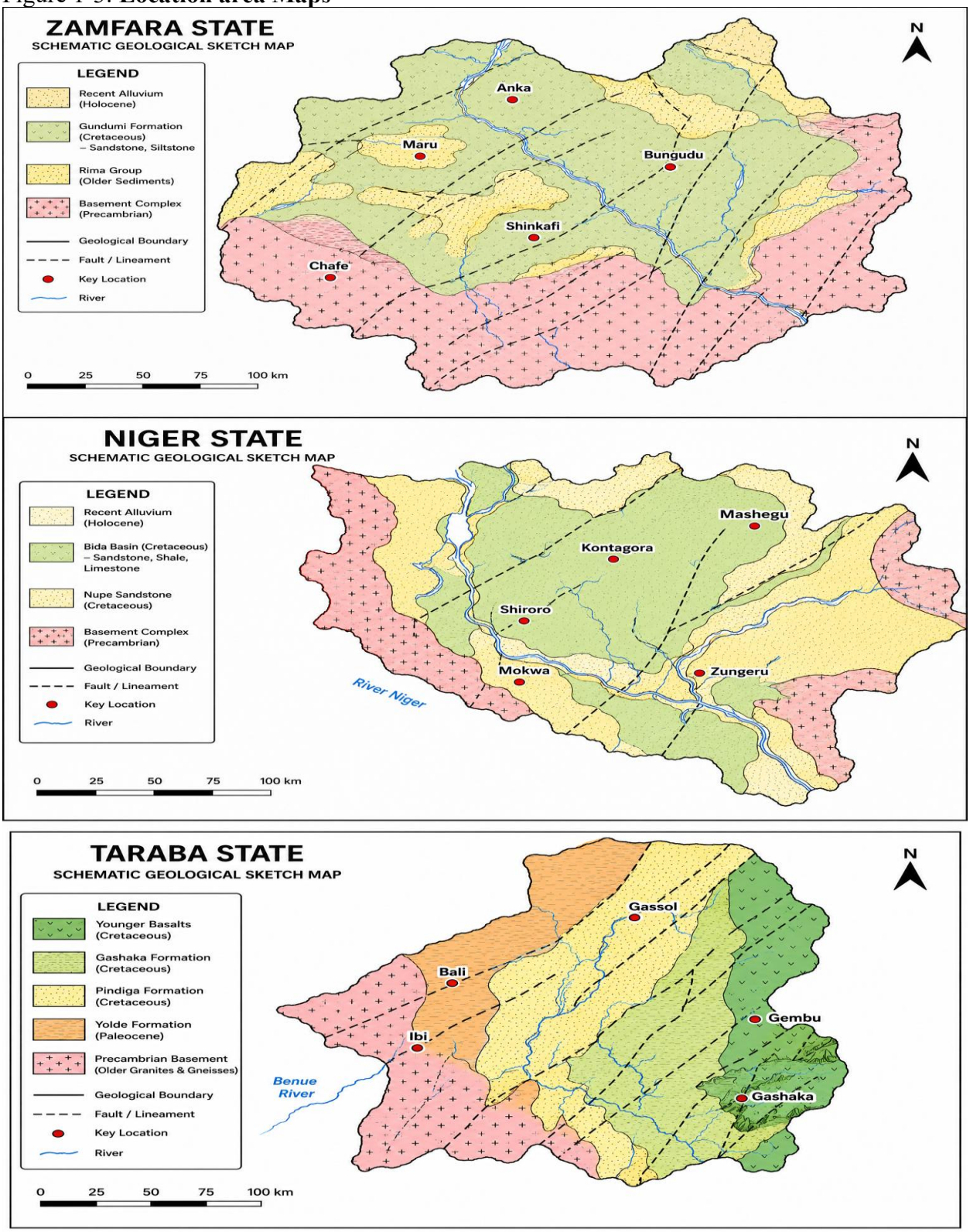

Table 1. Distribution of samples across the study localities.

© IIJP 2026 | Volume 2, Issue 4, pp. 1266-1286 | ISSN: 2636-4484

State	Local Government Areas sampled	LGAs	Samples per LGA	Samples collected	Samples in EF dataset
Taraba	Gashaka, Bali, Gassol, Sardauna (Gembu), Ibi	5	20	100	100
Zamfara	Anka, Maru, Bungudu, Tsafe (Chafe), Shinkafi	5	20	100	100
Niger	Shiroro, Kontagora, Wushishi (Zungeru), Mashegu, Mokwa	5	20	91	91
Total		15		291	291

EF = enrichment factor.

Materials and methods

Sample preparation and ICP-OES analysis

Samples were air-dried at 40 °C for 72 h (a temperature chosen to minimise the loss of mercury), disaggregated, sieved through a 2-mm stainless-steel mesh, homogenised by cone-and-quarter splitting, and about 50 g was pulverised in an agate planetary ball mill to < 75 µm; the mill was cleaned between samples with ethanol, a quartz-sand blank and air purging. A 0.500 g portion was digested in PTFE vessels with 9 mL HNO_3, 3 mL HCl and 2 mL HF by microwave digestion following a modified USEPA Method 3052 (ramp to about 200 °C over 15 min, held 15 min); boric acid was added after cooling to complex residual HF and the solution was made to 50 mL with 2% (v/v) HNO_3. Each digestion batch included a method blank, a certified reference material (OREAS-25 or SARM-36) and a sample duplicate.

Fifty-five elements were determined with a PerkinElmer optima 5300DV ICP-OES Dual view (Axial + Radial using the software Varian ICP-OES (ICP Expert II) at the D.K Olukoya Central Research & Reference Laboratories, University of Lagos, Akoka, Lagos State, Nigeria. The ICP-OES operated at 1,300 W radiofrequency power with plasma, auxiliary and nebuliser argon flows of 15, 1.5 and 0.8 L/min, an uptake rate of 1.5 mL/min and CCD detection (3 s integration, three replicates). The full multi-element suite, including the rare earth elements, the platinum-group elements Pt and Pd and the high-field-strength elements Nb and Ta, was recorded so that every element could be evaluated as a gold predictor (Table 2). Calibration used at least five points prepared from NIST-traceable stock solutions in 2% HNO_3, with $R^2 > 0.9999$; control standards at 10, 50 and 90% of the calibration range were run after every 25 samples and the instrument was recalibrated if any deviated by more than ±10%. Data were accepted if reference-material recoveries were within 85–115% (80–120% for elements of higher natural variability), duplicate RPD was < 20% for analytes above ten times the method detection limit, and method blanks were below the detection limit.

Table 2. Classification of the 55 (other than Fe) elements determined by ICP-OES.

