This is the peer-reviewed, final accepted version for American Mineralogist, published by the Mineralogical Society of America.
The published version is subject to change. Cite as Authors (Year) Title. American Mineralogist, in press.
DOI: https://doi.org/10.2138/am-2023-9164. http://www.minsocam.org/

Revision 1.0

# A Multivariate Statistical Approach for Mineral Geographic Provenance Determination using Laser-Induced Breakdown Spectroscopy and Electron Microprobe Chemical Data: A Case Study of Copper-bearing Tourmalines

Barbara L. Dutrow[1], Nancy J. McMillan[2], Darrell J. Henry[1]

1 Department of Geology & Geophysics, Louisiana State University, Baton Rouge, LA, 70803 USA; Dutrow@lsu.edu, glhenr@lsu.edu

2 Department of Geological Sciences, New Mexico State University, Las Cruses, NM 88003 nmcmilla@nmsu.edu

## ABSTRACT

The geographic provenance of minerals provides key insights into a range of geologic problems including the source of gem materials. The tourmaline supergroup is unparalleled in its ability to record and preserve extensive chemical signatures of its formational environment. To evaluate the likelihood that tourmalines of similar compositions from separate geographic localities could be differentiated, a multivariate statistical approach has been utilized on two complementary datasets. These chemical analytical datasets of copper-bearing "paraíba" tourmaline include a qualitative Laser Induced Breakdown Spectroscopy (LIBS) data set and a quantitative electron microprobe (EMP) data set.

Fifty-four samples of copper-bearing tourmalines from known source locations from Brazil (São José de Batalha of Paraíba state and the neighboring Rio Grande do Norte state), from Mozambique, and from Nigeria, were analyzed using LIBS with a subset of these samples analyzed by EMP. Datasets obtained by each method were evaluated with multivariate statistics (PCA, PLSR). Although the sample set is limited, sequential PLSR modeling of the spectra clearly distinguished the four localities with high success: > 95% for LIBS and > 87% for EMP. The statistical analyses of the two techniques, LIBS and EMP, suggest that each technique emphasizes different elements for discrimination when considered in context of the available data. The elements Cu, Mn, Fe, Mg, Ti, Zn, K, H, Co, and V were significant in LIBS chemometric models. Statistically significant elements in EMP models were Mn, Cu, Al, Ca, K, and F. Each technique results in a robust determination for geographic provenance of tourmalines with comparible compositions. The significant distinguishing chemical elements reflect geochemical distinctions in each host environment that are imparted on the tourmaline.

Multivariate statistics applied to LIBS and EMP data provide an effective tool for provenance discrimination of paraíba tourmalines, distinguishing Brazilian-sourced samples from African-sourced materials. These data provide new methods for separating the geographic origin of minerals with very similar composition such as demonstrated here for copper-bearing tourmalines.

Keywords: tourmaline, provenance, Paraiba, LIBS, EMP

## INTRODUCTION

The determination of the geographic origin (provenance) of minerals separated from their original host rock can provide significant insights to various geological processes. Provenance studies can relate to a geographic origin or locality, which may be associated with a spatially restricted geologic unit or to a host rock environment. For example, provenance elucidates shifting patterns of modern and ancient sedimentation (e.g., Morton et al. 2005), provides key information on paleogeographic/tectonic reconstructions (e.g., von Eynatten and Gaupp 1999), establishes a basis for identification of valuable minerals mined in conflict zones (e.g., Hark et al. 2012; McManus et al. 2020) or the likely sources of some gemstones (e.g., Palke et al. 2018) and refines exploration strategies key to identifying sources of needed critical materials (e.g., Lohmeier et al. 2021). Additionally, geographic origin of gem materials is a complex and important problem in the world economy as companies and organizations strive to maintain and certify a supply chain free of conflict minerals. In other cases, substantial price differences of gemstones result from their different geographic origins. Commonly, mineral chemistry is utilized to provide provenance information. This chemical distinction is challenging when

differences among possible source areas are subtle or exhibit considerable overlap in chemical parameters or when age criteria alone are insufficient.

Many minerals retain chemical signatures of their formational environment, but no mineral embeds the range of chemical fingerprints better than the minerals of the tourmaline supergroup. Even during a complex, multistage geologic history that can include crystallization, weathering, reburial, metamorphosis, regrowth and deformation, tourmaline retains textural and chemical signatures of its earlier evolutionary history (e.g., Henry and Guidotti 1985; Henry and Dutrow 1996; van Hinsberg et al. 2011a,b). Tourmaline's utility as a petrogenetic indicator stems, in part, from its (1) complex crystal chemistry, providing structural and chemical flexibility to incorporate a wide range of chemical constituents of multiple valence states and sizes, to imprint a signature of its chemical environment of formation, (2) stability over an extensive range pressures (P) and temperatures (T) encompassing nearly all crustal and upper-mantle conditions, (3) ability to form in widely varying rock and fluid compositions, and (4) minimal volume diffusion such that its imprinted chemical signature remains intact (see summaries by Henry and Dutrow 1996; Dutrow and Henry 2011; van Hinsberg et al. 2011b).

The rich chemical signatures, coupled with its mechanical and chemical stability, make tourmaline a unique target for establishing *new* methodologies for provenance studies. In some instances, chemical distinctions among sources are subtle, yet critical to define. An excellent test case, and one of economic interest, is the sourcing of copper-bearing tourmalines. Determining their geographic origin, or provenance, is challenging and has important financial implications.

Copper-bearing elbaitic or liddicoatitic tourmaline is widely prized as a gemstone due to its vivid, saturated, 'neon' blue hues that are caused by the incorporation of $Cu^{2+}$ as a chromophore (Fig. 1; e.g., Rossman et al. 1991). Originally discovered in the 1980s in Brazil

near the São José da Batalha Mine in the state of Paraíba (Koivula and Kammerling 1989) and later in the 1990s in the nearby state of Rio Grande do Norte (e.g., Fritsch et al. 1990; Shigley et al. 2001), these exquisite Cu-bearing specimens became known as Paraíba tourmalines (Fig. 1). Subsequently, other localities hosting similarly colored Cu-bearing tourmalines were found as elbaitic tourmaline in Nigeria in 2001 (Smith et al. 2001) and Mozambique in 2004 (Wentzell 2004; Abduriyim and Kitawaki 2005; Laurs et al. 2008; Katsurada and Sun 2017). The African tourmalines were found originally in secondary alluvial deposits. Chemically, all of these tourmalines are classified as elbaite or fluor-elbaite species, with a general formula of Na $(Li_{1.5}Al_{1.5})$ $Al_6$ $(Si_6O_{18})$ $(BO_3)_3$ $(OH)_3$ (OH) or F replacing one OH (for species nomenclature see Henry et al. 2011; Henry and Dutrow 2018) and $Cu^{2+}$ substituting into the octahedral site that typically accommodates Li-Al. In 2017, Cu-bearing fluor-liddicoatites – Ca $(Li_2Al)$ $Al_6$ $(Si_6O_{18})$ $(BO_3)_3$ $(OH)_3$ (F) - were discovered and were attributed to a locality in Mozambique (Katsurada and Sun 2017). The varietal name, "paraíba" tourmaline, is used to refer to any of the saturated blue, green, and violet tourmalines containing $Cu^{2+} \pm Mn^{2+}$ as chromophores (LMHC 2012). Paraíba tourmaline sources for gemstones are difficult, if not impossible, to distinguish based on color alone. Yet, the Brazilian material from the original mine area can command prices that are 5-10 times higher than those of their African counterparts of comparable quality and size. Consequently, provenance is an essential component of the tourmaline's value as a gemstone.

Major-element tourmaline "environmental" diagrams such as the Al-Fe-Mg ternary (Henry and Guidotti 1985) are not effective for determination of paraíba tourmaline sources because most have elbaitic composition except for the liddicoatitic tourmalines which are easily distinguished based on their elevated Ca contents. Consequently, this necessitates the use of other criteria such as minor and trace elements to potentially fingerprint the likely source of

paraíba tourmalines. For gemmy paraíba tourmaline, most attempts at provenance evaluations rely on quantities of a limited number of trace and minor element constituents (e.g., Cu, Zn, Ga, Sr, Sn, Pb), obtained via LA-ICP-MS, or isotopes, obtained via Secondary Ion Mass Spectrometry (Ludwig et al. 2011), that are plotted in simple binary or ternary diagrams or in a serial combination of these diagrams as a means to deconvolute the overlapping chemical signatures distinctive of a source (e.g., Abduriyim et al. 2006; Peretti et al. 2009; Palke et al. 2018; Okrusch et al. 2016; see review by Katsurada et al. 2019). Although these types of provenance diagrams have met with varying degrees of success, they do not holistically consider the entire range of paraíba tourmaline chemistry available for provenance evaluation.

This contribution explores use of a multivariate statistical approach for enhanced provenance determination that considers a wider spectrum of chemical information available from two distinctively different, but complementary, newly acquired chemical analytical datasets of paraíba tourmaline: Laser Induced Breakdown Spectroscopy (LIBS) spectra and electron microprobe (EMP) chemical analyses. The purpose of this study is to determine if multivariate statistics can reveal whether one or both datasets can be more effective or, at least complementary, provenance indicators for minerals with very similar compositions.

## METHODS

The LIBS analytical sample set consists of 54 copper-bearing tourmalines with known provenance from four distinct localities (Fig. 2). Samples were obtained from highly reputable gem dealers specializing in paraíba tourmaline (See Appendix 1 for sample information). Representing Brazil are 24 grains from two localities: São José de Batalha, Paraíba state (SJdB; the original Paraíba locality; 6 grains, 5.93 carats (cts), color-zoned blue, purple) and Rio Grande

do Norte state (RGdN; 15 grains, color-zoned blue, purple; Fig. 2a,b). In addition, three samples displaying neon-blue colors are identified as from "Brazil" but with unknown specific localities, two samples are in matrix and one is a single crystal. Mozambique (Moz) is represented by 24 tourmaline grains with a spectrum of colors including pink, blue, purple and green (total weight of 51.73 cts; see Fig. 2c). Nigeria (Nig) is represented by 11 grains (totaling 28.28 cts; Fig. 2d). Nigerian grains are largely green to blue green. Most rough crystals measured less than 1 cm in size and were without the matrix material.

**LIBS analyses**

LIBS is a relatively recent analytical technique that is finding utility in the geosciences (e.g., see reviews by Fabre 2020; Harmon and Senesi 2021). The information-rich spectra contain signatures of all elements in concentrations above detection limits (e.g., Cremers and Radziemski 2013), molecular emissions, select isotopic ratios (e.g., Smith et al. 2002; Doucet et al. 2011: Russo et al. 2011), and some structural information (Serrano et al. 2015) resulting in a detailed chemical fingerprint of the material analyzed. To take advantage of the rich chemical dataset embedded in tourmaline, this LIBS study uses the spectrum of relative peak intensities of each tourmaline rather than absolute quantities of individual elements within the tourmaline.

Minimal sample preparation is required for LIBS (see e.g., McMillan et al. 2018; McMillan et al. 2019 for additional information). Rough samples were cleaned with isopropyl alcohol to remove oils and surface residue and air dried. Most tourmalines are individual grains or clusters of grains. Originally, samples were mounted on a plexiglass sheet with BlueTac to secure the grains; later the BlueTac was eliminated. The sheet was placed into the sample holder in the LIBS instrument chamber. LIBS data were acquired prior to EMP data analyses to avoid

any possible contamination from EMP sample preparation such as polishing and carbon coating of the grains.

Tourmalines were analyzed with an Applied Spectra J200 LIBS instrument, at Materialytics, Inc., fitted with a Q-switched Quantel ULTRA 100 Big Sky Nd:YAG laser operated at a fundamental wavelength of 266 nm and < 6 ns pulse width. The instrument utilized an Andor Mechelle ME 5000 spectrograph ($\lambda/\Delta\lambda = 5000$) and an Andor iStar ICCD (intensified charge coupled device) camera, model DH334T-18F-03. Analytical conditions were a laser power of 150 mJ, with a delay of 0.5 microseconds between the time of the laser shot and light collection, a gate width (time of light collection) of 10 microseconds, and a nominal spot size of 50 micrometers (subsequent analyses demonstrated a larger ablation pit of nearly 80 micrometers). Spectra were obtained at 1 atm at room temperature in an argon atmosphere to confine the LIBS plasma and thus enhance emission intensity. Where grain size allowed, 64 shots were obtained per sample in an 8 x 8 grid with spacing of 100 micrometers between shots – an area covering about 1 mm x 1 mm. An ancillary study suggested that 64 shots were optimal for characterizing the samples (McMillan et al. 2019). At each analytical location, a cleaning shot was done prior to the analytical shot. The spectral emission was collected over the 26,000+ channels of the detector/spectrometer system to assemble the spectrum in the wavelength range from 200 – 1000 nm for each analytical shot. Spectra were truncated at 771 nm which preserves the potassium peaks at 766.5 and 769.9 nm but masks the primary argon peaks at higher wavelengths. Multiple shots per sample and their corresponding spectra are averaged and normalized to the mean peak intensity to produce a single spectrum per sample. Averaging LIBS spectra helps mitigate variations caused by inherent shot-to-shot variability (McMillan and Dutrow, in press). Background correction was not applied. Intensities were converted to log

values for modeling purposes. Where necessary, identification of LIBS peak positions utilized the on-line NIST database of optical emission lines (Kramida et al. 2022).