Class	Elements	No.
Transition metals	Ti, V, Cr, Mn, Co, Ni, Cu, Zn, Y, Zr, Nb, Mo, Ru, Pd, Ag, Cd, Ta, W, Re, Ir, Pt, Au, Hg,	24
Lanthanides	La, Ce, Pr, Nd, Sm, Eu, Gd, Tb, Dy, Er, Yb, Lu	12
Alkali metals	Li, Na, K, Rb, Cs	5
Metalloids	B, Si, As, Sb, Te	5
Post-transition metals	Ga, In, Pb, Bi	4
Alkaline earth metals	Be, Mg, Ba	3

Class	Elements	No.
Actinides	Th, U	2
Reactive non-metal	P	1

Enrichment factors

Concentrations were transformed to enrichment factors (EF) to separate crustal contributions from enrichment by mineralisation processes, using Fe as reference element and the average upper continental crust of Taylor and McLennan (1985) as background:

$EF = (C_x / C_{Fe})_{sample} \div (C_x / C_{Fe})_{background}$ (1)

where C_x and C_{Fe} are the concentrations (mg/kg) of element x and of Fe. EF values were calculated for every element in each sample and interpreted with the classes of Sutherland (2000): < 2 minimal, 2–5 moderate, 5–20 significant, 20–40 very high and > 40 extremely high enrichment. The EF approach is widely used although its limitations, notably the dependence on the choice of reference element and background, are recognised (Reimann and de Caritat, 2000, 2005). The EF of Au was the response variable and the EFs of the other elements the predictors.

Random Forest and SHAP

A Random Forest regression (Breiman, 2001) was fitted to each state dataset (n = 100 (Taraba), n = 100 (Zamfara) and n = 91 (Niger) and to the combined dataset (n = 291). Each tree was grown on a bootstrap sample with a random subset of predictors tried at each split (*mtry*), and *mtry* were tuned using out-of-bag error and cross-validation (Kohavi, 1995). Performance was evaluated with R^2 and root-mean-square error and with plots of predicted against observed Au. Predictors were ranked by importance, and their contributions were shown with SHAP summary plots (Lundberg and Lee, 2017), in which elements are ordered from most to least influential, point colour indicates whether the element's EF is high (red) or low (blue), and horizontal position shows whether it raises or lowers the predicted Au. Because Random Forest importance can be distorted by correlated predictors (Strobl et al., 2008), the ranked list was used to screen, not to decide, the final predictors.

Stepwise multiple regression

Multiple linear regression models of the general form

$Au = \beta_0 + \beta_1X_1 + \beta_2X_2 + \ldots + \beta_kX_k + \varepsilon$ (2)

(where Au is the EF of gold, $X_1 \ldots X_k$ the EFs of the retained predictors, β the coefficients and ε the error) were built in Minitab 17 by stepwise regression, in which predictors enter and leave the model according to their statistical significance [see supplementary appendix] until no further change is justified (Draper and Smith, 1998). The Taraba, Niger and Zamfara models were obtained by subjecting the 20 highest-ranked Random Forest predictors of each state to stepwise regression. The combined model, TANIZA (from Taraba, Niger and Zamfara), was obtained by subjecting all variables of the combined dataset (291 samples) directly to stepwise regression, without prior Random Forest selection.

This combined Random Forest model served as a comparison only; the TANIZA regression model was obtained directly from the combined dataset by stepwise regression.

Models were assessed by the standard error of the regression (S), R^2, adjusted R^2 (which penalises non-contributing predictors) and predicted R^2 (which measures how well a model predicts observations not used to build it) (Montgomery et al., 2012).

Assumption tests and graphical performance count

The assumptions of multiple linear regression were tested at $\alpha = 0.05$ for each dataset: linearity (Harvey–Collier test; Harvey and Collier, 1977), homoscedasticity (Breusch–Pagan test; Breusch and Pagan, 1979), normality of residuals (Shapiro–Wilk test; Shapiro and Wilk, 1965)

and independence of errors (Durbin–Watson statistic; Durbin and Watson, 1950). A test was passed when its p-value exceeded 0.05 or, for the Durbin–Watson statistic, when its value lay between 1.5 and 2.5. For each state, the entire cleaned EF dataset (all samples and all 56 elements, with none removed) was used: the Au EF was regressed on the EFs of the other 55 elements, and the four tests were applied to this regression and its ordinary-least-squares residuals; the same four tests were then applied to the residuals of each of the four final stepwise models in Table 3, which are the models reported (Section 4.5). For the combined dataset (291 samples), the three states, which differ greatly in the level and spread of their EFs, were treated as groups in a pooled regression in which each state keeps its own intercept and coefficients and is weighted by the inverse of its residual variance (the TANIZA model itself, with one intercept and common coefficients, was also tested directly, as described above), and the tests were applied to the weighted (standardised) residuals; the Harvey–Collier test used the recursive residuals of the three groups together (126 residuals). The tests were run in a program written with the aid of an AI tool (Claude, Anthropic). The tests were run in Python 3 (NumPy, pandas, SciPy and statsmodels); the Harvey–Collier statistic was computed from recursive residuals coded in NumPy.

The performance of each model was also assessed graphically. For a selected set of samples from each model's dataset (20 samples for each state model and 34 for TANIZA), the predicted Au values from the regression equation were plotted against the actual values in Microsoft Excel, a linear trend line was fitted, and its R^2 was taken as the *performance count* of the model, expressed as a percentage. Because the equations were fitted to these data, the count measures goodness of fit and is not a validation; out-of-sample performance was assessed with the predicted R^2 (Table 3) and with repeated (20 times) 10-fold cross-validation, which gave R^2 values of 88.9%, 76.0%, 55.8% and 97.3% for the Taraba, Niger, Zamfara and TANIZA models respectively. Model performance was evaluated using both graphical comparison and statistical criteria (minitab result), following Moriasi et al (2007), where NSE (Nash-Sutcliffe Efficiency) > 50% indicates satisfactory performance Plots of predicted against observed values are a standard graphical check of regression models (Montgomery et al., 2012; Hastie et al., 2021). Sampling locations were mapped in ArcGIS.

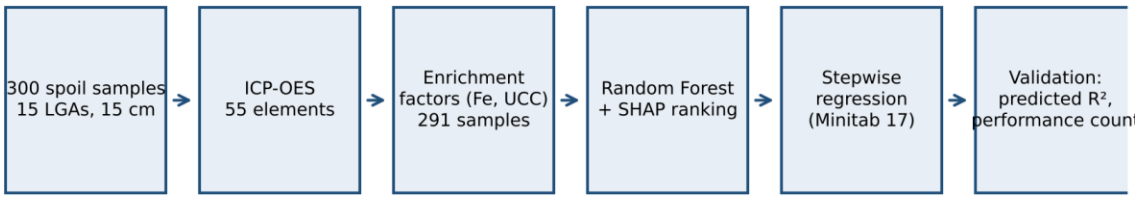

Fig. 1. Analytical and modelling workflow of the study.

Results

Enrichment-factor dataset

Fifty-five elements (23 transition metals, 12 lanthanides, 5 alkali metals, 5 metalloids, 4 post-transition metals, 3 alkaline earth metals, 2 actinides and 1 reactive non-metal; Table 2) were included in the EF dataset, which comprised 91 samples from Niger State and 100 each from Taraba and Zamfara States (291 samples).

Random Forest and SHAP ranking

The Random Forest models of the three states reproduced the observed Au values (Fig. 2a–c), and the SHAP summary plots (Figs. 3 and 4) ranked the elements from most to least influential. The five elements most strongly associated with gold differed from state to state: Ta, Lu, Cr, Sm and Nd in Taraba State; W, Cu, Pr, Y and Er in Niger State; and Pt, Li, Re, Sm and Pd in Zamfara State. Only samarium appeared among the five highest-ranked predictors of more than one state. In the SHAP plots, high values of most of the leading elements pushed the predicted gold value upward. When the EFs of the three states were combined, the Random Forest model reproduced the observed Au values with a cross-validated R^2 of 0.689 (Fig. 2d), and the SHAP analysis (Fig. 4d) ranked Mo, Lu, Ta, Ga, Pt and W as the most influential elements, followed by Mg, Zr, Pr, Rb, Nb, Zn, Pb, La, Cd, Ag, Th, Ce, Ba and Re. This combined Random Forest model served as a comparison only; the TANIZA regression model was obtained directly from the combined dataset. All the predictors retained in the final Taraba (Sm, Er, Gd, Dy), Niger (W, Si, Hg, Sm, Gd) and Zamfara (Pt, Pd, Sm, Ce) models appear among the elements ranked by SHAP (Figs. 3 and 4).

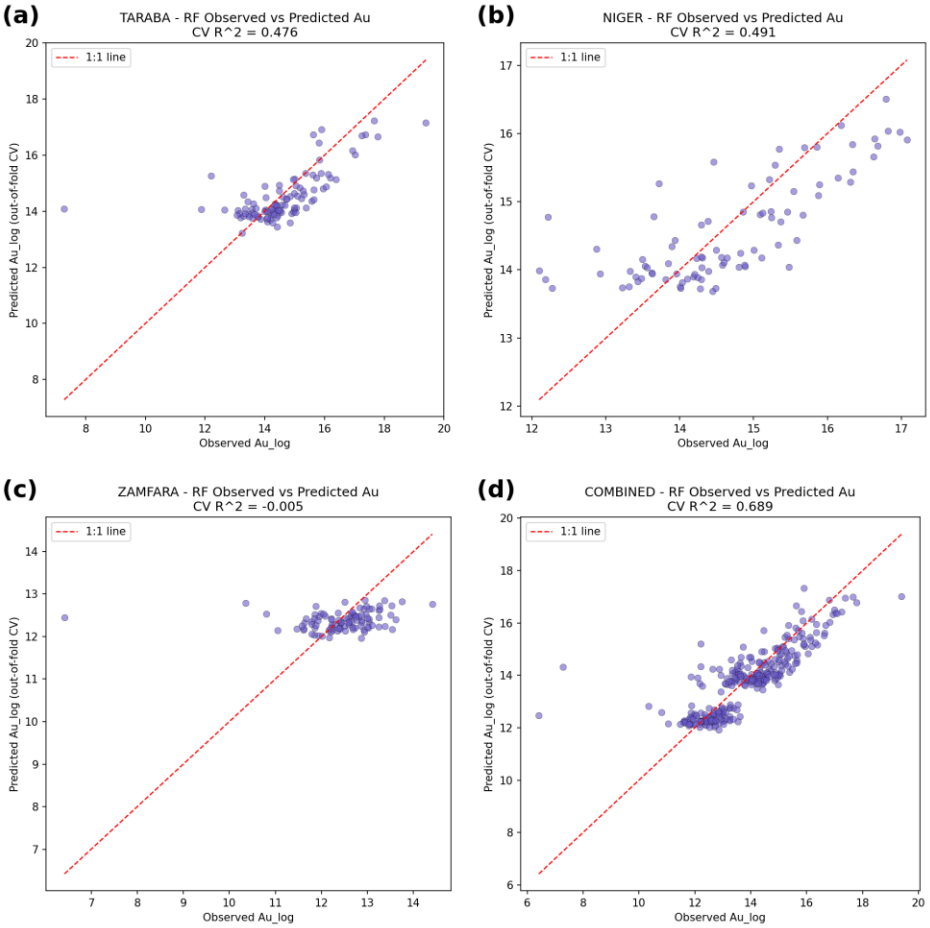

Fig. 2. Random Forest predicted versus observed Au (EF) values: (a) Taraba; (b) Niger; (c) Zamfara; (d) combined dataset of the three states.