Acquisition of such a large dataset requires statistical methods and/or machine-learning techniques for data analyses and interpretation. This study employs the multivariate statistical techniques Principal Component Analysis (PCA, Esbensen 2004) and Partial Least Squares Regression (PLSR; Wold et al. 2001; Esbensen 2004) to quantitatively classify spectra with reference to the geographical source of the tourmaline. The strong emission response of some major elements required masking of select peaks from the spectra to allow subtler chemical variations to be enhanced. For these tourmalines, masking of peaks for the elements Si, Al, Li, Na as well as the Ca peaks at 393.3, 396.8, and 422.7 nm resulted in improved models. While other multivariate statistical techniques may be advantageous, for this test case, methods used previously were followed (e.g., McMillan et al. 2018).

**Multivariate Statistical Modeling**

PCA is a dimension-reducing multivariate technique that calculates linear regressions, or Principal Components (PCs), through the data set in multivariate space (24,350 variables). A PCA score plot (sample analyses in n-dimensional space projected onto the plane of two principal components, e.g., PC 1 and PC 2) displays the spectral/compositional relationships of the dataset in the two directions of the principal components. This comparison is used to determine the order in which the geographic localities (SJdB, RGdN, Moz, Nig) are modeled beginning with the compositionally most distinct group modeled first (Multari et al. 2010; Kochelek et al. 2015; McMillan et al. 2018).

PLSR models were used to quantitatively discriminate between the samples of the locality of interest and all other localities. PLSR is similar to PCA but includes the value of an

independent variable, in this case the Provenance Variable (PV), in the regression. Spectra of samples from the locality of interest were assigned a Provenance Variable value of 1; spectra of samples from all other localities were assigned a PV value of 0. To calibrate the model, 50% of the spectra from the geographic localities were selected; spectra from the 50% remaining samples were used for test-set validation in a later step. Because the database contained one spectrum per sample, no individual sample was present in both the calibration and validation sets, although samples from a given geographic locality were present in both sets. Statistical modeling was accomplished using the Unscrambler ® software by Camo. The NIPALS (Nonlinear Iterative Partial Least Squares) algorithm was applied with 15 PLSR components; no weighting was applied to variables. All models are mean-centered (see also McMillan et al. 2018 for further discussion).

To quantitatively assign a spectrum to a locality group, a numerical value that separates calculated Provenance Variable values for the two groups in the calibration set is defined: the Value of Apparent Distinction (VAD, Kochelek et al. 2015). The VAD is calculated as the value that gives the highest number of correctly assigned samples during calibration. Any sample with a calculated PV value greater than or equal to the VAD is classified as a tourmaline within the group of interest; those with calculated locality variables less than the VAD are classified as belonging to the group of the remaining localities. Once a VAD is assigned, it does not change during validation.

PLSR models were validated using test-set validation. PV values are calculated for tourmaline spectra *not* used to calibrate the PLSR model. The VAD determined during calibration is used to predict whether each spectrum in the validation set belongs to the locality of interest or the group of the remaining localities. The prediction accuracy is calculated as the

percent of correctly assigned test-set spectra for which locality information is known. For example, Model 1 evaluates São José da Batalha (SJdB) samples. Applying the VAD of 0.45 to the spectra not used in the calibration set, all of the São José da Batalha samples are predicted to be from this locality as well as one African sample, and the other samples are predicted to belong to the group of remaining samples (Fig. 3). Thus, Model 1 is 96% successful, one sample is miscategorized. Once validated, the decision tree of PLSR plots is developed for each remaining group of samples (RGdN, Moz, Nig).

Each PLSR model identifies spectra that belong to one group (i.e. the geographic locality). After a group is distinguished, those samples are removed from the dataset and all subsequent models. In this case, São José da Batalha samples in Model 1 are removed. The order of the models may be critical to obtaining sufficient separation of samples. Each model is determined by choosing the compositionally most distinct group at each step, as defined by the relationships on a PCA score plot. Because the most distinct group is always eliminated, the samples near the final decision tree are those with the most compositional similarities. Typically, samples in those groups are indistinguishable from each other when modeled in the presence of the other samples, but the small differences between them can be extracted and used to separate these groups when they are modeled in isolation after the other groups are removed.

**Electron Microprobe Analysis (EMP)**

To test the applicability of the multivariate statistical approach on widely available tourmaline compositional data from EMP, a subset of 15 tourmaline samples for which LIBS data were obtained (Fig. 4a-c; 5 grains – Brazil; 6 Mozambique; 4 Nigeria; and two additional samples), were analyzed by wavelength dispersive spectrometry using the JEOL 8230 electron microprobe at LSU. Quantitative compositional analyses for major and minor elements were

obtained at an accelerating potential of 15kV and a 10 nA beam current using a 2-micrometer spot size, with Na analyzed first. Natural minerals and synthetic materials were used as standards including andalusite (Al), diopside (Ca, Mg, Si), fayalite (Fe), chromite (Cr), kaersutite (Ti), rhodonite (Mn), willemite (Zn), chalcopyrite (Cu), galena (Pb), albite (Na), sanidine (K), fluorite or fluor-phlogopite (F), tugtupite (Cl) with synthetic $Bi_2Te_3$ (Bi), V-diopside glass (V), and GaAs (Ga). EMP detection limits are given in the Supplemental Data. Lithium, H or B cannot be effectively analyzed by the EMP and were not included in the data modeled. Two well-characterized elbaite tourmalines served as secondary standards. Count times for major elements were 10 s on the peak, 20 s on the background and for minor and trace elements 60 s peak, 30 s background. Analytical precision is estimated to be ±1 percent relative for the major elements and ± 5 % for the minor elements. Where color zoning is apparent, analytical traverses were made across the samples; in other cases, 10 – 30 analytical spots per grain were randomly selected.

Mineral formulae were normalized following the recommended procedures Henry et al. (2011) permitting B, H and $Fe^{3+}$ to be calculated based on stoichiometry and charge balance and Li estimated by the procedures of Pesquera et al. (2016). Calculating atoms per formula unit (apfu) served as an additional quality check for EMP data but the normalized data are not used for the statistical analysis. To avoid calculation artifacts, oxide weight percentages of measured elements were used for multivariate statistical modeling and are given in the Supplemental Data.

Evaluating the efficacy of multivariate statistical models for separating provenance of paraíba tourmaline using EMP data followed the same methodology as for separating the LIBS data. However, only 18 variables per chemical analysis are available for modeling. Although the data set comprised 295 analyses, only 15 samples were analyzed. All analyses for each sample

were restricted to either the calibration or the validation set to ensure that the models focus on fundamental characteristics of the tourmalines rather than simply identifying analyses from the same sample. Because of the low number of samples, calibrations were based on analyses from 2-4 samples per country and models were validated with two samples from each country. As a result, the calibration set comprised analyses from four (Mozambique), three (Brazil), or two (Nigeria) samples and the validation set comprised analyses from two samples from each country.

## RESULTS

Copper-bearing tourmalines analyzed in this study included elbaite or fluor-elbaite species; no samples of the rare Cu-bearing fluor-liddicoatite species were included. Representative EMP analyses for each geographic locality are given in Table 1. Cu-bearing fluor-liddicoatites are Ca-dominant from Mozambique (Katsurada and Sun 2017) and their geographic origin is easily determined based on the Ca-dominance of the tourmaline.

### Multivariate statistics using LIBS data

LIBS spectra (unmasked) for the Cu-bearing elbaites display prominent Na, Al, Si, Li, and B peaks, in addition to Cu and Mn peaks as expected (Fig. 5). In several samples, LIBS detected minor and trace elements such as K, Mg, Bi, Zn, Ga, and Sr. The presence of these elements were confirmed by previous LA-ICP-MS analyses of paraíba tourmaline (Z. Sun, personal communication). Although Ca and Mg are minor components, the high intensity of these emission lines reflects the relatively low ionization energy of the alkaline earth elements (Cremers and Radziemski 2013).

The decision tree for these sample suites consists of three models (Fig. 6; Dutrow et al. 2019). In an initial PCA that includes all the tourmaline spectra from the four localities (São

José da Batalha, Brazil, SJdB; Rio Grande do Norte, Brazil, RGdN; Mozambique; and Nigeria), no single group clustered tightly and the groups overlapped in PC1-PC2 space (Fig. 7). The SJdB spectra were chosen as the first group to model because the São José da Batalha, Brazil PLSR model had the highest success rate of all possible first models. Model 1, which classifies spectra as either belonging to the SJdB group or to the group of all other tourmalines, is excellent (Fig. 6), despite the overlap of groups in PCA space (Fig. 7). The calibration shows separation between the groups with a VAD of 0.45 (Fig. 3). The validation is 96% successful, correctly classifying 25 of 26 samples. The one false positive is a sample of Nigerian tourmaline classified as SJdB.

The spectra of SJdB tourmalines were removed from all subsequent models. Model 2 classifies spectra as belonging to RGdN or to the group of all other tourmalines (Mozambique and Nigeria). There is clear separation between the two groups in the calibration of Model 2 (Fig. 3), which used a VAD = 0.50 value. The validation is 96% successful, correctly predicting the provenance of 22 of 23 samples (Fig. 3,6). Again, one Nigerian sample yielded false positive results. This sample is the same as that which was incorrectly classified as SJdB in Model 1.

Finally, Model 3 discriminates between tourmaline spectra from Nigeria and Mozambique (Fig. 6). Spectra are well-separated in the calibration with a VAD= 0.52 (Model 3; Fig. 3). The calibration is 94% successful, correctly classifying 16 of 17 samples. One Nigerian sample was misclassified as belonging to the Mozambique group; however, it is a different sample than the false positive sample in Models 1 and 2. The consistent misclassification of Nigerian samples suggests that the sample set is too small to be representative of the actual dispersion of compositions. Alternatively, on visual examination, this sample has a saw mark which may have left a surface contamination or varied the surface texture of the sample that

affected plasma properties. Overall, the decision tree correctly classified 63 of 66 spectra (one spectrum per sample) resulting in an cumulative prediction accuracy of 95%. The overall true positive rate (only considering the location assigned to PV 1) is 94% (16 of 17; Fig. 3).

Based on the success of the previous geographic modeling, the geographic origin of two unknown Brazilian samples was predicted. Using the LIBS decision tree developed, both unknown samples classify as being from the Rio Grande do Norte, Brazil, locality.

**Multivariate statistics using EMP data**

A more widely used analytical technique for characterizing tourmaline mineral chemical is by electron probe microanalyses (EMP). As such, this multivariate statistical approach was developed using an EMP analytical dataset obtained for a subset of the tourmalines for which LIBS data had been acquired (see Supplemental Data for all oxide weight percentages used for multivariate statistics). Importantly, in addition to the major elements, Cu, Mn and F are present in the tourmalines in amounts readily analyzed by the EMP. F is not easily detected by LIBS but is with EMP. Vanadium, Cr and Pb are at, or below, EMP detection limits (Supplemental Data).

Modeling EMP data with multivariate statistics followed similar procedures as the modeling for the LIBS data. Because of the smaller sample set size, both Brazilian localities were combined. The character of the EMP data set is different than the LIBS data set, in which each sample is represented by a single spectrum. For EMP data, 10-30 points were analyzed for each of the 15 tourmaline samples (Brazil: 5; Mozambique: 6; Nigeria: 4) resulting in a total of 295 analyses. This data set captures the variability within each sample well, but there are too few samples to be representative of the variability within each country of origin.

A PCA score plot for the calibration EMP analyses in the models shows good clustering for analyses of each tourmaline sample but lacks distinct clustering of samples from each country

(Fig. 8). Some relationships are consistent with those found via LIBS. For example, the Brazilian samples plot at negative values of PC2 over a large range of PC1 values. The Nigerian and Mozambican samples cover broad areas that intersect near the origin of the score plot. Analysis of more samples could help the PCA discern different relationships that might provide better separation of the groups.

PLSR is a supervised method, where the variables (EMP analyses) are correlated with known Provenance Variables (PV). Because of this, PLSR models can be successful, regardless of messy relationships in PCA. Model 1 in the EMP decision tree (Fig. 9) separates Brazilian tourmaline analyses from the group of Mozambican and Nigerian samples. The calibration is 96% successful, correctly predicting the origin of 173 of 180 calibration samples with a VAD of 0.51 (Fig. 10). The validation is also 96% successful, correctly predicting the origin 110 of 115 analyses. Five Brazilian analyses are predicted to belong to the group of all others; there are no false positives.