© IIJP 2026 | Volume 2, Issue 4, pp. 1266-1286 | ISSN: 2636-4484

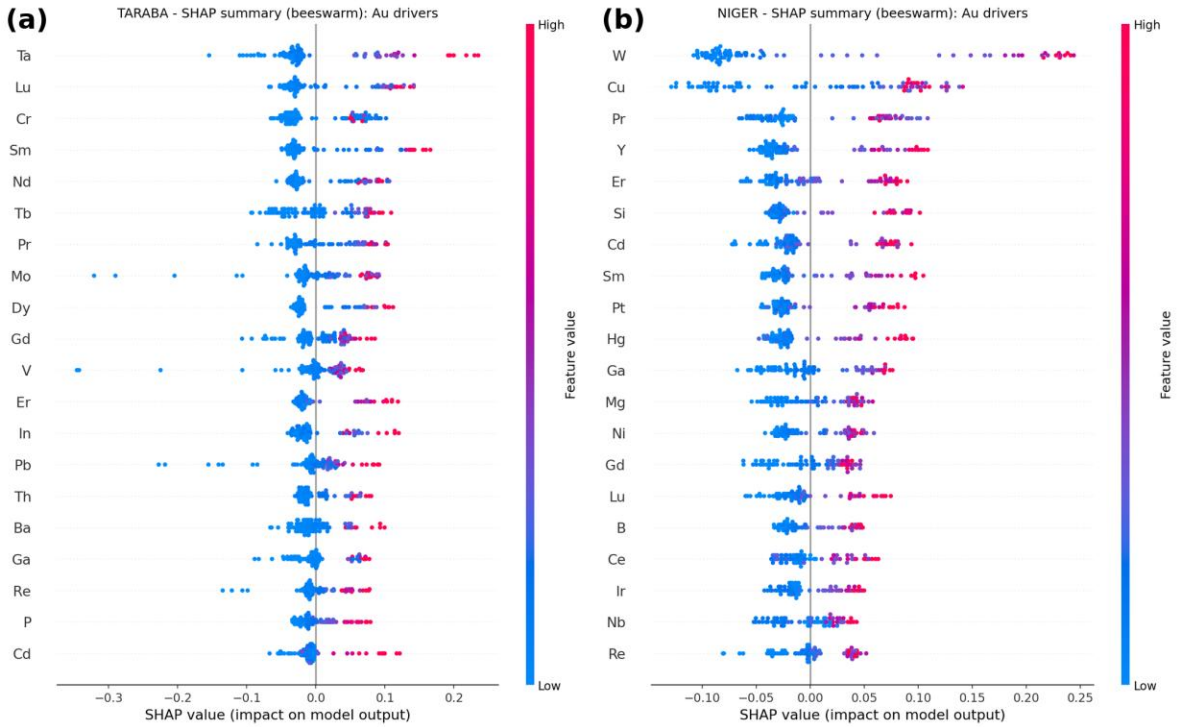

Fig. 3. SHAP summary plots of the impact of each element on the predicted Au value: (a) Taraba State; (b) Niger State. Elements are ordered from most (top) to least influential; colour shows the element's EF (red high, blue low) and horizontal position its effect on the predicted Au.

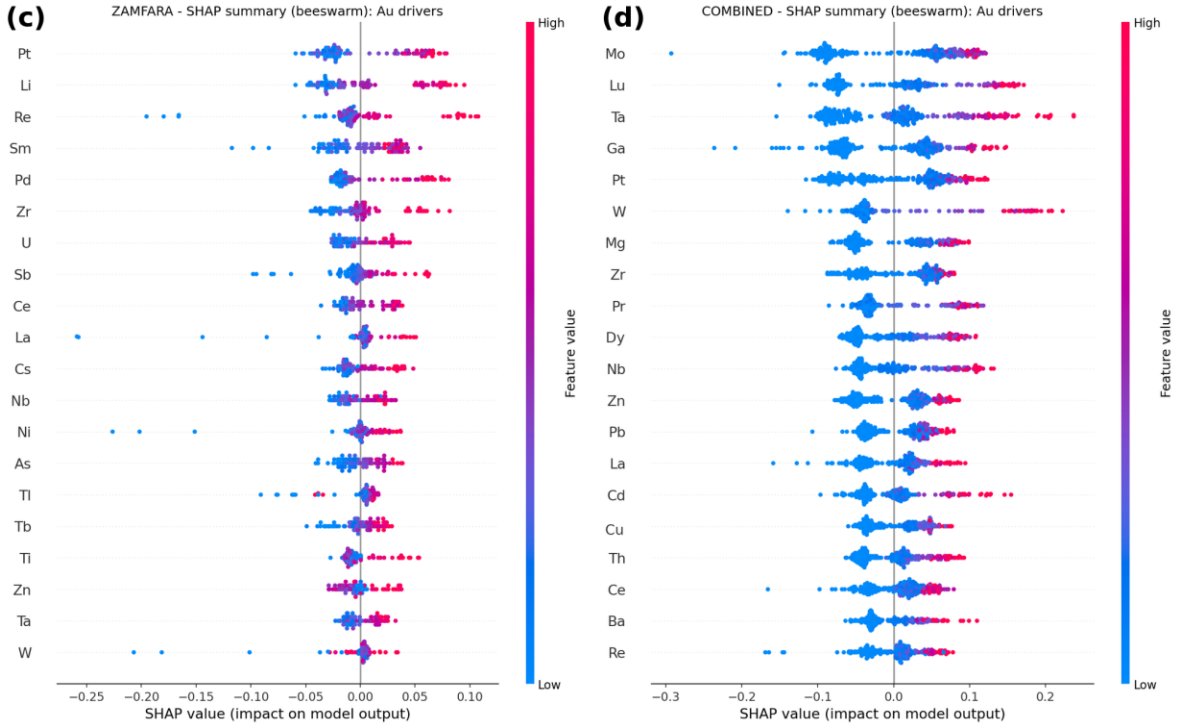

Fig. 4. SHAP summary plots (continued): (c) Zamfara State; (d) combined dataset of the three states.

Stepwise regression models

The summary statistics and the retained predictors of the four models are given in Table 3. The Taraba model retained four predictors (Sm, Er, Gd, Dy), all rare earth elements, and explained 98.77% of the variation in Au (adjusted R^2 = 98.72%, predicted R^2 = 88.35%). The Niger model retained W, Si, Hg, Sm and Gd (R^2 = 94.48%, predicted R^2 = 81.83%). The Zamfara model retained Pt, Pd, Sm and Ce and was the weakest (R^2 = 62.45%, adjusted R^2 = 60.87%, predicted R^2 = 55.72%). The TANIZA model retained eight predictors (Ba, W, B, Hg, La, Si, Tb, Sm) and explained 98.55% of the variation in Au (adjusted R^2 = 98.51%, predicted R^2 = 97.49%). The regression equations of the models are given in Table 4. The gap between R^2 and predicted R^2, an indicator of overfitting, was 10.42 percentage points for the Taraba model, 12.65 for Niger, 6.73 for Zamfara and only 1.06 for TANIZA. Samarium was retained in all four models, Gd in the Taraba and Niger models, and W, Hg and Si in the Niger and TANIZA models.

Table 3. Summary statistics and predictors of the four gold predictive models developed by stepwise regression.

Model	N	Predictors retained	S	R^2 (%)	R^2 (adj) (%)	R^2 (pred) (%)
Taraba	100	Sm, Er, Gd, Dy	3,098,830	98.77	98.72	88.35
Niger	91	W, Si, Hg, Sm, Gd	2,718,453	94.48	94.15	81.83
Zamfara	100	Pt, Pd, Sm, Ce	147,207	62.45	60.87	55.72
TANIZA (combined)	291	Ba, W, B, Hg, La, Si, Tb, Sm	2,132,431	98.55	98.51	97.49

N = number of samples; S = standard error of the regression, in the units of the Au response variable (EF); adj = adjusted; pred = predicted.

Table 4. Regression equations of the gold predictive models (Au and all element terms are enrichment factors; Minitab 17 output).

Model	Regression equation
Taraba	Au = 957389 + 11.552 Sm − 1.702 Er − 5.178 Gd + 10.321 Dy
Niger	Au = 1614850 + 1.7342 W − 280864 Si + 1.2299 Hg + 3.690 Sm − 4.921 Gd
Zamfara	Au = 117723 + 5.44 Pt + 0.570 Pd + 4.06 Sm + 47.5 Ce
TANIZA (combined)	Au = 376569 + 354.5 Ba + 1.2617 W − 4.278 B + 1.3067 Hg − 13.97 La − 154914 Si − 0.8090 Tb + 7.637 Sm

The EFs of the retained predictors from any new or independent mine dataset, calculated with Fe as reference and the upper continental crust as background (Eq. 1), can be substituted into the equation of the relevant model to predict the Au EF. The EFs are dimensionless ratios (Eq. 1); their large magnitude (mean Au EF of the order of 10^5–10^7) reflects the very low Au content of the upper continental crust used as background relative to the Au concentrations of the spoils. Variance inflation factors (VIF) were 2.3–8.0 in the Taraba model, 2.0–6.9 in Niger, 1.5–2.1 in Zamfara and 4.3–10.2 in TANIZA (highest for Sm, B, Hg and Si), which indicates moderate to high multicollinearity among the rare earth elements and with Hg, B and Si in the Taraba and TANIZA models. The negative coefficients of Er, Gd, La and Tb are therefore conditional on the other predictors in the equation and should not be read as evidence that these

elements are negatively associated with gold; each equation should be used with all of its terms together.

Graphical performance count

The performance counts of the Taraba, Zamfara, Niger and TANIZA models were 85.6%, 87.7%, 98.1% and 98.9% respectively (Fig. 5; Table 5), all above the 50% criterion, so all four were considered good models. The Zamfara model, although its R^2 of 62.45% was the lowest, achieved a performance count of 87.7%.

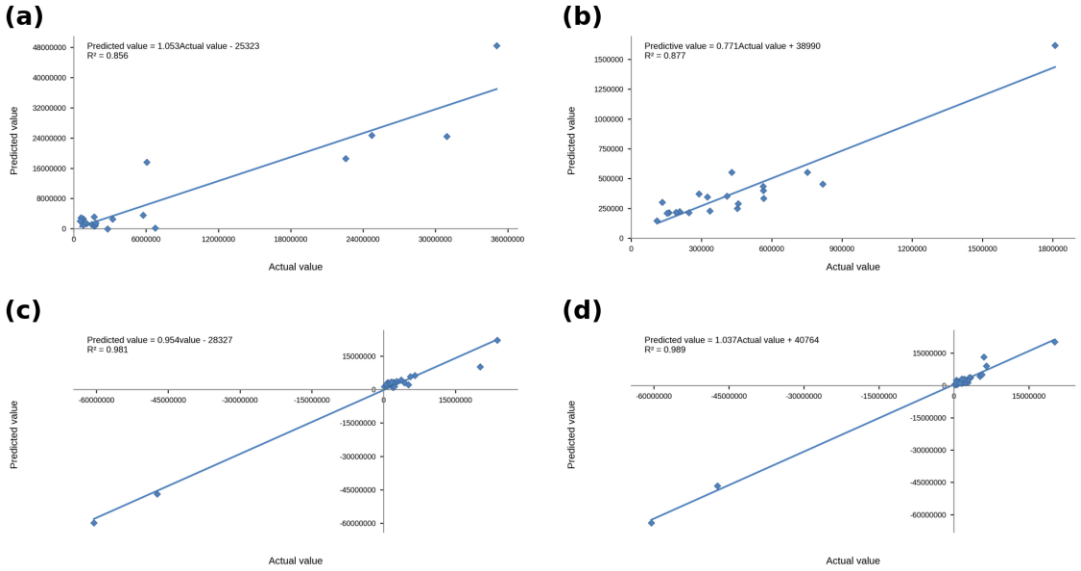

Fig. 5. Performance counts of the gold predictive models: predicted versus actual Au (EF) for selected samples with fitted trend line and R^2: (a) Taraba (n = 20; R^2 = 0.856); (b) Zamfara (n = 20; R^2 = 0.877); (c) Niger (n = 20; R^2 = 0.981); (d) TANIZA (n = 34; R^2 = 0.989).

Table 5. Graphical performance count of the four gold predictive models.

Model	n plotted	R^2 of trend line	Performance count (%)	Verdict
Taraba	20	0.856	85.6	Good (> 50%)
Zamfara	20	0.877	87.7	Good (> 50%)
Niger	20	0.981	98.1	Good (> 50%)
TANIZA (combined)	34	0.989	98.9	Good (> 50%)

Performance count = R^2 of the trend line of predicted against actual Au, expressed as a percentage; a model was considered good when it exceeded 50%.

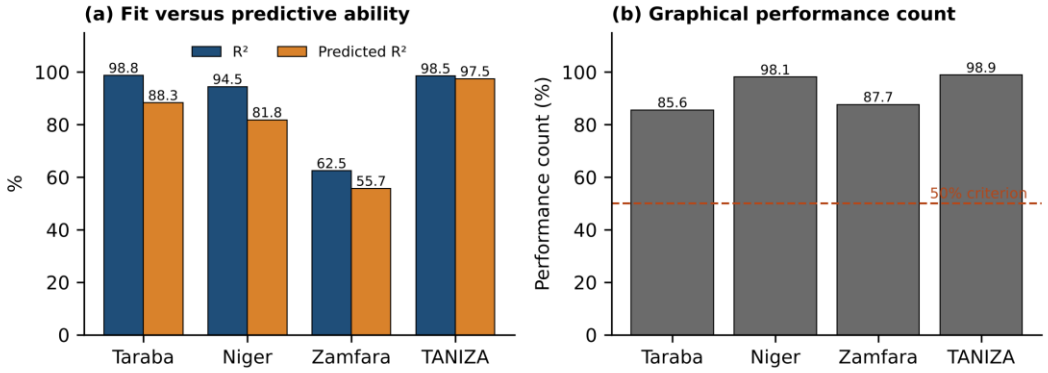

Fig. 6. Comparison of the four models: (a) R^2 and predicted R^2; (b) graphical performance count against the 50% acceptance criterion.

Regression assumption tests

All four regression assumptions were satisfied in all four datasets when the tests were applied to the full regressions of the Au EF on the EFs of the other elements. The Harvey–Collier tests gave no evidence of non-linearity ($p = 0.522–0.890$), the Breusch–Pagan tests none of heteroscedasticity ($p = 0.560–0.820$), the Shapiro–Wilk tests none of departure from normality of the residuals ($W = 0.979–0.993$; $p = 0.542–0.649$), and the Durbin–Watson statistics (1.884–2.314) lay within the 1.5–2.5 band, indicating no autocorrelation of errors.