Model 2 is more complex. The calibration (Fig. 10) establishes relatively consistent Provenance Variable values for Mozambican calibration analyses with an average near 1 (average = 0.91, range = 0.36-1.36, standard deviation = 0.18). In contrast, the PV values calculated for Nigerian samples, while less than 1, are different from each other. One sample clusters at an average of 0.45 and the other with an average of 0.03 (Fig. 10). Because one Nigerian sample has relatively high calculated PV values, the VAD that results in the best model success is 0.62. This VAD value results in a calibration accuracy of 97% (112 correct predictions of 115), with two false negatives and one false positive. However, this VAD value is not the best choice for the validation (Fig. 10). A higher VAD would have yielded a higher success, as all of the Mozambican validated analyses have fairly high calculated PV values, as do

19 of the 40 Nigerian validation analyses. This results in a prediction accuracy of 75% for this model (Fig. 9, 10; 56 of 75 analyses). More samples with analytical data are needed to calculate more successful models. Overall, the EMPA decision tree correctly predicts the country of origin of 87% of the analyses.

## DISCUSSION

These combined results underscore the utility of multivariate analyses for separating likely geographic source localities of compositionally similar minerals as demonstrated by elbaitic tourmalines. Significantly, these outcomes result in separation of geographic localities using considerably different mineral chemical acquisition techniques. For both techniques, the high prediction accuracy of modeling suggests that even with a limited dataset, subtle variations in chemical components, when taken as a whole, can provide important signatures of the source region. While the power of the data-rich LIBS spectra coupled with multivariate statistics has been previously demonstrated for separating locality information (e.g., Hark et al. 2012; McMillan et al. 2012; Schenk and Almirall 2012; Kochelek et al. 2015; Gyftokostas et al. 2020), multivariate statistics has not been demonstrated as a useful tool for separating localities using the widely available EMP data. For the LIBS technique, intensity of the emission lines reflects a combination of the elemental abundance and the emissivity properties. Separating localities in this dataset required masking peaks from select major elements in part, because they hid more subtle and meaningful chemical variations. In contrast, for the quantitative EMP data, subtle differences in minor elements facilitated separation of geographic localities. However, because of the smaller sample suite, only broad categories could be distinguished. More EMP data from additional samples of each locality would further refine this procedure.

Loading plots (Fig.11) exhibit the influence of each variable (elemental concentration for

EMPA and wavelength intensity for LIBS) on the direction of the principal component through the data set. Variables with values close to zero have minimal impact on the PC and are approximately the same in the samples as in the model. Variables with high positive values strongly influence the direction of the PC, have different values in the samples, and increase in concentration/intensity in the positive direction of the PC on a score plot. Variables with high negative values are similar except that they increase in concentration/intensity in the negative direction of the PC. PCA models for each pair of localities were calculated for both LIBS and EMP datasets (Figs. 7 and 8, respectively); representative loading plots are presented in Figure 11. For the EMP data, score plots of PCA data indicated a more significant influence of elements Mn, Cu, Al, Ca, and F with lesser influences of K when separating sources (Fig. 11). Previous paraíba provenance determinations typically use the quantities of six elements (Cu, Zn, Ga, Sr, Sn, Pb), obtained by LA-ICP-MS, for discrimination of geographic source (e.g., Katsurada et al. 2019). While Cu-Zn-Pb are more readily acquired by EMP and Ga to some degree, Pb and Sn are generally below EMP detection limits (typically <0.001 wt% oxide). Although these elements are below detection in EMP datasets, they are not for LIBS data acquisition. The dominant elements in LIBS loading plots are Cu, Mn, Fe, Mg, Ti, Zn, K, H, Co, V, Li, and Na. Interestingly, Ca, Sr, Sn, and Pb were not observed in loading plots for the LIBS data, suggesting these elements did not exert a major influence on the separation of localities for tourmalines studied here (Fig. 11). That implies less that there is something missing from LIBS, but, perhaps, that different elements may enhance geographic discrimination (e.g. K, Bi, Mn, F). Such information allows development of alternative diagrams for facilitating provenance determination with compositions determined by LA-ICP-MS. Additionally, this study indicates that statistical analyses of the two techniques, LIBS and EMP, emphasize different elements.

Even with the analytical limitations of each technique, robust results for geographic provenance are attained.

While the high success for discriminating provenance of remarkably similar tourmaline compositions is encouraging, there are caveats. The paraíba sample set analyzed here is relatively small with limited variability, in part due to the rarity and cost of materials. No Ca-dominant Cu-bearing tourmalines were among those analyzed, although these are straightforward to distinguish chemically by their Ca concentrations. Not all Cu-bearing tourmalines analyzed display the characteristic "neon" blue hue of the prized Paraíba tourmalines (Fig. 1). Green, greenish-blue and violet hues were included in the sample sets to capture the likely range of chemical variability for Cu-bearing tourmalines. Additionally, a large area is needed for the optimal number of LIBS analyses coupled with the 80-micrometer spot size. If the sample is zoned, the LIBS analytical spot can include overlapping chemical zones unlike data obtained with the EMP. Other multivariate techniques such as Bayesian Statistics (e.g., McManus et al. 2018) or machine learning algorithms might enhance the discrimination further.

Overall, these data demonstrate that spectra obtained by LIBS can be used to provide provenance discrimination when coupled with multivariate statistics. Analyses are rapid with minimal required sample preparation. Loading plots facilitate identification of important elements in discriminating sample localities and can be used to decipher potentially new criteria for provenance determination. Moreover, multivariate analyses of EMP data also allow categories to be differentiated based on more readily obtained chemical data. Application of the multivariate statistics to EMP data suggest that K, Bi, Mn and F may be additional provenance discriminators. Together these data elucidate elements most useful for geographic discrimination of localities and the sourcing of paraíba tourmaline.

## GEOLOGIC IMPLICATIONS

Determining the provenance of mineral grains separated from their host rock has, for example, revolutionized paleogeographic reconstructions, and provided new data on uplift histories and drainage basin development. While many provenance studies rely on zircon ages, expanding the types of detrital minerals used for provenance determination adds new, unexpected opportunities for past geologic reconstructions – the tourmaline source rock types, and for some compositions, the geographic locality can be distinguished. Additionally, in this time of conflict minerals, it is critically important to be able to source conflict gems and metals. This study provides a case study for new methods that allow minerals of very similar compositions to be separated based on chemical parameters. This study shows, for the first time, the power of multivariate statistics applied to EMP data for separating tourmaline localities. Multivariate statistics applied to LIBS and EMP data provide a robust tool for provenance discrimination of Paraiba tourmalines, distinguishing Brazilian-sourced samples from African-sourced materials. Accurate sourcing of gemstones has economic implications as does the sourcing of conflict stones, particularly when economic sanctions may be in place.

## ACKNOWLEDGMENTS

Funding for this project was provided through NSF-IF 1551434 (to Dutrow and Henry) and 1551415 (to McMillan) and is gratefully acknowledged. Robert Wagner- Beija Flor Gems donated select Brazilan samples from São José da Batalha, Paul Wild donated the Nigerian and Mozambican materials and Brendan Laurs facilitated this connection for African material; all are thanked for their generosity. Materialytics kindly allowed usage of their LIBS instrument, with special thanks to Catherine McManus and her team. Shoshauna Farnsworth-Pinkerton and Janelle Hansen facilitated data acquisition. Brian Cook supplied additional information on the Brazilian localities, his insights are appreciated. Matt Wortel, Univ Iowa, is thanked for excellent

craftsmanship in making and polishing grain mounts. Insightful reviews by Russ Harmon and Beatrice Celata helped clarify and improve the paper and are gratefully appreciated.

## REFERENCES:

Abduriyim, A., and Kitawaki, H. (2005) Gem News International: Cu- and Mn-bearing tourmaline—More production from Mozambique. Gems & Gemology, 41, 360–361.

Abduriyim, A., Kitawaki, H., Furuya, M., and Schwarz, D. (2006) "Paraiba"-type copper-bearing tourmaline from Brazil, Nigeria, and Mozambique: Chemical Fingerprinting by LA-ICP-MS. Gems & Gemology, 42, 4–21.

Cremers, D.A., and Radziemski, L.J. (2013) Handbook of Laser-Induced Breakdown Spectroscopy. 283 p. John Wiley & Sons Ltd. Chichester UK.

Doucet, F.R., Lithgow, G., Kosierb, R., Bouchard, P., and Sabsabi, M. (2011) Determination of isotope ratios using Laser-Induced Breakdown Spectroscopy in ambient air at atmospheric pressure for nuclear forensics. Journal of Analytical Atomic Spectrometry, 26, 536-541.

Dutrow, B.L., Farnsworth-Pinkerton, S., Henry, D.J., McMillan, N.J. and Niepagen, N. (2019) Copper-bearing tourmaline sources: evidence from Laser-Induced Breakdown Spectroscopy (LIBS) and electron microprobe analyses (EMP). GSA Abstracts with Programs. 51, No. 5, doi: 10.1130/abs/2019AM-340108

Dutrow, B.L., and Henry, D.J. (2016) Fibrous Tourmaline: A sensitive probe of fluid compositions and petrologic environments. The Canadian Mineralogist, 54, 311-335.

Esbensen, K.H. (2004) Multivariate Data Analysis in practice: An introduction to multivariate

data analysis and experimental design. Multivariate Data Analysis, 598 p. CAMO Process AS, Oslo, Norway.

Fabre, C. (2020) Advances in Laser-Induced Breakdown Spectroscopy analysis for geology: A critical review. Spectrochimica Acta Part B, 166, 105799.

Fritsch, E., Shigley, J.E., Rossman, G.R., Mercer, M.E., Muhlmeister, S.M., and Moon, M. (1990) Gem-quality cuprian-elbaite tourmalines from São José da Batalha, Paraíba, Brazil. Gems & Gemology, 26, 189–205. http://dx.doi.org/10.5741/GEMS.26.3.189

Hark, R.R., Remus, J.J., East, L.J., Harmon, R.S., Wise, M.A., Tansi, B.M., Shughrue, K.M., Dunsin, K.S., and Liu, C. (2012) Geographical analysis of "conflict minerals" utilizing laser-induced breakdown spectroscopy. Spectrochimica Acta Part B, B74-75, 131-136.

Gyftokostas, N., Stefas, D., and Couris, S. (2020) Olive oils classification via Laser-Induced Breakdown Spectroscopy. Applied Sciences, 10, 3462.

Harmon, R.S. and Senesi, G.S. (2021) Laser-Induced Breakdown Spectroscopy – A geochemical tool for the 21st century. Applied Geochemistry, 128, 104929.

Henry, D.J., and Dutrow, B.L. (1996) Metamorphic tourmaline and its petrologic applications. In: E.S. Grew and L.M. Anovitz, L.M., Eds., Boron: mineralogy, petrology and geochemistry, 33, p. 503-557, Reviews of Mineralogy, Mineralogical Society of America, Washington, D.C

Henry, D.J., and Dutrow, B.L. (2018) Tourmaline studies through time: contributions to scientific advancements. Journal of Geosciences, 63, 77-98.

Henry, D.J., and Guidotti, C.V. (1985) Tourmaline as a petrogenetic indicator mineral - an

example from the staurolite-grade metapelites of NW Maine. American Mineralogist, 70, 1-15.

Henry, D.J., Novák, M., Hawthorne, F.C., Ertl, A., Dutrow, B.L., Uher, P., and Pezzotta, F. (2011) Nomenclature of the tourmaline-super- group minerals. American Mineralogist, 96, 895–913. http://dx.doi.org/10.2138/am.2011.3636

Katsurada, Y., and Sun, Z. (2017) Cuprian liddicoatite tourmaline. Gems & Gemology, 53, 34–41. http://dx.doi.org/10.5741/GEMS.53.1.34

Katsurada, Y., Sun, Z., Breeding, B.L., and Dutrow, B.L. (2019) Geographic origin determination of Paraíba tourmaline. Gems & Gemology, 55, 648-659.

Kochelek, K.A., McMillan, N.J., McManus, C.E., and Daniel, D.L. (2015) Provenance determination of sapphires and rubies using laser-induced breakdown spectroscopy and multivariate analysis. American Mineralogist, 100, 1921-1931.

Koivula, J.J. and Kammerling, R.C. (1989) Gem news: Paraíba tourmaline update. Gems & Gemology, 25(4), 248.

Kramida, A., Ralchenko, Yu., Reader, J. and NIST ASD Team (2022). *NIST Atomic Spectra Database* (version 5.10), [Online]. Available: https://physics.nist.gov/asd [Tue Nov 07 2023]. National Institute of Standards and Technology, Gaithersburg, MD. DOI: https://doi.org/10.18434/T4W30F

Laboratory Manual Harmonisation Committee (LMHC) (2012) LMHC Information Sheet #6: "Paraíba tourmaline" Version 7, https://static1.squarespace.com/static/5bfbb7e6cc8fed3bb9293

Laurs, B.M., Zwaan, J.C., Breeding, C.M., Simmons, W.B., Beaton, D., Rijsdijk, K.F., Befi, R.,

and Falster, A.U. (2008) Copper-bearing (Paraíba-type) tourmaline from Mozambique. Gems & Gemology, 44, 4–30. http://dx.doi.org/10.5741/GEMS.44.1.4

Lohmeier, S., Lottermoser, B.G., Strauss, K., Adolffs, T., Sindern, S, and Gallhhofer, D. (2021) Nearshore marine garnet and magnetite placers in the Erongo and S-Kunene regions, Namibia. Journal of African Earth Sciences, 180. https://doi.org/10.1016/j.jafrearsci.2021.104221

Ludwig, T., Marschall, H.R., von Strandmann, P.P., Shabaga, B.M., Fayek, M., and Hawthorne, F. C. (2011) A secondary ion mass spectrometry (SIMS) re-evaluation of B and Li isotopic compositions of Cu-bearing elbaite from three global localities. Mineralogical Magazine, 75, 2485-2494.

McManus, C.E., Dowe, J., and McMillan, N.J. (2018) Quantagenetics® analysis of laser-induced breakdown spectroscopic data: Rapid and accurate authentication of materials: Spectrochimica Acta Part B, B145, 79-85.

McManus, C.E., McMillan, N.J., Dowe, J., and Bell, J. (2020) Diamonds certify themselves: Multivariate statistical provenance analysis. Minerals, 10, 916, http://doi:10.3390/min10100916

McMillan, N.J., Montoya, C., and Chesner, W.H. (2012) Correlation of limestone beds using laser-induced breakdown spectroscopy and chemometric analysis. Applied Optics, 51, B213-B222.

McMillan, N.J., Curry, J., Dutrow, B.L., and Henry, D.J. (2018) Identification of the host lithology of tourmaline using Laser-Induced Breakdown Spectroscopy for application in sediment provenance and mineral exploration. The Canadian Mineralogist, 56, 393-410.

McMillan, N.J., Dutrow, B.L., Henry, D.J., and McManus, C.E. (2019) Methods for tourmaline

analyses by laser-induced breakdown spectroscopy. Abstracts with Programs - Geological Society of America. Geological Society of America (GSA): Boulder, CO, United States, Phoenix, AZ.

Morton, A.C., Whitham, A.G., and Fanning, C.M. (2005) Provenance of Late Cretaceous to Paleocene submarine fan sandstones in the Norwegian Sea: Integration of heavy mineral, mineral chemical and zircon age data. Sedimentary Geology, 182, 3-28.

Multari, R.A., Cremers, D.A., Dupre, J.M., and Gustafson, J.E. (2010) The use of laser-induced breakdown spectroscopy for distinguishing between bacterial pathogen species and strains. Applied Spectroscopy, 64, 750-9.

Okrusch, M., Ertl, A., Schüssler, U., Tillmanns, E., Bratz, H., and Bank, H. (2016) Major-and trace-element composition of Paraíba-type tourmaline from Brazil, Mozambique and Nigeria. The Journal of Gemmology, 35, 120-139.

Palke, A.C., Wong, J., Verdel, C., and Avila, J.N. (2018) A common origin for Thai/Cambodian rubies and blue and violet sapphires from Yogo Gulch, Montana, USA? American Mineralogist, 103, 469-479.

Peretti, A., Bieri, W., Reusser, E., Hametner, K., and Gunther, D. (2009) Chemical variations in multicolored “Paraíba”-type tourmalines from Brazil and Mozambique: implications for origin and authenticity determination. Contributions to Gemology, 9, 1-84.

Pesquera, A., Gil-Crespo, P.P., Torres-Ruiz, F., Torres-Ruiz, J., and Roda-Robles, E. (2016) A multiple regression method for estimating Li in tourmaline from electron microprobe analyses. Mineralogical Magazine, 80, 1129-1133.

Rossman, G.R., Fritsch, E., and Shigley, J.E. (1991) Origin of color in cuprian elbaite from São José da Batalha, Paraíba, Brazil. American Mineralogist, 76, 1479–1484.

Russo, R.E., Bol'shakov, A.A., Mao, X., McKay, C.P., Perry,. D.L., and Sorkhabi, O. (2011) Laser Ablation Molecular Isotopic Spectrometry: Spectrochimica Act Part B, 66, 99-104.

Schenk, E.R., and Almirall, R. (2012) Elemental analysis of cotton by laser-induced breakdown spectroscopy. Applied Optics, 49. C153-C160.

Serrano, J., Moros, J., and Laserna, J.J. (2015) Sensing signatures mediated by chemical structure of molecular solids in laser-induced plasmas. Analytical Chemistry, 87, 2794-2801.

Shigley, J.E., Cook, B.C., Laurs, B.M. and de Oliveira Bernardes, M. (2001) An update on "Paraíba" tourmaline from Brazil. Gems & Gemology, 37, 260–276. http://dx.doi.org/10.5741/GEMS.37.4.260

Smith, C.P., Bosshart, G., and Schwarz, D. (2001) Gem News International: Nigeria as a new source of copper-manganese-bearing tourmaline. Gems & Gemology, 37, 239-240.

Smith, C.A., Martinez, M.A., Veirs, D.K., and Cremers, D.A. (2002) Pu-239/Pu-240 isotope ratios determined using high resolution emission spectroscopy in a laser-induced plasma. Spectrochimica Acta Part B: Atomic Spectroscopy, 57, 929-937.

van Hinsberg, V.J., Henry, D.J., and Marschall, H.R. (2011a) Tourmaline: An ideal indicator of its host environment. Canadian Mineralogist, 49, 1-16.

van Hinsberg, V.J., Henry, D.J., and Dutrow, B.L. (2011b) Tourmaline as a petrologic forensic mineral: A unique recorder of its geologic past. Elements, 7(5), 327-332.

von Eynatten, H., and Gaupp, R. (1999) Provenance of Cretaceous synorogenic sandstones in the Eastern Alps: constraints from framework petrography, heavy mineral analysis and mineral chemistry. Sedimentary Geology, 124, 81-111.

Wentzell, C.Y. (2004) Lab Notes: Copper-bearing color-change tourmaline from Mozambique. Gems & Gemology, 40, 50–251.

Wold, S., Sjöström, M., and Eriksson, L. (2001) PLS-regression: a basic tool of chemometrics. Chemometrics and Intelligent Laboratory Systems, 58, 109-130.

## List of Figure Captions

**Figure 1**. Photo of a rough and facetted Paraíba tourmaline displaying the desired brillant neon blue color. Crystal weighs 8.80 ct, from Brazil. The facetted stone is a 10.91 ct neon blue Paraiba tourmaline (no heat) from the Batalha mine, Brazil. Cut gem courtesy of a Private Collector and Mona Lee Nesseth, Custom Estate Jewels. Photo composite: Robert Weldon/© GIA. Used by permission of GIA.

**Figure 2**. Selection of rough Paraiba tourmaline samples from the four localities used for LIBS investigation. **(a)** São José de Batalha, Brazil: Paraíba tourmaline samples with blue, green, purple and pink colors, and with notable color zoning, **(b)** Rio Grande do Norte, Brazil: Paraíba tourmaline samples with similar blue, green, purple, and pink colors, **(c)** Mozambique: Paraíba tourmaline samples with blue, green, lavender and pink colors, and **(d)** Nigeria: Blue and green Paraíba tourmaline samples (see Appendix 1 for more details).

**Figure 3.** Graphs showing the calibration **(left column)** and validation **(right column)** results for the LIBS decision tree based on geographic locality. Samples used in the calibration set were not used in the validation set, leading to a different data distribution. The value of apparent distinction (VAD) determined in the calibration, shown as a dashed line, remains the same in the validation set. Note the change in scale. Each model indicates the number of samples correctly identified of the total number of samples from that locality and is given as a percent success. Because Mozambique is the final locality distinguished, it has the same success as the final model – Nigeria. Overall, 95% of validation spectra were correctly classified. See text for details.

**Figure 4.** Selection of polished paraíba tourmaline samples, in epoxy, used for EMP data collection with sample numbers. **(a)** Color-zoned Brazilian paraíba tourmaline grain from the São José de Batalha (SJdB) and Rio Grande do Norte (BZ) localities. White arrow shows the location of a detailed EMP traverse. **(b)** Variety of colored Paraiba tourmaline grains from Mozambique. **(c)** Paraíba tourmaline grains from Nigeria. (see Table 1; Appendix 1; and Supplemental Data for more information).

**Figure 5**. Representative LIBS spectrum from each of the four different localities for paraíba tourmaline, stacked to show alignment of peaks. Brazilian localities are separated into: Brazil, SJdB for São José de Batalha; and Brazil, RGdN for Rio Grande do Norte. Selected major and minor elements are labeled. The black vertical line combines a number of features, two Ca and two Al emission lines within the labeling line.

**Figure 6.** Decision tree for PSLR modeling of LIBS spectra for paraíba tourmalines shown with percent correctly predicted (success) for each locality. After locality samples are modeled, they are removed from all subsequent models. Numerical value of 1 refers to samples belonging to that model dataset, 0 indicates all others. See Fig. 3 and text for details.

**Figure 7.** PCA score plot calculated using LIBS spectra of paraíba tourmaline samples from four localities, shown by different symbols. Samples from different localities overlap and lack distinct data clustering per locality. PC1 accounts for 36% of the variance in the dataset; PC2 accounts for 13%. PCA plots are used to determine the sequence of PLSR models.

**Figure 8.** PCA score plot for all EMP analyses used in this study. Specific samples and localities are given in different colors and symbols by sample number. Analyses from each sample plot in discrete clusters, but clear distinctions between tourmalines from different countries are not apparent.

**Figure 9.** PLSR decision tree based on EMP data with percent of the samples correctly identified as belonging to the known locality shown as success. Both localities in Brazil were grouped together because of the small sample set and represented by “Brazil”. See Fig. 6 for details.

**Figure 10.** Calibration **(left)** and validation **(right)** results for EMP decision tree. The dashed line indicates the selected VAD, value of apparent distinction. Each colored symbol represents a different locality as given. Overall, 87% of validation analyses were correctly classified.

**Figure 11.** Representative loading plots for PCA models of tourmaline compositions by EMP **(A. and B.)** and LIBS **(C. and D.)**. Variables (elements for EMP and peaks for LIBS) that influence the direction of the principal component through the data set have high positive or negative values, depending on the direction of the influence. These elements exist at different concentrations in the samples modeled. Variables with values close to zero do not vary significantly among the samples.

**Appendix Text**

**Appendix 1:** Tourmaline samples used in the LIBS and EMP study

| Sample location / region | Sample number (all used for LIBS) | EMP | Sample type | Description | Source |
|---|---|---|---|---|---|
| **Brazil** | | | | | |
| Rio Grande del Norte | 18-BZ-Rgdn-bd-1 | | Crystals (xls) | chip, prismatic zoned crystals with pink core and blue rims; In Ab | Dutrow |
| Rio Grande del Norte | 18-BZ-Rgdn-bd-2 | | xls | billet with 2 lg xls, prismatic zoned xl - pink-blue, other chips with pink/blue/black, all low peaks | Dutrow |
| Rio Grande del Norte | 18-BZ-Par-1 | | xl | 1.5 cm xl, zoned - blue-pink, cats eye fibers on one side | Brian Cook (purchase) |
| Rio Grande del Norte | 18-BZ-Par-2 | **x** | xl | 1 cm xl, zoned - blue-pink | Brian Cook (purchase) |
| Rio Grande del Norte | 18-BZ-Par-3 | | xl | 1 cm xl, zoned - blue-pink. Half xl | Brian Cook (purchase) |
| Rio Grande del Norte | 18-BZ-Par-4 | | xl | 1 cm thin xl, zoned - blue-pink | Brian Cook (purchase) |
| Rio Grande del Norte | 18-BZ-Par-5 | | grain | 0.5 cm xl, zoned - mostly pink some blue | Brian Cook (purchase) |
| Rio Grande del Norte | 18-BZ-Par-6 | | grain | 0.5 cm grain mostly dk pink | Brian Cook (purchase) |
| Rio Grande del Norte | 18-BZ-Par-7 | | grain | 0.5 cm grain mostly dk pink | Brian Cook (purchase) |
| Rio Grande del Norte | 18-BZ-Par-8 | | grain | 0.5 cm grain mostly dk pink | Brian Cook (purchase) |
| Rio Grande del Norte | 18-BZ-Par-9 | **x** | grain | 0.3 cm grain mostly dk pink | Brian Cook (purchase) |
| Brazil – unknown region | 18-BZ-Par-unk - 1 | **x** | 4 grains | t.s. billet, thin section | Dutrow |
| Brazil – unknown region | 18-BZ-Par-unk - 2 | **x** | grain | 2x2 cm single xl; zoned with pink/blue | Dutrow |
| | | | | | |
| São José de Batalha | 19-BZ-SJdB-par-10 | **x** | grain | 1 cm thin xl, zoned - blue-pink | Beija Flor Gems - Robert Van Wagoner (donation) |
| São José de Batalha | 19-BZ-SJdB-par-11 | | grain | 3 mm thin xl, zoned - blue-pink | Beija Flor Gems - Robert Van Wagoner (donation) |