Applied to the residuals of the four final stepwise models (Table 3), linearity (Harvey–Collier $p = 0.949$, 0.388, 0.453 and 0.924 for the Taraba, Niger, Zamfara and TANIZA models) and independence of errors (Durbin–Watson = 1.933, 2.270, 1.894 and 2.076) were satisfied in all four models. Homoscedasticity was satisfied in the three state models (Breusch–Pagan $p = 0.144$, 0.843 and 0.173) but not in the TANIZA model ($p < 0.001$). Normality of the residuals was satisfied only in the Zamfara model (Shapiro–Wilk $W = 0.982$, $p = 0.175$); the residuals of the Taraba ($W = 0.853$), Niger ($W = 0.900$) and TANIZA ($W = 0.832$) models departed from normality ($p < 0.001$) because of a few large residuals (excess kurtosis 6.0, 2.1 and 6.9). Because the TANIZA residuals were heteroscedastic, its coefficients were re-tested with heteroscedasticity-consistent (HC3) standard errors: seven of the eight predictors remained significant at $\alpha = 0.05$ ($p \leq 0.044$), and B was marginal ($p = 0.061$). The coefficients in Table 4 are unchanged, but the standard errors and p-values of the Taraba, Niger and TANIZA models should be read with these departures in mind.

Discussion

Predictor suites differ by state, as pathfinder theory predicts

The most important predictors of gold differed among the three states, both in the Random Forest ranking and in the final regression models. This agrees with the principle that pathfinder suites depend on the deposit and the terrain and must be established locally (Rose et al., 1979). It is also consistent with the contrasting geological settings: Garba (2000) attributed the Pan-African mesothermal gold at Bin Yauri (Kebbi State) to hydrothermal fluids depositing gold in structurally controlled veins, Obaje (2009) described the contrast between the schist belts of the north-west and the basement of eastern Nigeria, and Salawu et al. (2023) demonstrated structural control of gold in the Zuru Schist Belt of Kebbi State. Differences in lithology, structure and weathering environment have probably produced distinct surface geochemical signatures of gold (Webster and Mann, 1984; Butt et al., 2000).

Taraba: A rare-earth signature

The Taraba model was composed entirely of rare earth elements (Sm, Er, Gd, Dy) and explained 98.77% of the variation in gold. Rare earth element patterns of scheelite in hydrothermal gold deposits record the chemistry of the mineralising fluids (Ghaderi et al., 1999; Sciuba et al., 2020), rare earth elements are mobilised and deposited in hydrothermal systems (Lottermoser, 1992; Williams-Jones et al., 2012), and rare-metal-bearing pegmatites occur in the basement of northern Nigeria (Garba, 2003). It could probably mean that the rare earth elements are carried in resistant accessory minerals that crystallised from the same fluids that deposited gold and survived weathering with it, so that they remain with gold in the spoils; this requires mineralogical confirmation.

Niger: Tungsten, silicon and mercury

The retention of W, Si, Hg, Sm and Gd in the Niger model fits the quartz-vein character of orogenic gold systems, in which gold occurs with scheelite and sulphides (Groves et al., 1998), and the association of intrusion-related gold with tungsten–tin provinces (Thompson et al., 1999). Silicon probably reflects the quartz veins that host the gold, tungsten reflects scheelite or wolframite in the veins, and Sm and Gd the rare earth elements associated with scheelite and accessory minerals. Mercury, in contrast, reflects the amalgamation of the ore and not a natural geological association (Veiga et al., 2006; Odukoya et al., 2022; Anene et al., 2024): it is a useful predictor because it follows the processing of the ore (because Hg marks where ore has already been processed, it cannot point to unmined gold; the Niger and TANIZA models are therefore suited to screening mine spoils and not to locating unmined mineralisation), and must be measured in any case for health reasons. The model explained 94.48% of the variation in Au; Odukoya et al. (2022) and Okonkwo et al. (2021), who analysed soils of artisanal gold-mining sites in Niger State, are the closest regional precedents for the use of soil geochemistry at gold-mining sites in the state.

Zamfara: Platinum-group elements and a weaker model

The association of gold with Pt and Pd in the Zamfara model, together with Sm and Ce, is consistent with the mafic–ultramafic metavolcanic units of the north-western schist belts, which are the principal hosts of the platinum-group elements (Obaje, 2009). The Zamfara model was the weakest (R^2 = 62.45%, predicted R^2 = 55.72%). The most probable explanation is the erratic distribution of coarse free gold in quartz veins, a sampling-variability problem identified by Rose et al. (1979) as a major difficulty in gold geochemistry, possibly compounded in this state by the reworking of spoil by wind. The model is nonetheless acceptable, since its performance count (87.7%) exceeds the 50% threshold.

TANIZA: A regional screening equation

The TANIZA model retained Ba, W, B, Hg, La, Si, Tb and Sm. These are consistent with the geochemical signature of orogenic gold deposits, which associates gold with W, B, As, Sb and Hg hosted by quartz veins within potassic, sericite-bearing alteration (Groves et al., 1998), and with the trace-element suite of scheelite, including the rare earth elements (Sciuba et al., 2020): Si probably reflects quartz veins; W, La, Tb and Sm reflect scheelite and other accessory minerals; B reflects tourmaline; Ba reflects potassic alteration; and Hg the amalgamation of the ore (Veiga et al., 2006). The model explained 98.55% of the variation in Au and had a predicted R^2 of 97.49% (this high value is not driven by differences among the states: they account for only 3% of the variance in the Au EF, and adding state terms to the model did not improve the fit, R^2 = 98.56% against 98.55%, F-test p = 0.48; the model nevertheless fits the states unequally, with within-state R^2 of 98.9% for Taraba, 95.9% for Niger and 36.0% for Zamfara, and a model

built on two states predicted the third well for Taraba (R^2 = 97.4%), moderately for Niger (62.3%) and not at all for Zamfara), only 1.06 percentage points lower, which indicates that it predicts samples not used in building it almost as well as it fits those from which it was built. Its predictors differ in part from those of the state models, which suggests that a regional model captures the signal common to the three states while the state models capture local signals. State-specific models are therefore preferable for detailed exploration within each state, and TANIZA for regional screening of mine spoils.

Novel and confirmed predictors

Table 6 compares the predictors with the gold indicators reported in the literature. Tungsten, mercury, silicon and boron confirm conventional indicators, whereas the rare earth elements Sm, Gd, Er, Dy, Ce, La and Tb, barium and the platinum-group elements Pt and Pd are not routinely used as gold indicators in the Nigerian basement complex. The recurrence of Sm in all four models suggests that rare-earth-bearing accessory minerals are a common component of the gold-bearing spoils of the three states. The conventional pathfinders As, Sb, Bi and Te, although widely regarded as primary indicators, were not retained by the stepwise procedure in any model: with EFs normalised to Fe and closely intercorrelated element groups, the algorithm selects the combinations that best explain Au in these datasets, which does not imply that As and Sb are unimportant in the primary mineralisation.

Table 6. Comparison of the predictors of the four models with gold indicators reported in the literature.