| | | | | | |
|---|---|---|---|---|---|
| São José de Batalha | 19-BZ-SJdB-par-12 | **x** | grain | 3 mm thin xl, zoned - blue-pink | Beija Flor Gems - Robert Van Wagoner (donation) |
| São José de Batalha | 19-BZ-SJdB-par-13 | | grain | 3mm thin xl, zoned - blue-pink | Beija Flor Gems - Robert Van Wagoner (donation) |
| São José de Batalha | 19-BZ-SJdB-par-14 | **x** | grain | 5 mm thin xl, zoned - blue-pink | Beija Flor Gems - Robert Van Wagoner (donation) |
| São José de Batalha | 19-BZ-SJdB-par-15 | | grain | 3mm thin xl, zoned - blue-pink | Beija Flor Gems - Robert Van Wagoner (donation) |
| | **Mozambique** | | | | |
| | 18-Moz-Par-1 | | grain | 1 cm pale blue with pink rim | Paul Wild (donation) |
| | 18-Moz-Par-2 | | grain | blue - 6 mm | Paul Wild (donation) |
| | 18-Moz-Par-3 | | grain | pale green-blue - 1 cm | Paul Wild (donation) |
| | 18-Moz-Par-4 | | grain | pale green - I cm | Paul Wild (donation) |
| | 18-Moz-Par-5 | **x** | grain | blue-green | Paul Wild (donation) |
| | 18-Moz-Par-6 | | grain | pink - 5 mm | Paul Wild (donation) |
| | 18-Moz-Par-7 | | grain | blue - green - 5 mm | Paul Wild (donation) |
| | 18-Moz-Par-8 | | grain | pale-blue - 3mm | Paul Wild (donation) |
| | 18-Moz-Par-9 | | grain | pale-blue green - 7 mm | Paul Wild (donation) |
| | 18-Moz-Par-10 | **x** | grain | light purple | Paul Wild (donation) |
| | 18-Moz-Par-11 | | grain | blue-green - 6 mm | Paul Wild (donation) |
| | 18-Moz-Par-12 | | grain | pink - 1 cm | Paul Wild (donation) |
| | 18-Moz-Par-13 | | grain | pale-green - 5 mm | Paul Wild (donation) |
| | 18-Moz-Par-14 | **x** | grain | blue-green | Paul Wild (donation) |
| | 18-Moz-Par-15 | | grain | light-blue- green - 7 mm | Paul Wild (donation) |
| | 18-Moz-Par-16 | | grain | light-green - 1 cm | Paul Wild (donation) |
| | 18-Moz-Par-17 | | grain | blue - 1 cm | Paul Wild (donation) |

| | | | | | |
|---|---|---|---|---|---|
| | 18-Moz-Par-18 | **x** | grain | blue | Paul Wild (donation) |
| | 18-Moz-Par-19 | | grain | pink - 7 mm | Paul Wild (donation) |
| | 18-Moz-Par-20 | **x** | grain | blue and pink | Paul Wild (donation) |
| | 18-Moz-Par-21 | | grain | green - 1 cm | Paul Wild (donation) |
| | 18-Moz-Par-22 | | grain | pale-blue - 1 cm | Paul Wild (donation) |
| | 18-Moz-Par-23 | **x** | grain | pink | Paul Wild (donation) |
| | 18-Moz-Par-24 | | grain | light-blue - 1 cm | Paul Wild (donation) |
| | | | | | |
| | **Nigeria** | | | | Paul Wild (donation) |
| | 18-Nig-Par-1 | | grain | dark green - 1.5 cm | Paul Wild (donation) |
| | 18-Nig-Par-2 | | grain | green - 1 cm | Paul Wild (donation) |
| | 18-Nig-Par-3 | **x** | grain | green | Paul Wild (donation) |
| | 18-Nig-Par-4 | | grain | light blue-green 1 cm | Paul Wild (donation) |
| | 18-Nig-Par-5 | | grain | light green - 1 cm | Paul Wild (donation) |
| | 18-Nig-Par-6 | | grain | light green - 5 mm | Paul Wild (donation) |
| | 18-Nig-Par-7 | **x** | grain | green | Paul Wild (donation) |
| | 18-Nig-Par-8 | | grain | light green - 5 mm | Paul Wild (donation) |
| | 18-Nig-Par-9 | **x** | grain | green | Paul Wild (donation) |
| | 18-Nig-Par-10 | **x** | grain | green | Paul Wild (donation) |
| | 18-Nig-Par-11 | | grain | light green - 3 mm | Paul Wild (donation) |

**TABLE 1**. Representative Paraiba tourmaline analyses of blue regions, average grain compositions and Maximum/Minimum values of the grains.

| Location | Brazil, Rio Grande do Norte | | | | Brazil, São José de Batalha | | | | Mozambique | | | | Nigeria | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Sample # | 18-BZ-Par-2bc | | | | 19-BZ-SJbB-12 | | | | 18-Moz-Par-20 | | | | 18-Nig-Par-10 | | | |
| Analysis | #12 | Grain average | Max | Min | #15 | Grain average | Max | Min | #8 | Grain average | Max | Min | #10 | Grain Average | Max | Min |
| $B_2O_3$ [a] | 11.07 | 11.03 | | | 11.04 | 11.04 | | | 11.05 | 11.08 | | | 10.99 | 11.05 | | |
| $SiO_2$ | 37.32 | 37.39 | 37.96 | 36.93 | 37.45 | 37.46 | 38.22 | 36.95 | 38.09 | 38.16 | 38.51 | 37.53 | 37.60 | 37.93 | 37.93 | 37.16 |
| $Al_2O_3$ | 41.75 | 40.79 | 42.00 | 38.82 | 41.25 | 41.47 | 42.23 | 40.92 | 40.72 | 41.10 | 42.79 | 40.56 | 41.08 | 41.06 | 41.66 | 40.56 |
| $TiO_2$ | 0.00 | 0.03 | 0.10 | 0.00 | 0.03 | 0.01 | 0.06 | 0.00 | 0.01 | 0.01 | 0.06 | 0.00 | 0.08 | 0.06 | 0.08 | 0.00 |
| $V_2O_3$ | 0.00 | 0.00 | 0.03 | 0.00 | 0.00 | 0.00 | 0.02 | 0.00 | 0.00 | 0.01 | 0.03 | 0.00 | 0.00 | 0.00 | 0.02 | 0.00 |
| $Cr_2O_3$ | 0.00 | 0.00 | 0.02 | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 |
| $Fe_2O_3$ [a] | 0.00 | 0.05 | 0.00 | 0.00 | 0.00 | 0.00 | 0.20 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| FeO | 0.00 | 0.00 | 0.18 | 0.00 | 0.00 | 0.03 | 0.00 | 0.00 | 0.01 | 0.02 | 0.04 | 0.00 | 0.02 | 0.19 | 0.19 | 0.00 |
| MnO | 0.59 | 1.44 | 2.34 | 0.06 | 0.14 | 0.73 | 1.47 | 0.01 | 1.02 | 0.86 | 1.02 | 0.56 | 1.41 | 1.82 | 1.82 | 1.36 |
| MgO | 0.00 | 0.01 | 0.08 | 0.00 | 0.00 | 0.01 | 0.09 | 0.00 | 0.01 | 0.00 | 0.01 | 0.00 | 0.00 | 0.01 | 0.01 | 0.00 |
| ZnO | 0.00 | 0.45 | 2.24 | 0.00 | 0.00 | 0.08 | 0.51 | 0.00 | 0.15 | 0.15 | 0.19 | 0.04 | 0.03 | 0.00 | 0.05 | 0.00 |
| CuO | 1.47 | 1.03 | 1.89 | 0.34 | 1.41 | 0.98 | 1.67 | 0.44 | 0.40 | 0.32 | 0.42 | 0.18 | 0.36 | 0.23 | 0.50 | 0.20 |
| $Li_2O$ [b] | 1.84 | 1.82 | | | 1.99 | 1.86 | | | 1.99 | 1.97 | | | 1.82 | 1.76 | | |
| CaO | 0.37 | 0.23 | 0.51 | 0.06 | 0.26 | 0.31 | 0.55 | 0.08 | 0.76 | 0.71 | 0.82 | 0.31 | 0.35 | 0.30 | 0.38 | 0.28 |
| PbO | 0.02 | 0.01 | 0.05 | 0.00 | 0.00 | 0.01 | 0.03 | 0.00 | 0.07 | 0.08 | 0.12 | 0.04 | 0.00 | 0.00 | 0.05 | 0.00 |
| $Na_2O$ | 2.08 | 2.14 | 2.57 | 1.87 | 2.05 | 1.98 | 2.19 | 1.82 | 1.99 | 1.97 | 2.08 | 1.90 | 1.97 | 2.09 | 2.15 | 1.97 |
| $K_2O$ | 0.02 | 0.01 | 0.02 | 0.00 | 0.01 | 0.01 | 0.02 | 0.00 | 0.01 | 0.01 | 0.02 | 0.01 | 0.02 | 0.01 | 0.02 | 0.00 |
| $Bi_2O_3$ | 0.03 | 0.05 | 0.34 | 0.00 | 0.14 | 0.06 | 0.28 | 0.00 | 0.01 | 0.02 | 0.06 | 0.00 | 0.15 | 0.08 | 0.16 | 0.07 |
| F | 1.10 | 1.03 | 1.34 | 0.68 | 0.80 | 0.94 | 1.33 | 0.75 | 1.22 | 1.08 | 1.29 | 0.71 | 0.95 | 1.07 | 1.20 | 0.80 |
| Cl | 0.00 | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 |
| $H_2O$ [a] | 3.25 | 3.32 | | | 3.43 | 3.33 | | | 3.09 | 3.14 | | | 3.20 | 3.12 | | |
| Subtotal | 100.89 | 100.83 | | | 99.99 | 100.31 | | | 100.60 | 100.68 | | | 100.02 | 100.76 | | |
| O=F | 0.46 | 0.43 | | | 0.34 | 0.40 | | | 0.51 | 0.46 | | | 0.40 | 0.45 | | |
| Total | 100.43 | 100.40 | | | 99.66 | 99.91 | | | 100.09 | 100.23 | | | 99.62 | 100.32 | | |
| | 15 Y+Z+T cation normalization | | | | | | | | | | | | | | | |
| B site: B [a] | 3.000 | 3.000 | | | 3.000 | 3.000 | | | 3.000 | 3.000 | | | 3.000 | 3.000 | | |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| T site: Si | 5.861 | 5.893 | 5.897 | 5.898 | 5.990 | 5.985 | 5.944 | 5.947 |
| B | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| Al | 0.139 | 0.107 | 0.103 | 0.102 | 0.010 | 0.015 | 0.056 | 0.053 |
| T site total | 6.000 | 6.000 | 6.000 | 6.000 | 6.000 | 6.000 | 6.000 | 6.000 |
| | | | | | | | | |
| Al (total) | 7.726 | 7.575 | 7.656 | 7.693 | 7.548 | 7.597 | 7.654 | 7.652 |
| | | | | | | | | |
| Z site: Al | 6.000 | 6.000 | 6.000 | 6.000 | 6.000 | 6.000 | 6.000 | 6.000 |
| $Cr^{3+}$ | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| $V^{3+}$ | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| $Fe^{3+}$ [a] | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| Mg | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| Z site total | 6.000 | 6.000 | 6.000 | 6.000 | 6.000 | 6.000 | 6.000 | 6.000 |
| | | | | | | | | |
| Y site: Al | 1.587 | 1.468 | 1.553 | 1.591 | 1.538 | 1.582 | 1.598 | 1.599 |
| Ti | 0.000 | 0.003 | 0.003 | 0.002 | 0.001 | 0.001 | 0.009 | 0.003 |
| $V^{3+}$ | 0.000 | 0.000 | 0.000 | 0.001 | 0.000 | 0.001 | 0.000 | 0.001 |
| $Cr^{3+}$ | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| $Fe^{3+}$ [a] | 0.000 | 0.006 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| $Fe^{2+}$ | 0.000 | 0.000 | 0.000 | 0.004 | 0.001 | 0.002 | 0.002 | 0.009 |
| $Mn^{2+}$ | 0.078 | 0.193 | 0.019 | 0.097 | 0.135 | 0.114 | 0.188 | 0.211 |
| Mg | 0.000 | 0.003 | 0.000 | 0.002 | 0.003 | 0.001 | 0.000 | 0.000 |
| Zn | 0.000 | 0.052 | 0.000 | 0.009 | 0.018 | 0.017 | 0.004 | 0.001 |
| Cu | 0.174 | 0.123 | 0.167 | 0.117 | 0.047 | 0.037 | 0.043 | 0.040 |
| Li [a] | 1.160 | 1.151 | 1.257 | 1.178 | 1.256 | 1.245 | 1.155 | 1.137 |
| Y-site total | 3.000 | 3.000 | 3.000 | 3.000 | 3.000 | 3.000 | 3.000 | 3.000 |
| | | | | | | | | |
| X site: Ca | 0.061 | 0.039 | 0.045 | 0.052 | 0.129 | 0.120 | 0.058 | 0.052 |
| Pb | 0.001 | 0.000 | 0.000 | 0.000 | 0.003 | 0.003 | 0.000 | 0.001 |
| Na | 0.633 | 0.653 | 0.627 | 0.603 | 0.607 | 0.599 | 0.605 | 0.628 |
| K | 0.004 | 0.003 | 0.001 | 0.002 | 0.002 | 0.003 | 0.003 | 0.003 |
| $Bi^{3+}$ | 0.001 | 0.002 | 0.006 | 0.002 | 0.000 | 0.001 | 0.006 | 0.005 |
| X-site vacancy | 0.300 | 0.303 | 0.322 | 0.340 | 0.259 | 0.275 | 0.328 | 0.311 |
| X-site total | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| V+W site: OH | 3.401 | 3.488 | 3.602 | 3.501 | 3.245 | 3.286 | 3.377 | 3.328 |
| V site: OH [a] | 3.000 | 3.000 | 3.000 | 3.000 | 3.000 | 3.000 | 3.000 | 3.000 |
| V site: O | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| W site OH [a] | 0.401 | 0.488 | 0.602 | 0.501 | 0.245 | 0.286 | 0.377 | 0.328 |
| W site: F | 0.546 | 0.512 | 0.398 | 0.467 | 0.605 | 0.538 | 0.474 | 0.505 |
| W site: Cl | 0.000 | 0.001 | 0.000 | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 |
| W site O | 0.053 | 0.000 | 0.000 | 0.031 | 0.150 | 0.176 | 0.149 | 0.166 |
| V,W-site total | 4.000 | 4.000 | 4.000 | 4.000 | 4.000 | 4.000 | 4.000 | 4.000 |
| species | Fluor-elbaite | Fluor-elbaite | Elbaite | Elbaite | Fluor-elbaite | Fluor-elbaite | Fluor-elbaite | Fluor-elbaite |

Notes: Minimum detection limits of minor and trace elements (in wt% oxides): $TiO_2$ = 0.007; $V_2O_3$, $Cr_2O_3$ = 0.013; FeO, MnO, MgO = 0.016; ZnO = 0.026; PbO= 0.021; $K_2O$ = 0.009.