Element	Model(s)	Reported in the literature as gold indicator	Status in this study
W	Niger, TANIZA	Yes: scheelite and wolframite in orogenic and intrusion-related gold systems (Groves et al., 1998; Thompson et al., 1999)	Confirmed predictor
Hg	Niger, TANIZA	Yes, but anthropogenic: introduced by amalgamation of gold ore (Veiga et al., 2006)	Confirmed predictor (processing-related)
Si	Niger, TANIZA	Indirect: quartz veins host gold (Groves et al., 1998)	Confirmed predictor
B	TANIZA	Yes: part of the characteristic orogenic-gold association, as tourmaline (Groves et al., 1998)	Confirmed predictor
Ba	TANIZA	Indirect: enrichment in potassic (sericite-bearing) alteration (Groves et al., 1998)	Predictor, indirect association
Sm	Taraba, Niger, Zamfara, TANIZA	No (not routine): REE carried in accessory minerals and scheelite (Ghaderi et al., 1999)	New predictor (common to all four models)
Gd	Taraba, Niger	No (not routine): rare earth element	New predictor
Er, Dy	Taraba	No (not routine): rare earth elements	New predictors
Ce	Zamfara	No (not routine): rare earth element	New predictor

© IIJP 2026 | Volume 2, Issue 4, pp. 1266-1286 | ISSN: 2636-4484

Element	Model(s)	Reported in the literature as gold indicator	Status in this study
La, Tb	TANIZA	No (not routine): REE recorded in scheelite and hydrothermal minerals (Sciuba et al., 2020; Williams-Jones et al., 2012)	New predictors
Pt, Pd	Zamfara	Rarely: platinum-group elements of mafic–ultramafic rocks (Obaje, 2009)	New predictors

Model validation and stepwise selection

Montgomery et al. (2012) recommended the predicted R^2 as a measure of a regression model's ability to predict new observations, because ordinary R^2 always increases as predictors are added. Linearity and independence of errors were satisfied in all four final models, but the residuals of the Taraba, Niger and TANIZA models were not normal and those of the TANIZA model were heteroscedastic (Section 4.5); the coefficients of these models are therefore best regarded as predictive and not as exact estimates of effect size. Stepwise regression is a practical way to select a small set of predictors from many candidates (Draper and Smith, 1998), but it can retain variables by chance and overestimate their effects (Whittingham et al., 2006); predicted R^2 and the graphical performance count were therefore used as additional checks.

Suggestions for further studies

i. The TANIZA model should be tested on new samples from the three states and from other gold mining states of Nigeria.
ii. Mineralogical studies, for example by scanning electron microscopy, should be carried out to identify the rare-earth-bearing minerals that associate with gold in the spoils.
iii. Samples should be collected at several depths through the spoil heaps to study the vertical distribution of gold and its predictors.
iv. Other machine learning algorithms, such as gradient boosting and support vector machines, should be compared with Random Forest for predictor screening.
v. Physicochemical properties should be determined on individually located samples, together with depth and spoil type, to relate them directly to the element concentrations.
vi. Lead isotope and bioaccessibility studies should be carried out in Zamfara State to separate ore-derived lead from other sources and to assess health risk.

Conclusions and recommendations

Random Forest with SHAP effectively screened a 55-element dataset from the gold mine spoils of Niger, Taraba and Zamfara States, and stepwise regression converted the screening into four simple gold predictive models whose performance counts (85.6–98.9%) all exceeded the 50% criterion. The elements that best predict gold differ among the states (Sm, Er, Gd and Dy in Taraba; W, Si, Hg, Sm and Gd in Niger; Pt, Pd, Sm and Ce in Zamfara), probably reflecting differences in geology and environment, while the combined TANIZA model (Ba, W, B, Hg, La, Si, Tb, Sm; R^2 = 98.55%, predicted R^2 = 97.49%) provides a single equation for regional screening. Rare earth elements (Sm, Gd, Er, Dy, Ce, La, Tb) and the platinum-group elements (Pt, Pd), together with W, Si, B, Ba and Hg, may extend the conventional pathfinder suite for gold in Nigerian basement-complex mine spoils.

We recommend that (i) the TANIZA model be used as a regional screening tool by researchers, the Nigerian Geological Survey Agency and exploration companies, and the state models for detailed work within each state, after testing on new samples; (ii) rare earth and platinum-group elements be included with the conventional pathfinders in multi-element suites for gold exploration in the Nigerian basement complex; (iii) the rare-earth-bearing minerals

associated with gold be identified by mineralogical study (e.g. scanning electron microscopy) and sampling be extended to several depths in spoil heaps; (iv) other algorithms, such as gradient boosting and support vector machines, be compared with Random Forest; and (v) government agencies establish a geochemical and environmental database of mine spoils.

Declarations

Author contributions. B. S. Ezekiel: conceptualisation, investigation, data curation, formal analysis, visualisation and writing of the original draft. W. O. Medjor, O. M. Othniel, S. Shikaa, F. Egharevba and E. A. Medjor: supervision, methodology and review and editing.

Declaration of AI use. An AI tool (Claude, Anthropic) was used to assist in writing the program for the regression-assumption tests; the authors reviewed and edited all content and take full responsibility for it.

Funding. This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests. The authors declare that they have no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

Data availability. The raw concentrations, enrichment factors and the full Minitab model outputs, including the regression equation of each model, are provided in the Supplementary Appendix and in the appendices of the first author's doctoral thesis, and are available from the corresponding author on reasonable request.

Acknowledgements. The authors thank the staff of the Lagos University Central Laboratory for the ICP-OES analyses and the mining communities of the sampled Local Government Areas for access to the sites.

References

Anene, N. C., Dangulbi, B. M., & Veiga, M. M. (2024). Assessment of gold and mercury losses in an artisanal gold mining site in Nigeria and its implication on the local economy and the environment. *Minerals, 14*(11), Article 1131. https://doi.org/10.3390/min14111131

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Breusch, T. S., & Pagan, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation. *Econometrica, 47*(5), 1287–1294. https://doi.org/10.2307/1911963

Butt, C. R. M., Lintern, M. J., & Anand, R. R. (2000). Evolution of regoliths and landscapes in deeply weathered terrain — implications for geochemical exploration. *Ore Geology Reviews, 16*(3–4), 167–183. https://doi.org/10.1016/S0169-1368(99)00029-3

Carranza, E. J. M., & Laborte, A. G. (2015). Random forest predictive modelling of mineral prospectivity with small number of prospects and data with missing values in Abra (Philippines). *Computers & Geosciences, 74*, 60–70. https://doi.org/10.1016/j.cageo.2014.11.002

Draper, N. R., & Smith, H. (1998). *Applied regression analysis* (3rd ed.). Wiley.

Durbin, J., & Watson, G. S. (1950). Testing for serial correlation in least squares regression. I. *Biometrika, 37*(3/4), 409–428. https://doi.org/10.2307/2332391

Garba, I. (2000). Origin of Pan-African mesothermal gold mineralisation at Bin Yauri, Nigeria. *Journal of African Earth Sciences, 31*(2), 433–449.