[a] Calculated based on stoichiometry and/or charge balance.

[b] Calculated based on the Li-estimation procedures of Pesquera et al. (2016)

Figure 1

# Figure 2

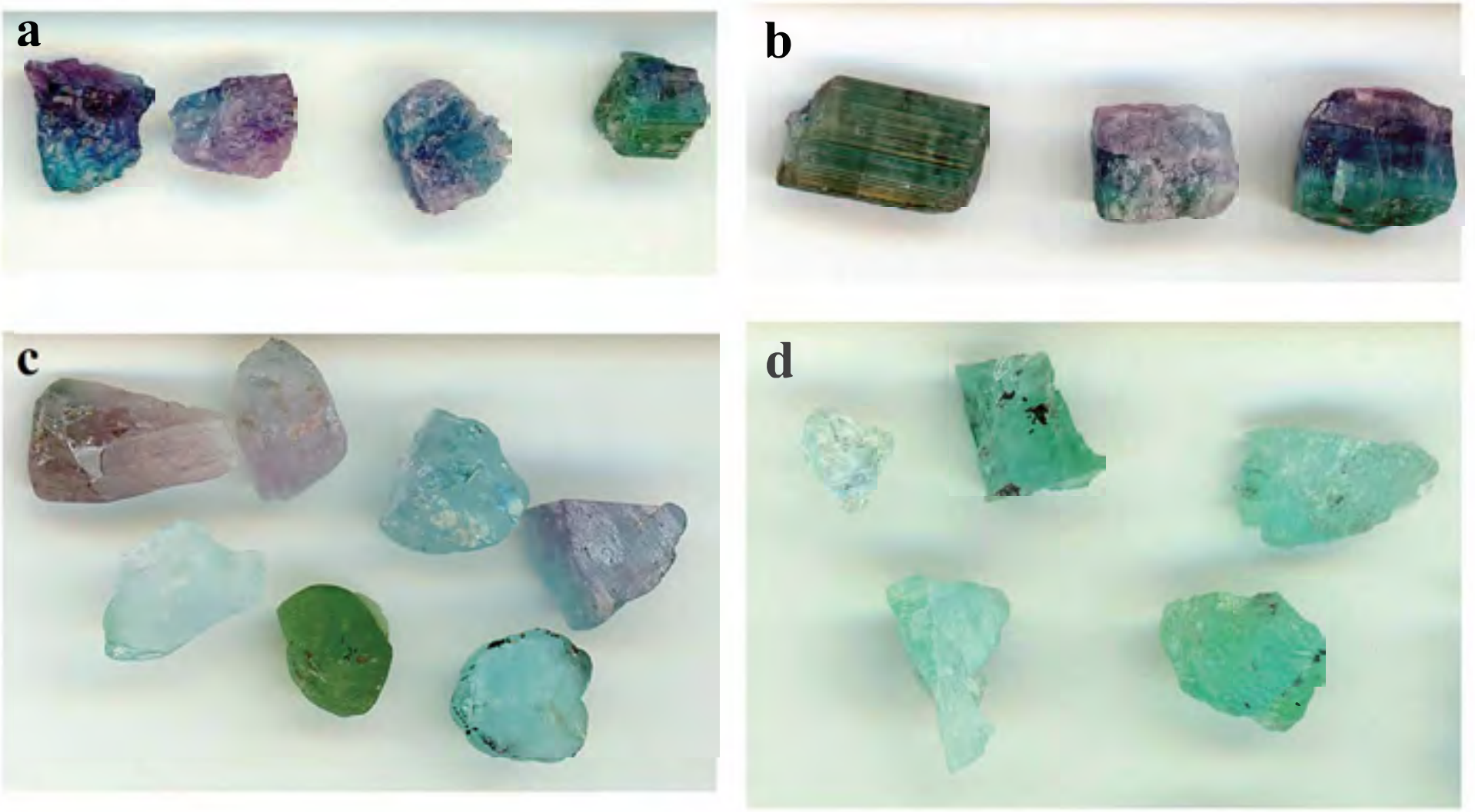


Figure 3

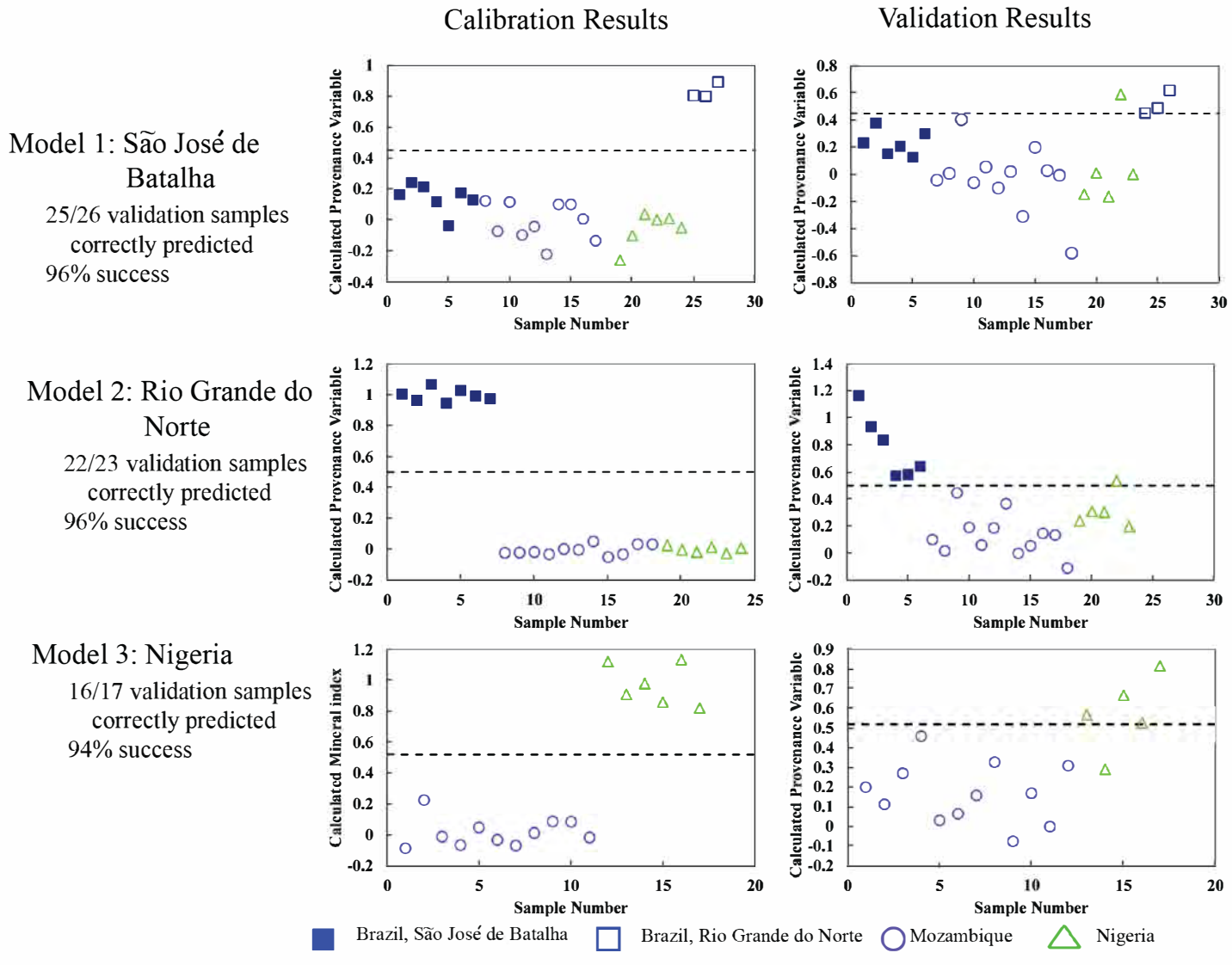

Calibration Results
Validation Results
Model 1: São José de Batalha
25/26 validation samples correctly predicted
96% success
Model 2: Rio Grande do Norte
22/23 validation samples correctly predicted
96% success
Model 3: Nigeria
16/17 validation samples correctly predicted
94% success
Calculated Provenance Variable
Calculated Mineral index
Sample Number
Brazil, São José de Batalha
Brazil, Rio Grande do Norte
Mozambique
Nigeria

**Figure 4**

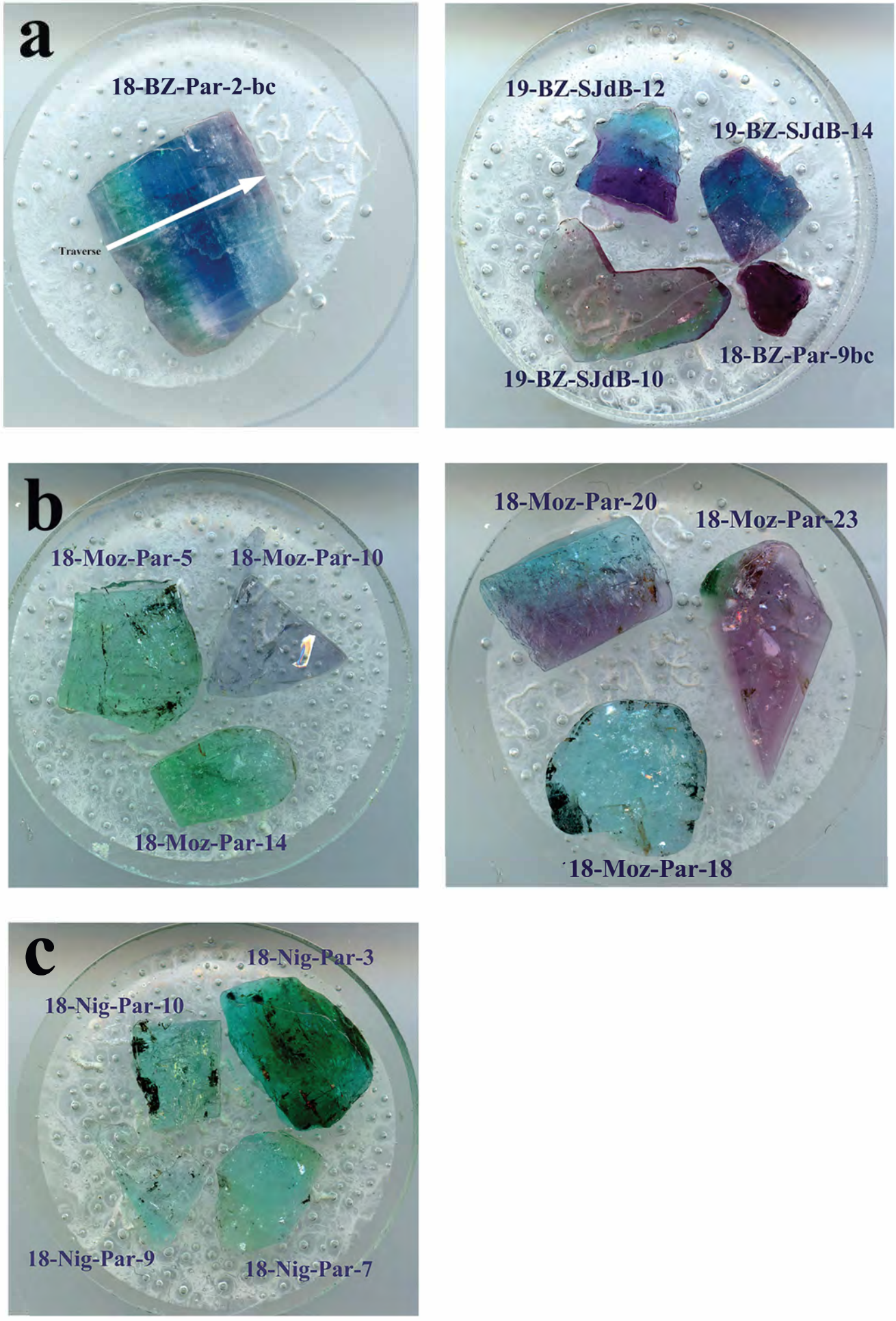

a
18-BZ-Par-2-bc
Traverse
19-BZ-SJdB-12
19-BZ-SJdB-14
19-BZ-SJdB-10
18-BZ-Par-9bc
b
18-Moz-Par-5
18-Moz-Par-10
18-Moz-Par-14
18-Moz-Par-20
18-Moz-Par-23
18-Moz-Par-18
c
18-Nig-Par-3
18-Nig-Par-10
18-Nig-Par-9
18-Nig-Par-7

Figure 5

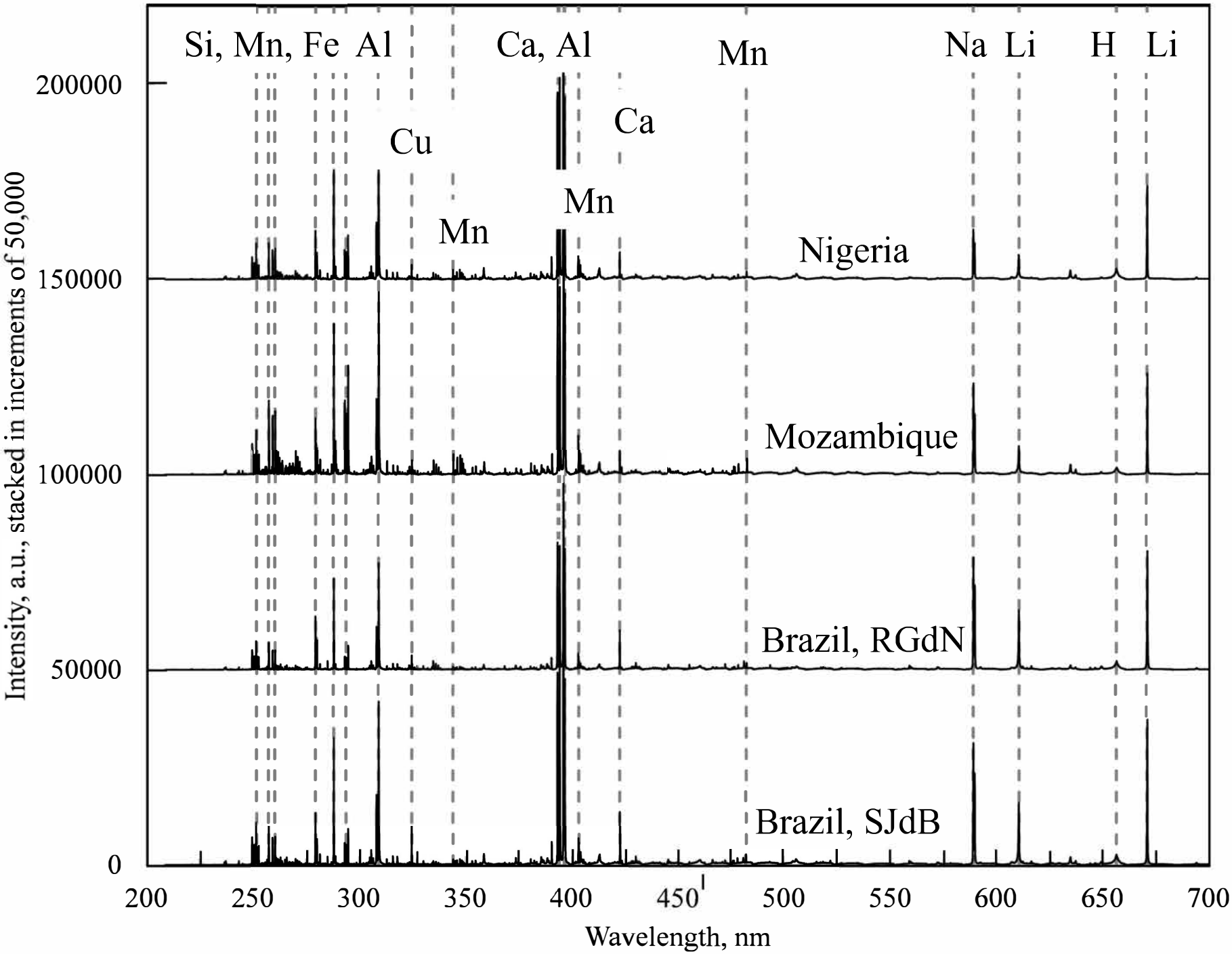
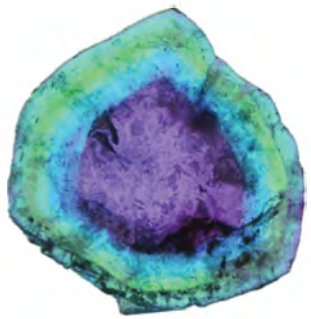

Si, Mn, Fe
Al
Cu
Mn
Ca, Al
Mn
Ca
Mn
Na
Li
H
Li
Nigeria
Mozambique
Brazil, RGdN
Brazil, SJdB
200000
150000
100000
50000
0
Intensity, a.u., stacked in increments of 50,000
200
250
300
350
400
450
500
550
600
650
700
Wavelength, nm

Figure 6

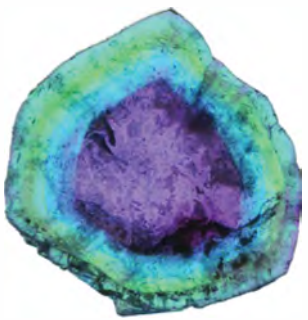


1

Model 1: ***São José de Batalha, Brazil***

0

***96% success***

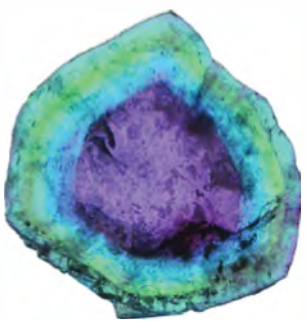


1

Model 2: ***Rio Grande do Norte, Brazil***

0

***96% success***



1

Model 3: ***Nigeria***

***94% success***

0

***Mozambique***

***94% success***

Figure 7

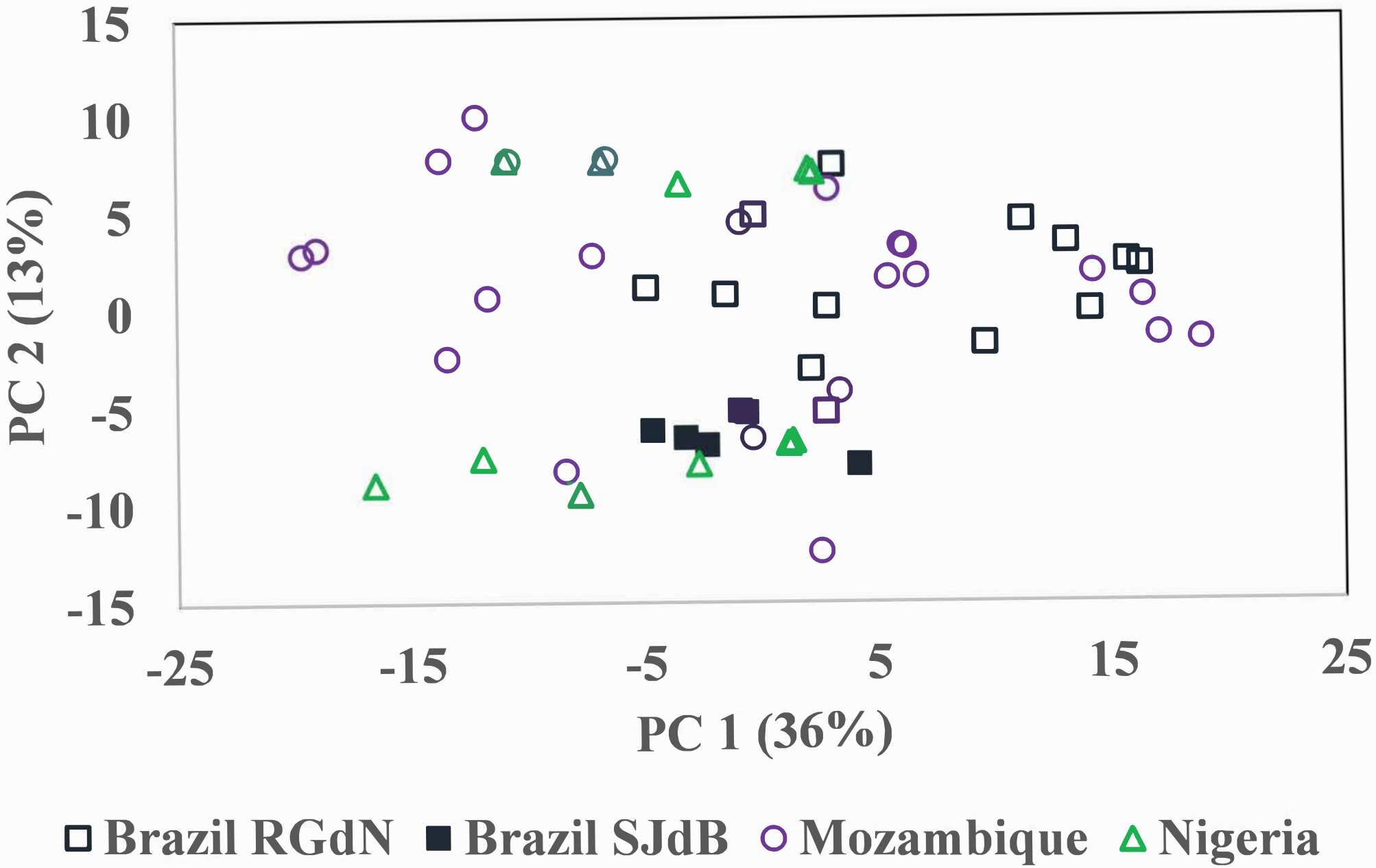

15
10
5
0
-5
-10
-15
PC 2 (13%)
-25
-15
-5
5
15
25
PC 1 (36%)
Brazil RGdN
Brazil SJdB
Mozambique
Nigeria

Figure 8

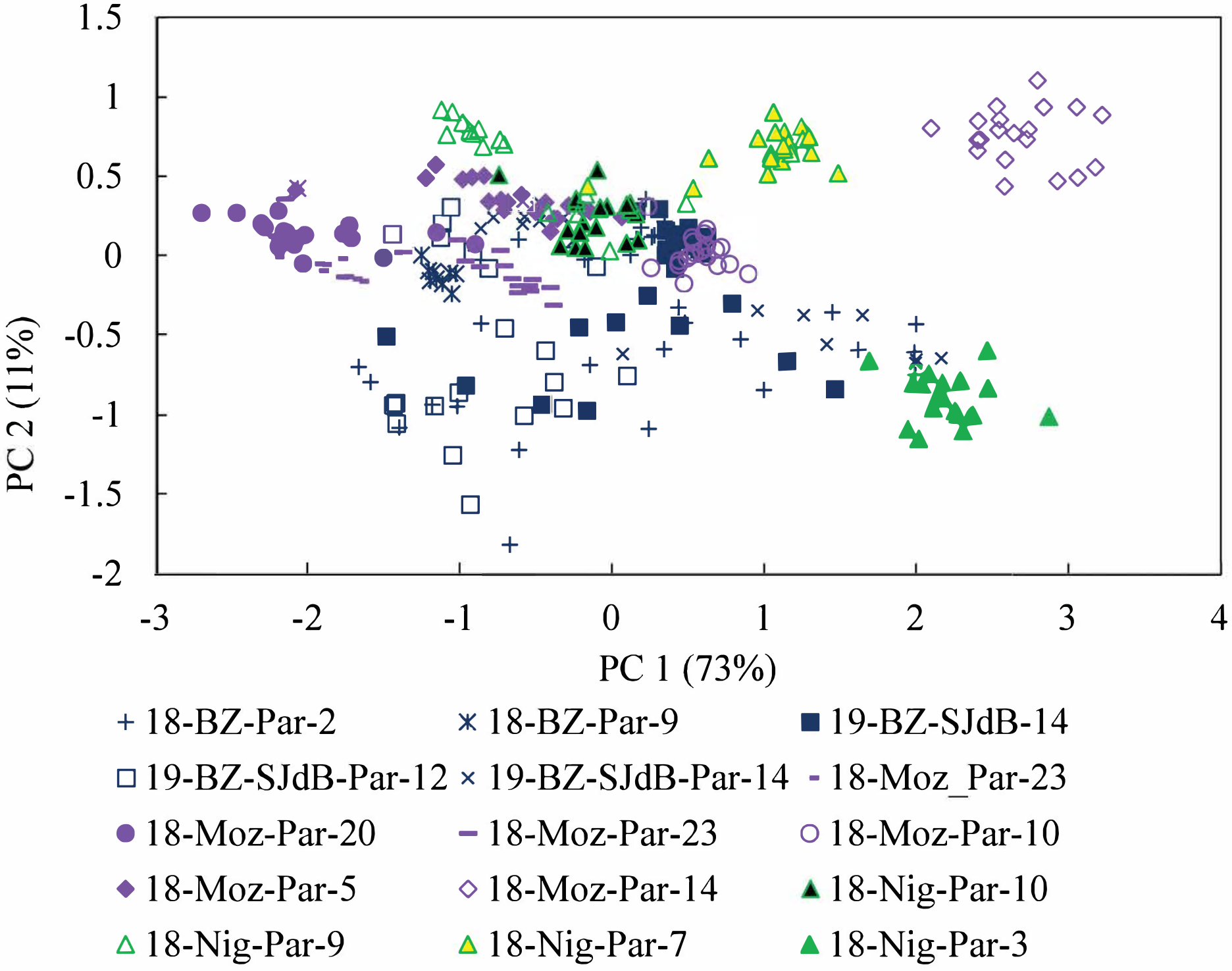

1.5
1
0.5
0
-0.5
-1
-1.5
-2
PC 2 (11%)
-3
-2
-1
0
1
2
3
4
PC 1 (73%)
18-BZ-Par-2
18-BZ-Par-9
19-BZ-SJdB-14
19-BZ-SJdB-Par-12
19-BZ-SJdB-Par-14
18-Moz_Par-23
18-Moz-Par-20
18-Moz-Par-23
18-Moz-Par-10
18-Moz-Par-5
18-Moz-Par-14
18-Nig-Par-10
18-Nig-Par-9
18-Nig-Par-7
18-Nig-Par-3