Garba, I. (2003). Geochemical discrimination of newly discovered rare-metal bearing and barren pegmatites in the Pan-African (600 ± 150 Ma) basement of northern Nigeria. *Applied Earth Science (Transactions of the Institution of Mining and Metallurgy B), 112*(3), 287–292.

Ghaderi, M., Palin, J. M., Campbell, I. H., & Sylvester, P. J. (1999). Rare earth element systematics in scheelite from hydrothermal gold deposits in the Kalgoorlie-Norseman region, Western Australia. *Economic Geology, 94*(3), 423–437. https://doi.org/10.2113/gsecongeo.94.3.423

Groves, D. I., Goldfarb, R. J., Gebre-Mariam, M., Hagemann, S. G., & Robert, F. (1998). Orogenic gold deposits: A proposed classification in the context of their crustal distribution and relationship to other gold deposit types. *Ore Geology Reviews, 13*(1–5), 7–27.

Harvey, A. C., & Collier, P. (1977). Testing for functional misspecification in regression analysis. *Journal of Econometrics, 6*(1), 103–119. https://doi.org/10.1016/0304-4076(77)90057-4

Hastie, T., Tibshirani, R., & Friedman, J. (2021). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed., 12th printing). Springer.

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. *Proceedings of the 14th International Joint Conference on Artificial Intelligence (IJCAI-95)*, 1137–1143.

Lottermoser, B. G. (1992). Rare earth elements and hydrothermal ore formation processes. *Ore Geology Reviews, 7*(1), 25–41. https://doi.org/10.1016/0169-1368(92)90017-F

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems 30* (pp. 4765–4774).

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). *Introduction to linear regression analysis* (5th ed.). Wiley.

Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., & Veith, T. L. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Transactions of the ASABE, 50(3), 885–900. https://doi.org/10.13031/2013.23153

Nunoo, S., Abu, M., Ayitey, E., Mvile, B. N., & Kalimenze, J. D. (2024). Multi-method machine learning techniques in gold pathfinder elements prediction in central parts of Tanzania using stream sediment geochemical data. Physics and Chemistry of the Earth, 136, Article 103766. https://doi.org/10.1016/j.pce.2024.103766

Obaje, N. G. (2009). *Geology and mineral resources of Nigeria* (Lecture Notes in Earth Sciences, Vol. 120). Springer.

Odukoya, A. M., Uruowhe, B., Watts, M. J., Hamilton, E. M., Marriott, A. L., Alo, B., & Anene, N. C. (2022). Assessment of bioaccessibility and health risk of mercury within soil of artisanal gold mine sites, Niger, North-central part of Nigeria. *Environmental Geochemistry and Health, 44*(3), 893–909. https://doi.org/10.1007/s10653-021-00991-2

Okonkwo, S. O., Jacob, J. O., Iyaka, Y. A., & Inobeme, A. (2021). Assessment of selected heavy metal concentrations in soils from a mining area in Minna, Niger state. Environmental Monitoring and Assessment, 193(3), Article 140. https://doi.org/10.1007/s10661-021-08903-8

Reimann, C., & de Caritat, P. (2000). Intrinsic flaws of element enrichment factors (EFs) in environmental geochemistry. *Environmental Science and Technology, 34*(24), 5084–5091. https://doi.org/10.1021/es001339o

Reimann, C., & de Caritat, P. (2005). Distinguishing between natural and anthropogenic sources for elements in the environment: Regional geochemical surveys versus enrichment factors. *Science of the Total Environment, 337*(1–3), 91–107. https://doi.org/10.1016/j.scitotenv.2004.06.011

Rodriguez-Galiano, V., Sanchez-Castillo, M., Chica-Olmo, M., & Chica-Rivas, M. (2015). Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines. *Ore Geology Reviews, 71*, 804–818. https://doi.org/10.1016/j.oregeorev.2015.01.001

Rose, A. W., Hawkes, H. E., & Webb, J. S. (1979). *Geochemistry in mineral exploration* (2nd ed.). Academic Press.

Salawu, N. B., Omosanya, K. O., Eluwole, A. B., Saleh, A., & Adebiyi, L. S. (2023). Structurally-controlled gold mineralization in the Southern Zuru Schist Belt NW Nigeria: Application of remote sensing and geophysical methods. *Journal of Applied Geophysics, 211*, Article 104969. https://doi.org/10.1016/j.jappgeo.2023.104969

Sciuba, M., Beaudoin, G., Grzela, D., & Makvandi, S. (2020). Trace element composition of scheelite in orogenic gold deposits. *Mineralium Deposita, 55*(6), 1149–1172. https://doi.org/10.1007/s00126-019-00913-4

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika, 52*(3/4), 591–611. https://doi.org/10.2307/2333709

Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T., & Zeileis, A. (2008). Conditional variable importance for random forests. *BMC Bioinformatics, 9*, Article 307. https://doi.org/10.1186/1471-2105-9-307

Sutherland, R. A. (2000). Bed sediment-associated trace metals in an urban stream, Oahu, Hawaii. *Environmental Geology, 39*(6), 611–627. https://doi.org/10.1007/s002540050482

Taylor, S. R., & McLennan, S. M. (1985). *The continental crust: Its composition and evolution*. Blackwell.

Thompson, J. F. H., Sillitoe, R. H., Baker, T., Lang, J. R., & Mortensen, J. K. (1999). Intrusion-related gold deposits associated with tungsten-tin provinces. *Mineralium Deposita, 34*(4), 323–334. https://doi.org/10.1007/s001260050207

Veiga, M. M., Maxson, P. A., & Hylander, L. D. (2006). Origin and consumption of mercury in small-scale gold mining. *Journal of Cleaner Production, 14*(3–4), 436–447.

Webster, J. G., & Mann, A. W. (1984). The influence of climate, geomorphology and primary geology on the supergene migration of gold and silver. *Journal of Geochemical Exploration, 22*(1–3), 21–42. https://doi.org/10.1016/0375-6742(84)90003-0

Whittingham, M. J., Stephens, P. A., Bradbury, R. B., & Freckleton, R. P. (2006). Why do we still use stepwise modelling in ecology and behaviour? *Journal of Animal Ecology, 75*(5), 1182–1189. https://doi.org/10.1111/j.1365-2656.2006.01141.x

Williams-Jones, A. E., Migdisov, A. A., & Samson, I. M. (2012). Hydrothermal mobilisation of the rare earth elements – a tale of "ceria" and "yttria". *Elements, 8*(5), 355–360. https://doi.org/10.2113/gselements.8.5.355

Zuo, R., & Xiong, Y. (2018). Big data analytics of identifying geochemical anomalies supported by machine learning methods. Natural Resources Research, 27(1), 5–13. https://doi.org/10.1007/s11053-017-9357-0