Figure 9

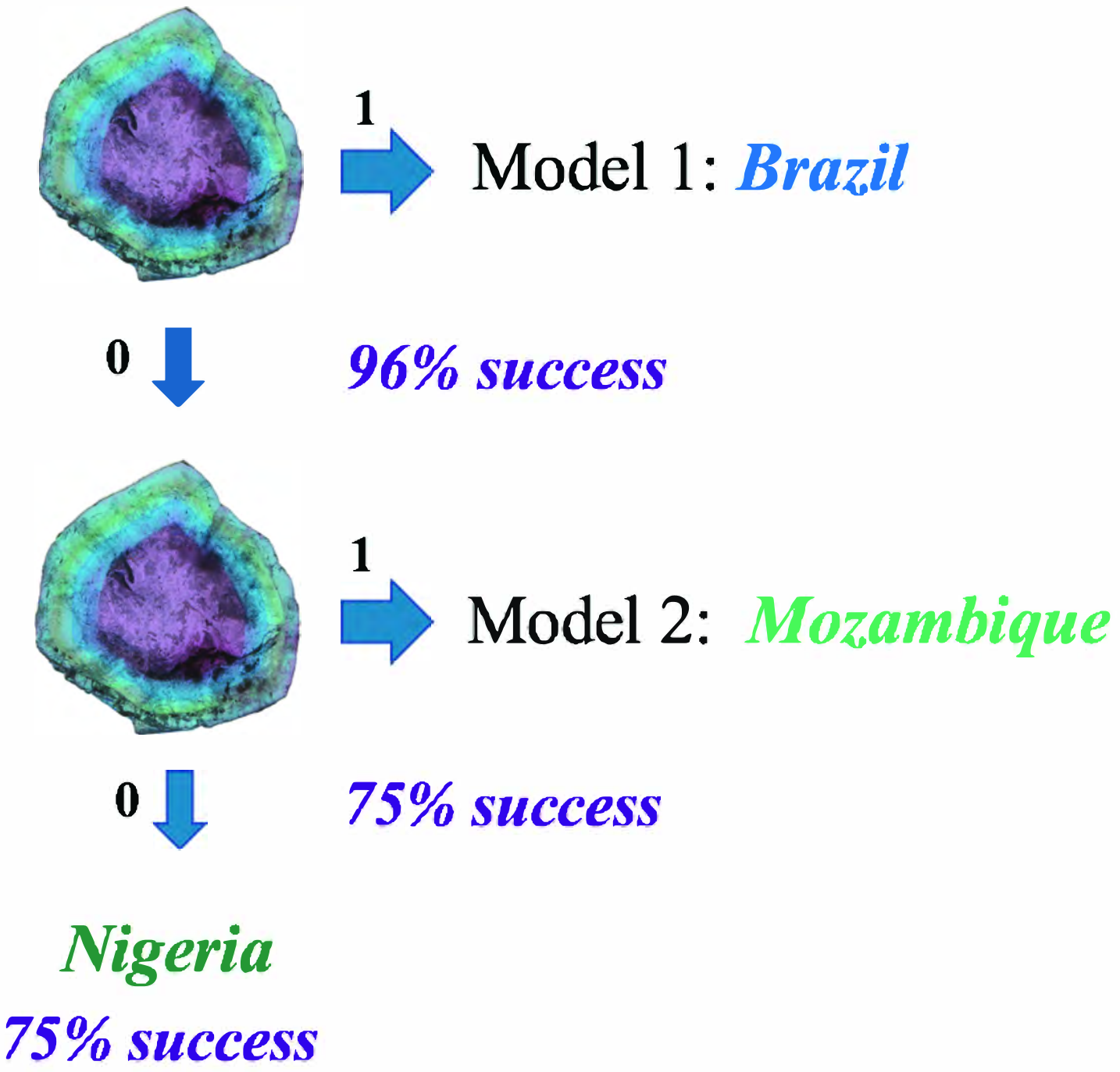

1
Model 1: Brazil
0
96% success
1
Model 2: Mozambique
0
75% success
Nigeria
75% success

Figure 10

Calibration Results

Validation Results

Model 1: Brazil

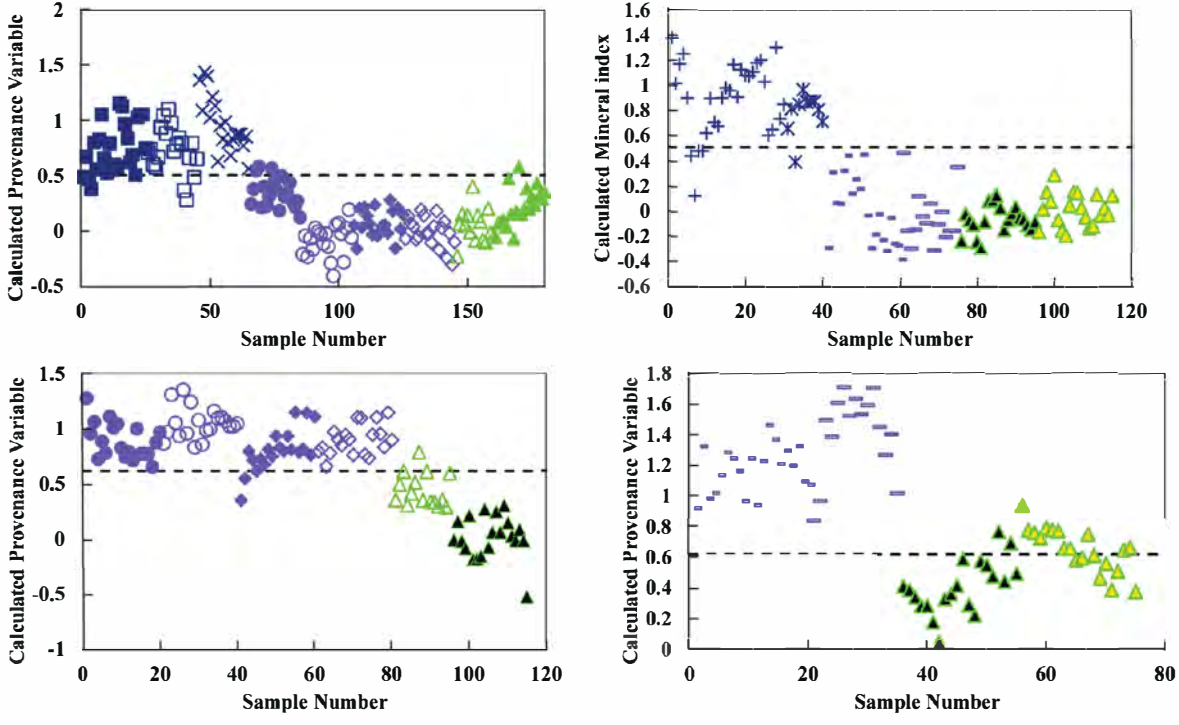
110/115 validation analyses correctly predicted

96% success

Model 2: Mozambique

56/75 validation analyses correctly predicted

75% success

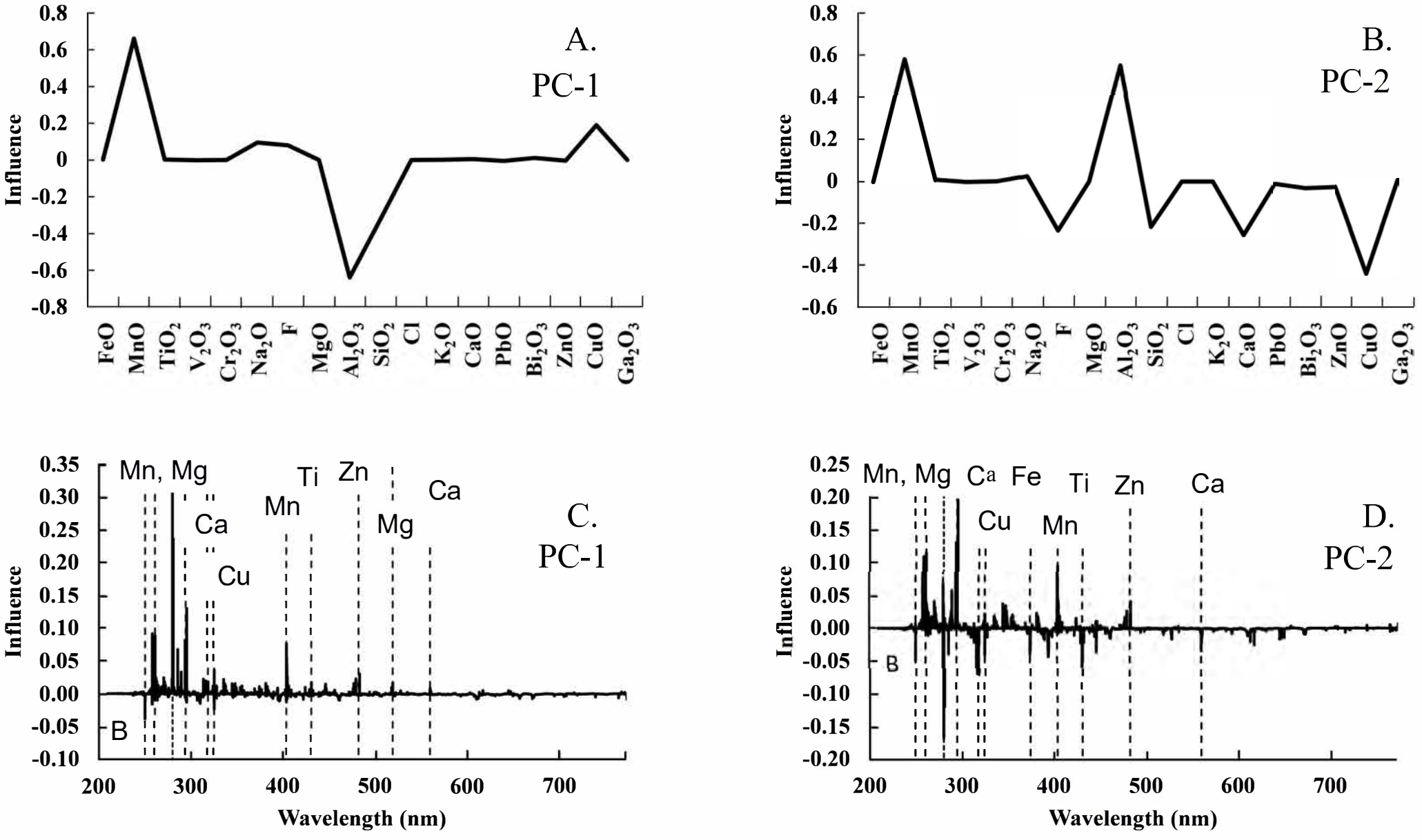



Brazil, both locations

Mozambique

Nigeria

Figure 11


A.
PC-1
Influence
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
FeO
MnO
TiO2
V2O3
Cr2O3
Na2O
F
MgO
Al2O3
SiO2
Cl
K2O
CaO
PbO
Bi2O3
ZnO
CuO
Ga2O3
B.
PC-2
Influence
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
FeO
MnO
TiO2
V2O3
Cr2O3
Na2O
F
MgO
Al2O3
SiO2
Cl
K2O
CaO
PbO
Bi2O3
ZnO
CuO
Ga2O3
C.
PC-1
Influence
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
-0.05
-0.10
Mn, Mg
Ca
Cu
Mn
Ti
Zn
Mg
Ca
B
200
300
400
500
600
700
Wavelength (nm)
D.
PC-2
Influence
0.25
0.20
0.15
0.10
0.05
0.00
-0.05
-0.10
-0.15
-0.20
Mn, Mg
Ca
Fe
Cu
Mn
Ti
Zn
Ca
B
200
300
400
500
600
700
Wavelength (nm)